OPTICAL MEASUREMENTS IN ELASTOHYDRODYNAMIC ROLLING-CONTACT BEARINGS

by Lavern Dale Wedeven

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

of the University of London

and also for the

DIPLOMA OF IMPERIAL COLLEGE

MARCH, 1970

Lubrication Laboratory
Department of Mechanical Engineering
Imperial College
London S.W. 7

ABSTRACT

Elastohydrodynamic oil film measurements for rolling point contact under starvation conditions are obtained using optical interferometry. Consideration of the measurement requirements and the requirements for good fringe visibility has led to the design of a duochromatic interferometric system. A calibration procedure provides the absolute thickness corresponding to each fringe and elucidates the effects of phase change on reflection.

The experimental measurements present a reasonably clear picture of the starvation phenomenon. Film thickness is determined by the pressure generated in the inlet region. A starved inlet region inhibits the generation of this pressure and, therefore, reduces the film thickness. It also causes the overall pressure, stress and elastic deformation to become more Hertzian. The experimental film thickness measurements under starved as well as flooded conditions are shown to agree with theoretical predictions based on Grubin assumptions. The results correlate with characteristic dimensions which describe the degree of starvation in the inlet region. These are used to predict the onset of starvation and are incorporated in a theoretical film thickness formula which enables film thickness under starvation conditions to be predicted.

Additional experiments with interferometry illustrates: the effect of ball spin, edge effects in line contact, oil emtrapment under 'rolling-normal approach' conditions, the behaviour of greases under elastohydrodynamic conditions, and cavitation and its influence on the inlet region.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the following in support of this work.

Dr. A Cameron for his willing and enthusiastic supervision during the course of this work.

Mr. D. Evans of the Imperial College Computer Dept. for his computation of the point contact problem under starvation conditions.

Dr. R. Gohar, Dr. C.A. Foord, and Mr. F.J. Westlake for their valuable discussions and Mr. B.L Klemz for the use of his photoelastic stress results in fig.63.

Mr. R. Dobson for his valuable advice and help on technical matters in the laboratory.

Mr. F. Seabury for assembling the electronic flash unit.

My dear wife Carol for her continued support and her willing efforts in typing the script and drawing some of the figures.

Mr. T. Ellis and the Hoffmann Bearing Company for the supply of the high quality steel balls and other bearing materials.

Mr. F.J. Wren and the British Timken Roller Bearing Company for their efforts in connection with the design and supply of the tapered roller thrust bearing.

Mr. J.F. Hutton and his colleagues at the Thornton Research Centre for their interesting discussions on greases and their supply of lubricants.

Science Research Council for a grant.

Miss M. Suter and the Regent Street Polytechnic School of Photography for photographing some of the equipment.

Dr. R.L. Johnson and the NASA Lewis Research Center for providing the colour print in fig.31.

Dr. K.C. Ludema of the University of Michigan for his encouragement to initiate this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 INTRODUCTION

1.1 Preface

The purpose of lubrication is to control friction and minimize failure. The thickness of the hydrodynamic film between bearing surfaces is a very important consideration, particularly for elastohydrodynamic (ehd) conditions where film thickness and surface irregularities are often of the same order of magnitude.

The present state of ehd, which is reviewed in Chapter 4, shows that the development of an ehd film is now quite well understood, and that the thickness of such films can be predicted theoretically with commendable accuracy. Presently, some investigations are concerned with isolated aspects of ehd and the explanations of small discrepancies between theory and experiment. Considerable emphasis is also being placed on the application of ehd principles to establish the most favourable operating conditions in machine elements.

Many applications, however, do not operate under the ideal conditions which are normally assumed or provided in most theoretical and experimental investigations. An important practical consideration which is often omitted is the supply and distribution of lubricant in the vacinity of the contact region. Lubricant starvation, which may often exist but not always be recognized, can have an overriding effect on film thickness and other ehd aspects. This has received very little attention in the present state of ehd and forms the subject of most of the work presented here.

Most of the experiments to be described were performed under point contact conditions pertaining to a ball thrust bearing which is described in Chapter 2. The ehd

films developed in this bearing were measured with optical interferometry. Although this technique has been used before, the relevant interferometric principles pertaining to its successful application as a technique for film thickness measurement have not always been clearly defined. Some of these principles are presented in Chapter 3 in conjunction with the description of the interferometer which was used.

Preliminary measurements pertaining to the starvation experiments are presented in Chapter 5 while the starvation results are presented and analyzed in Chapter 6. The final chapter deals with additional experiments involving ball spin, edge effects in line contact, entrapment under 'rolling-normal approach' conditions, greases, and cavitation.

1.2 Conception of Project

The conception of the project was a sequal to the sliding experiments of Gohar(1) and the rolling experiments of Foord(2). In these experiments the oil film thicknesses were measured with interferometry under conditions where the ball was rotated about an axis fixed in space. In most common ball bearings the balls are usually contained within a cage and travel between a stationary and rotating race. The axis of ball rotation is determined by the kinetics and dynamics of the system which governs the amount of roll and spin the balls may acquire. Sliding induces heat within the bearing. This will lower the ehd film thickness and may lead to early bearing failure.

It was proposed, therefore, to investigate the oil film thickness in a simple ball thrust bearing where the balls have curvilinear as well as rotational motion. Experimentally

this requires the interference fringes of the contact region of a single ball to be observed stroboscopically and recorded with high-speed flash photography. This technique was achieved. However, it was discovered that the film thickness was generally far more dependent on lubricant supply conditions than on effects attributed to ball spin.

It was then realized that the equipment, with only slight modifications,could conveniently be used to investigate film thickness under lubricant starvation conditions. High-speed flash photography facilitates the recording of the interference fringes as a function of a transient inlet boundary. By operating the bearing in a counterrotation manner a ball can be rotated in a fixed position and thus be observed continuously. Most of the results presented here were obtained with this counterrotation arrangement.

1.3 Nomenclature

a	radius of Hertzian contact zone for point contact
B	$\frac{3w}{2\pi a E' h_o}$, point contact parameter
b	one-half of Hertzian contact width for line contact
E'	reduced elastic modulus, $\frac{1}{E'} = \frac{1}{2}\left(\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right)$
E_1, E_2	elastic modulus of materials
F_c	centrifugal force
F_R, F_S	rolling and sliding friction
G	$\alpha E'$, materials parameter
H_m	$\frac{h_m}{R}$, film thickness parameter
h	thickness in z-direction
h_H	thickness of the gap outside the Hertzian region for dry contact
h_b	film thickness at the inlet lubricant boundary
h_m	minimum film thickness
h_o	central film thickness
$(h_o)_f$	film thickness under fully flooded conditions
h^*	h/h_o, dimensionless film thickness
$\bar{h}$	local film thickness where $dp/dx = 0$
I	monent of inertia
$I_{max,min}$	maximum and minimum intensity of fringes
$I_{1,2}$	intensity of rays 1 and 2
i	angle of incidence
K	fringe order
K_s	$h_o/(h_o)_f$
M_g	gyroscopic moment
m	mass
n	refractive index
n_o	refractive index at atmospheric pressure
P	p/E', dimensionless pressure
p	pressure
p_{max}	maximum Hertzian pressure
q	reduced pressure, $q = \frac{1}{\alpha}(1-e^{-\alpha p})$
q^*	$\frac{h_o^2 q}{12 u \eta_0 a}$, dimensionless reduced pressure
R	effective radius of curvature, $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$; also reflectivity

R_c	radius of pitch circle of cage
r	angle of refraction
r_f	radius of circular fringe
r^*	r/a, dimensionless polar coordinate
S	'inlet distance' measured between Hertzian radius and the inlet lubricant boundary
S_f	required inlet distance to obtain a flooded condition
s	additional coordinate in x-direction
U	$\frac{u\eta_0}{E'R}$ speed parameter
u	$\frac{1}{2}(u_1 + u_2)$; also fluid velocity relative to the contact region
u_r, u_s	rolling and sliding velocity
u_1, u_2	velocity of surface 1 and 2
V	$\frac{I_{max} - I_{min}}{I_{max} + I_{min}}$; fringe visibility
v	elastic displacement in z-direction
W	$\frac{w}{E'R}$, load parameter
w	applied load for point contact and load per unit width for line contact
w_c	cage loading due to centrifugal force
X	x/b, dimensionless x-coordinate
x,y,z	coordinates
x_b	'boundary distance' measured from the centre of the Hertzian region to the inlet lubricant boundary
α	pressure-viscosity coefficient
$\gamma_{1,2}$	degree of coherence between rays 1 and 2
δ	optical path difference between two rays
η	viscosity
η_0	viscosity at atmospheric pressure and test temperature
θ	limiting source or aperture size
λ	wavelength of light
μ	effective coefficient of friction
ρ	density
ρ_0	density at atmospheric pressure
σ_1, σ_2	Poisson's ratio of materials

ϕ	phase change on reflection; also, angular dimension
ψ	angular dimension
ω_b	angular velocity of ball
ω_c	angular velocity of cage

Chapter 2 MECHANICAL SYSTEM

2.1 Introduction

The experimental apparatus provides a system whereby the elastohydrodynamic film thickness in a ball thrust bearing can be measured using optical interferometry. The experimental bearing has two operational modes which provide different dynamic conditions of ball rotation.

The primary mechanical variables associated with the shape and thickness of the lubricant film are speed and load. The mechanical system must be designed so that these can be freely varied and controlled. To obtain a stable and steady film thickness the system must also be free of vibration.

2.2 Stroboscope Arrangement

The complete stroboscopic arrangement is shown schematically in fig.1. The ball thrust bearing is shown in fig.2. It consists of two flat races separated by three one-inch diameter steel balls. The lower race was made of steel and was lapped to a surface roughness of 2 micro-inches c.l.a. before running in. It is supported by an air bearing which is driven by an electric motor. The upper race is an optical crown glass disc, similar in design to that used by Foord(2). It is 6 inches in diameter and 1/2 inch thick. A partially reflecting layer of chromium is coated on its lower surface.

The three balls are equally spaced by a steel cage at a pitch diameter of 3.76 inches. The clearance between the balls and cage pockets is 0.008 inch, which

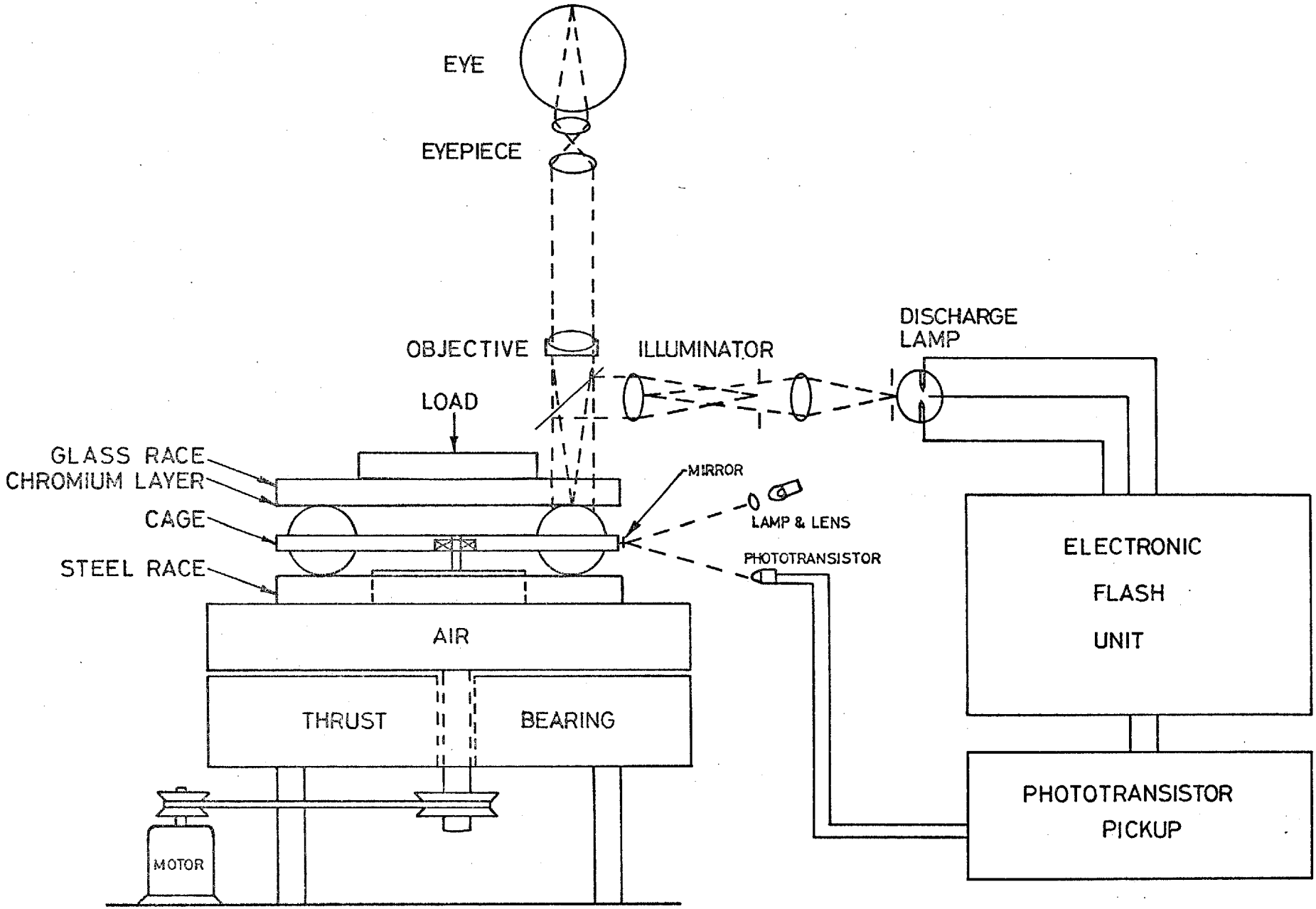


Fig.1 Schematic diagram of apparatus

Fig.2 Stroboscope arrangement

is similar to that found in commercial bearings using one-inch diameter balls. The cage is supported by a small ball bearing located in its centre and fixed to a shaft attached to the air bearing. This method fixes the axis of cage rotation and relieves the balls of some of the cage friction. Under high speeds, where the balls are centrifugally forced against the cage pockets, it was found that the steel cage generated scratches on the balls. The cage was later replaced by one made of a bonded cotton fabric material (Tufnol) which overcame this problem.

Load is applied to the top race by means of a lever arm which has a mechanical advantage of 2.75. A spring balance and turnbuckle attached to the free end of the arm provide a continuous range of loads. The load is distributed over the glass race by a steel loading block.

The bearing is operated by driving the lower race. The upper race is held stationary by a thin metal strip attached to the loading block. The cage, therefore, rotates (neglecting slip) at 1/2 the lower race speed. The interference fringes produced by a particular ball contact are observed stroboscopically.

The bearing is lubricated by a gravity feed system. A plastic tube feeds lubricant into the bearing from a lubricant bottle. The tube passes through a hole in the loading block and terminates at the lower surface of the block. The lubricant drips through a hole in the centre of the glass race and into a reservoir machined in the steel cage, (see fig.3). Three grooves, extending radially from the reservoir, channel lubricant to the leading edge of each ball. The lubricant adheres to the

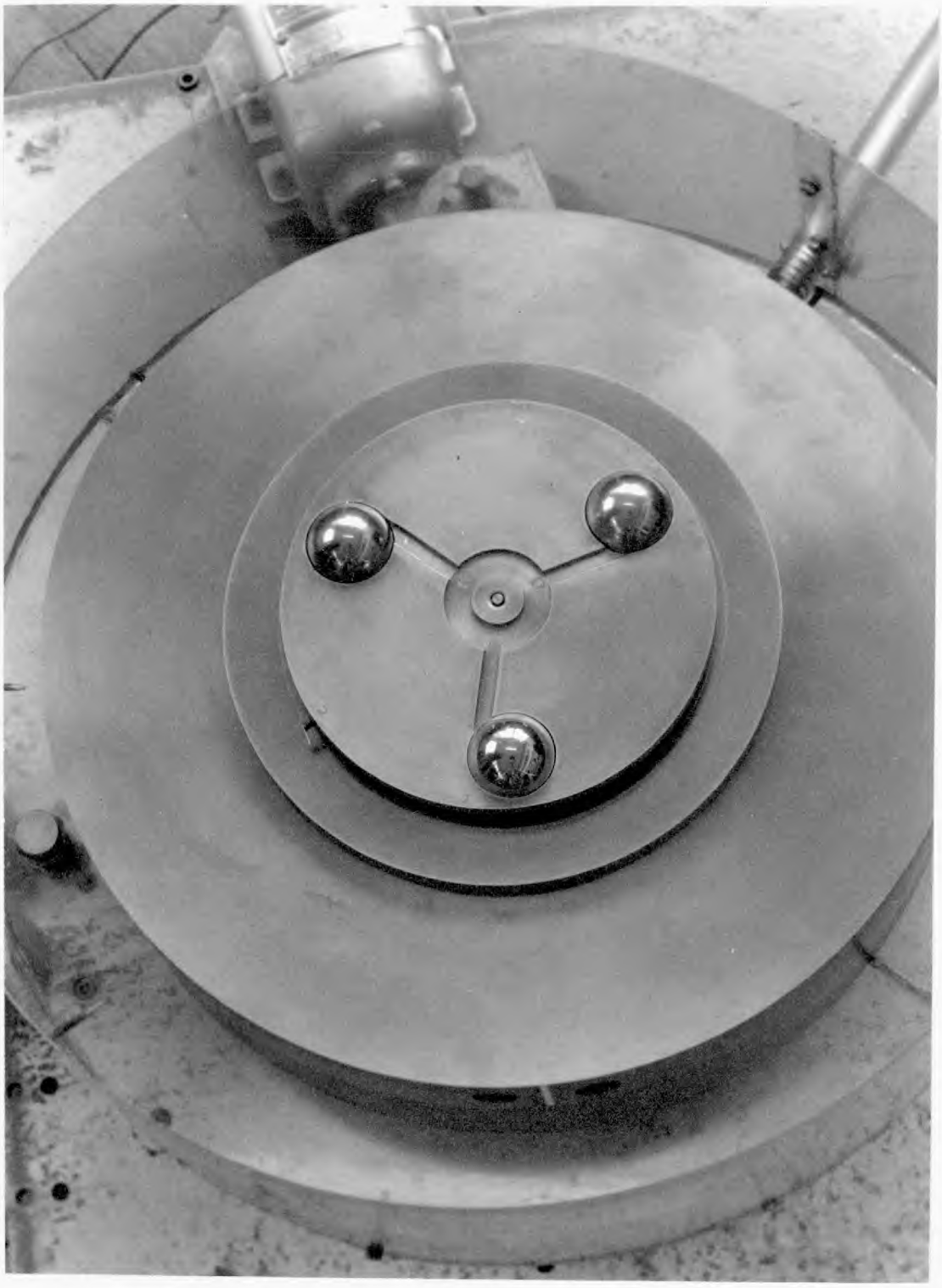


Fig.3 Cage for stroboscope arrangement

ball surface and is carried into the conjunction region by the rotation of the ball. Excess lubricant which is thrown off the bearing is collected by a plastic shield and trough attached to the air bearing.

2.3 Counterrotation Arrangement

The counterrotation arrangement is shown in fig.4. It is similar to the stroboscopic arrangement except for its mode of operation. In this case the cage is held stationary by a thin metal strip attached to it. Thus, when the lower race is driven in one direction, the upper race is rotated in the counter direction by traction transmitted through the three balls.

A small shaft attached to the air bearing fixes the axis of rotation of both the cage and the glass race. The glass race has a nylon insert in its centre, and the shaft rides against a ptfe bush which is pressed into the nylon insert. A loading block which is centred on the nylon insert rests on the glass race and rotates with it. Load from the lever arm is transmitted through a small ball thrust bearing located in the loading block.

The cage is shown in fig.5. It is made of Tufnol and supported in the centre by a small ball bearing. The cage has three pockets separated by 120^{o}. Two of the pockets are located at a pitch diameter of 5.2 inches. The third pocket, which contains the test ball, is adjustable in the radial direction and always located at a pitch diameter less than 5.2 inches. In this situation the load applied to the test ball is not the same as the other two. It is easily calculated by taking the sum of the moments about the bearing centre.

Fig.4 Counterrotation arrangement

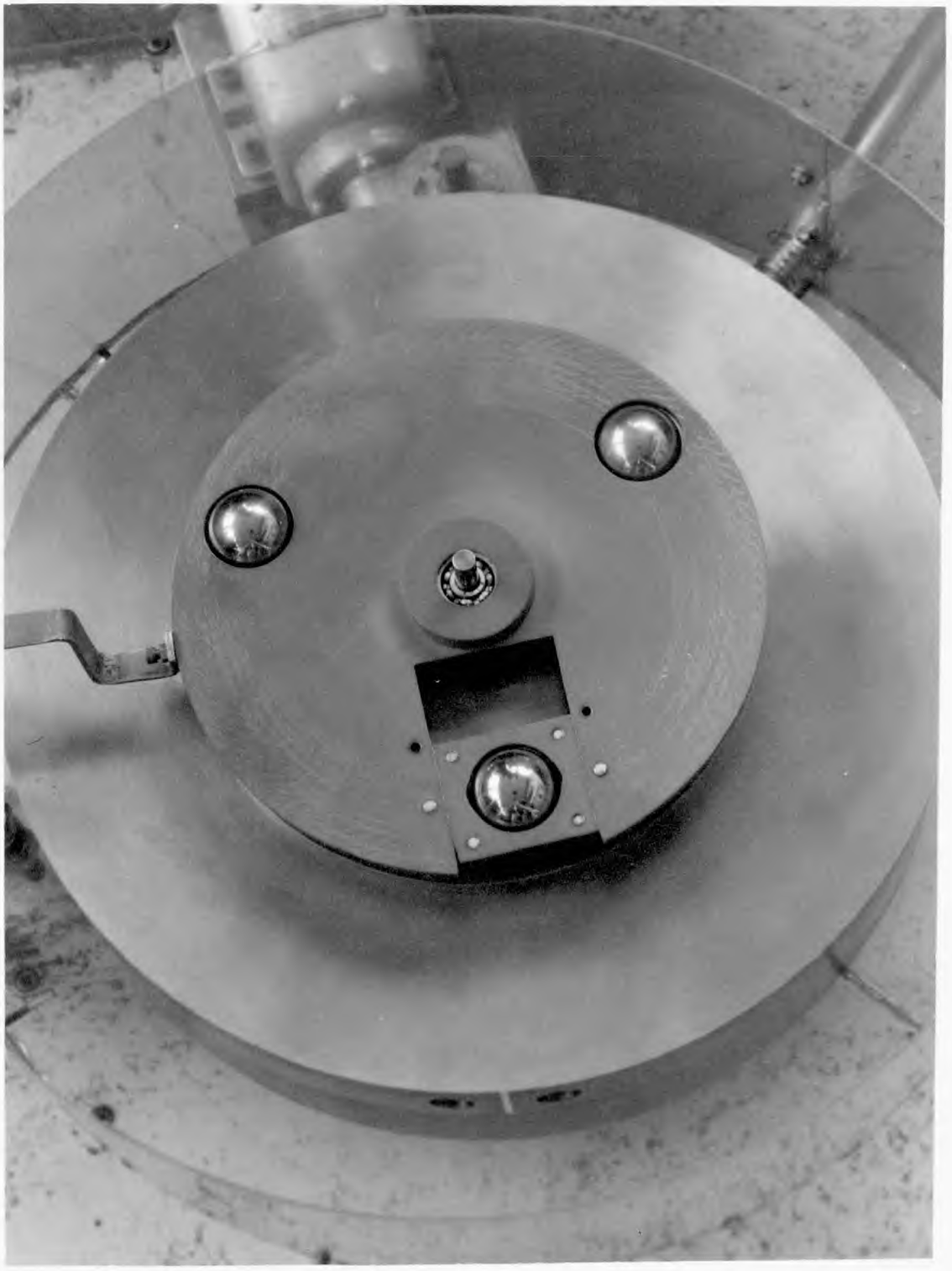


Fig.5 Cage for counterrotation arrangement

The adjustable hole is important for two reasons. First, for lubricant starvation tests, the distribution of lubricant on the bearing surfaces plays a vital role. By operating the test ball at a smaller pitch diameter, the lubricant tracks of the other two balls do not interfere. Any excess lubricant associated with these two balls is transported radially away from the test ball owing to the race rotation. Second, if the glass race or chrome layer become pitted or scratched, the test ball can easily be moved to another track.

The frictional torque associated with the ptfe bearing in the glass race and the ball thrust bearing in the loading block causes the balls to ride against the leading or trailing edges of the cage pockets. The leading and trailing edges of the adjustable cage pocket have been relieved as shown in fig.5. Thus, the test ball is controlled by the sides of the cage pocket. In the counterrotation mode the axis of ball rotation remains essentially constant. Hence, any viscous heating of the lubricant at the cage pocket is associated with the lubricant on the sides of the ball and not with the main load bearing region.

2.4 Supporting Air Bearing

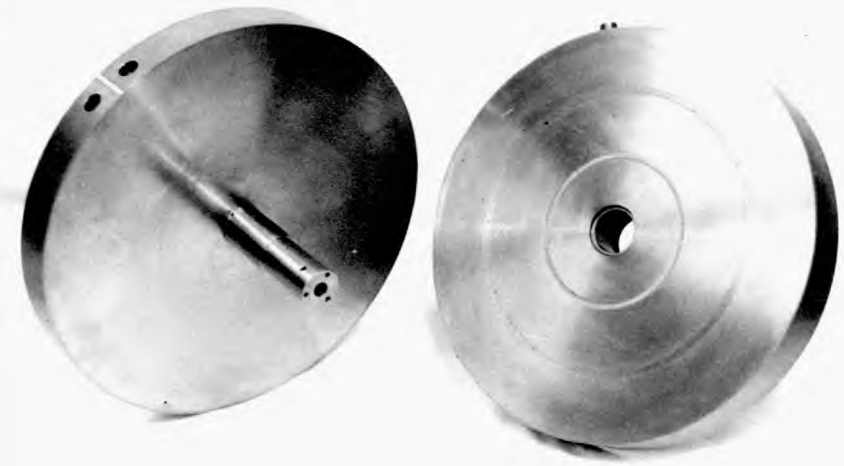
Under normal experimental loads the Hertzian diameter of contact between the ball and the glass race is on the order of 12×10^{-3} inches which requires a low-power microscope for its observation. The elastohydrodynamic films measured by interferometry are on the order of 1000 times smaller than this. To observe these films in a dynamically operated ball thrust bearing, the bearing must be operated

about a fixed axis with a minimum of undulation from the bearing races.

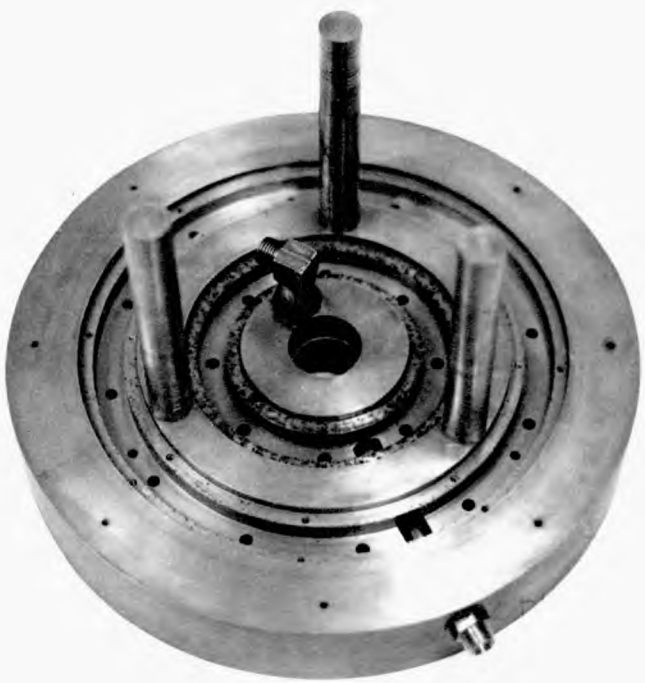
To satisfy these requirements an externally pressurized air bearing support was designed. The air bearing was designed for a thrust capacity of over 2000 lbf. so that it could accomodate the loads of a tapered roller thrust bearing as well as the ball thrust bearing. Simplicity and ease of construction were carefully considered in the design so that the bearing could easily be made on the machines available to the author.

The basic components of the bearing are shown in fig.6 and consist of two meehanite discs with a centrally located steel shaft. The discs which are 11 inches in diameter provide the required bearing area for the high thrust capacity. The disc faces were ground flat and parallel, and the rotating member was dynamically balanced to one gram-inch. The details of the air jets are shown in fig.7 and follow after the design criteria given by Grassam and Powell(3), Lock(4) and Constantenestu(5). The thrust bearing has two rings of eight jets. The jets are 0.020 inch in diameter and are connected by an annular compensating space 0.005 inch deep and 0.1 inch wide. Compressed air enters the side of the lower disc and feeds the jets from two annular channels shown in fig.6.

The journal bearing (see fig.7) is 1.125 inches in diameter and 1 inch long. Its load capacity is about 10 lbf. which is sufficient for the radial load requirement. A ptfe bush is inserted into the bearing housing, which provides added protection in case the bearing is overloaded. The journal bearing has a single row of six jets 0.032 inch



(a)



(b)

Fig.6 Basic components of air bearing: (a) upper and lower discs; (b) underside of lower disc showing annular feed channels

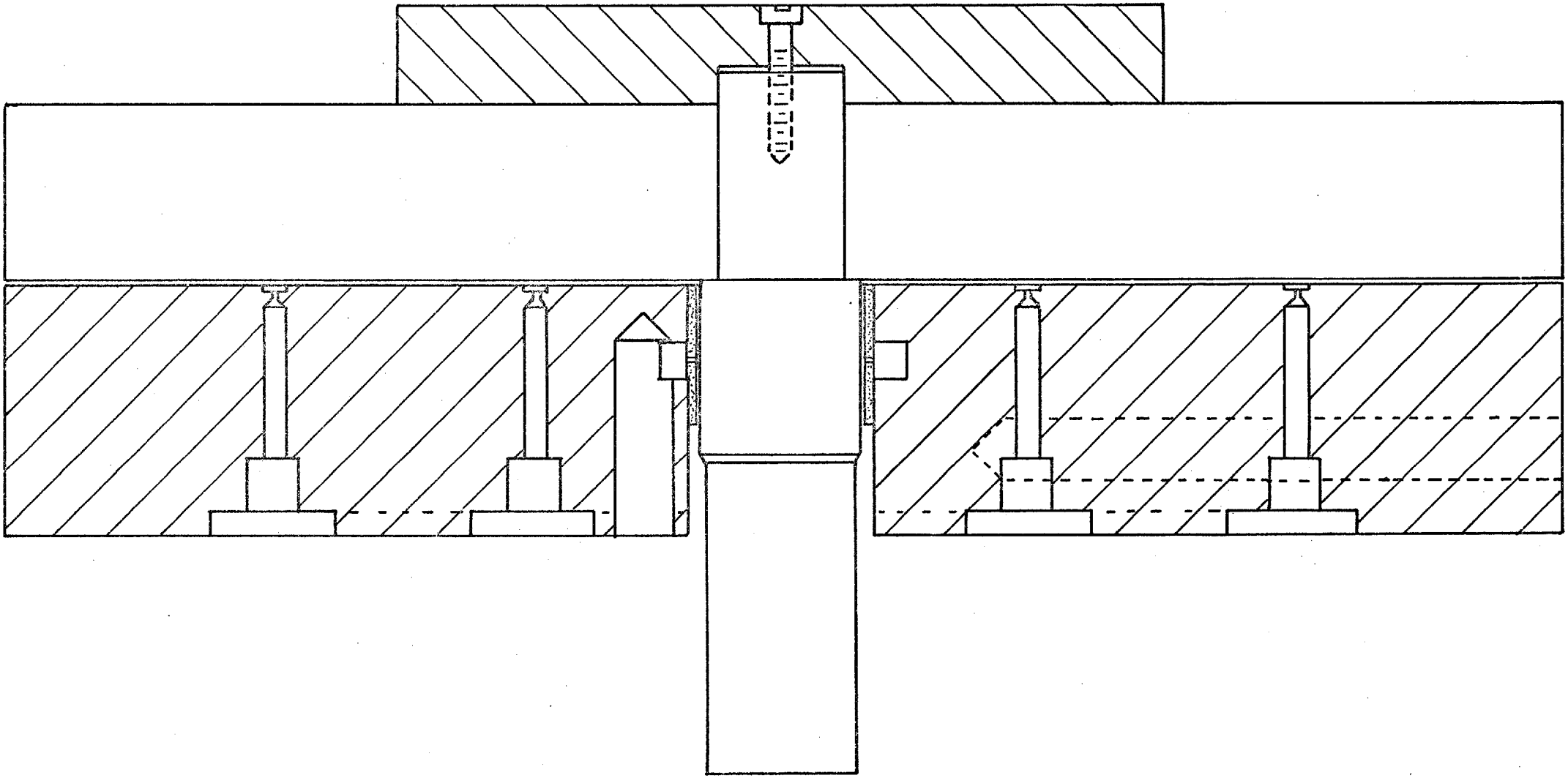


Fig.7 Supporting air bearing

in diameter. Compressed air enters the bottom of the lower disc and is distributed to the jets via an annular channel.

Under no load the air bearing can operate with only a few psi., gauge pressure. A maximum flow rate of 8 ft^3/ min. occurs at full line pressure of 80 psig. Full line pressure was necessary for the roller bearing tests where loads up to 1000 lbf. were used. A pressure of 20 psig. was used for the ball bearing tests where loads below 50 lbf. were encountered.

2.5 Drive System

The experimental bearing is driven by an electric motor and pulley system shown in fig.8. The air bearing shaft is connected to a 3 inch diameter pulley by a rubber coupling. The pulley rotates about a ball bearing located in its centre and is supported by a pedestal fixed to the base plate.

The electric motor is a fractional horsepower (1/12 hp at 5,500 rpm) universal series-wound type. The motor is connected to a regulator which supplies pulsed D.C. by a half-wave silicon controlled rectifier. The motor acts as its own tachometer and changes in torque are detected by a bridge circuit which alters the firing angle of the silicon controlled rectifier. The resulting speed-torque characteristics are most favourable at low speeds. The relatively constant load torque of the experimental bearing coupled with the large moment of inertia (15.7 lbm.-ft.-$sec.^2$) of the air-bearing aided this system in providing adequate speed regulation. The motor speed

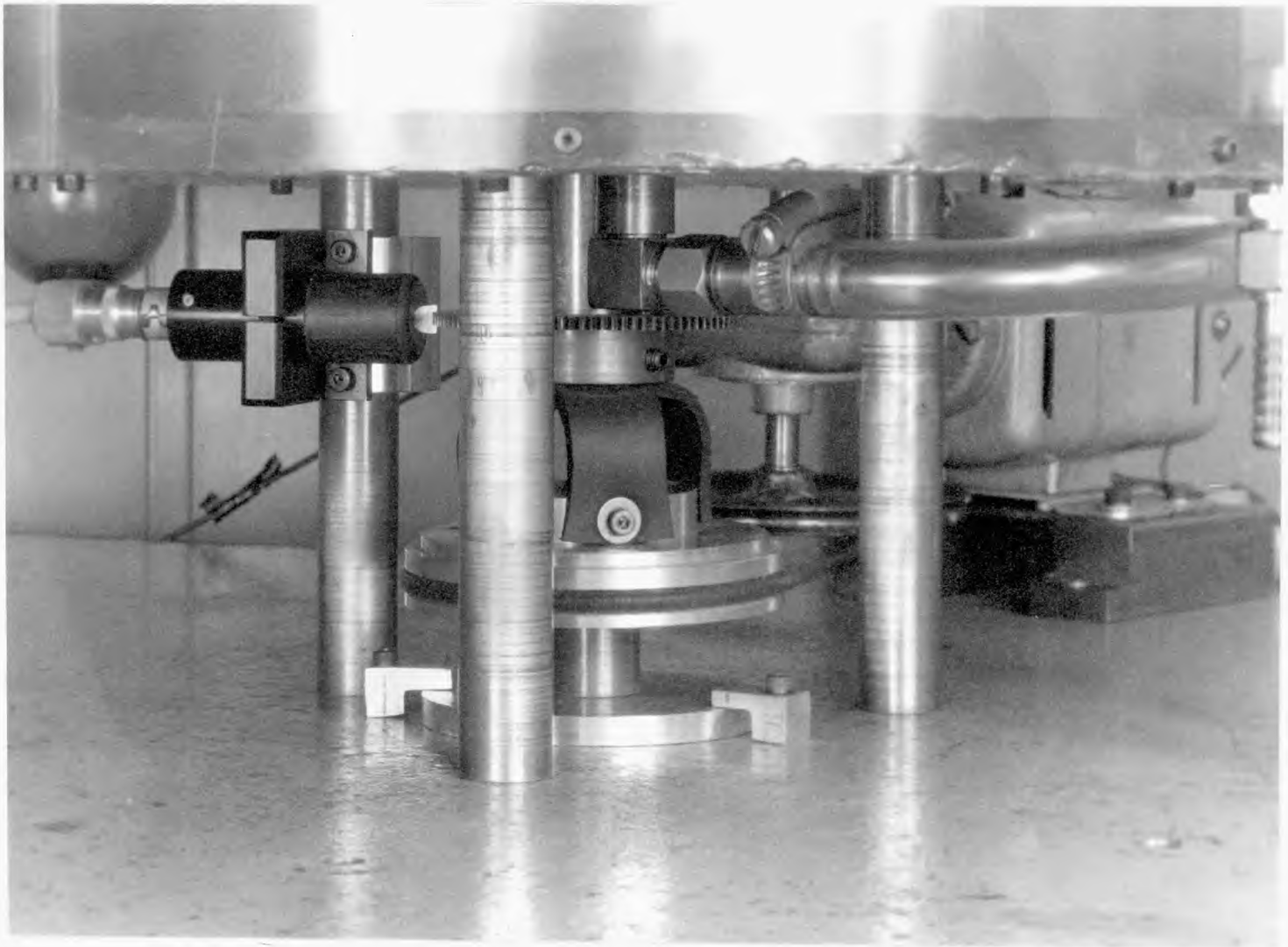


Fig.8 Drive system

is manually controlled by adjusting the input voltage to the regulator. This was accomplished by connecting a continuously-variable transformer between the mains and the regulator.

The speed range required was not very demanding. The lubricant starvation tests required a ball speed range of 80-250 rpm. The ball rotational speed is about 5 times the air bearing speed. Gross changes in the speed-viscosity parameter $(\frac{\alpha\eta_0 u}{R})$ were obtained by changing the lubricant viscosity as well as the bearing speed. The largest speed range was for some of the film thickness vs speed tests where ball speeds between 0 and 1000 rpm were required. The lower speed requirements were usually carried out with a 3.8 inch diameter pulley attached to a 20:1 reduction gear box, while a 2 inch diameter pulley directly attached to the motor shaft was used for the higher speed requirements.

2.6 Speed Measurement

The surface velocities of the test ball and the glass race were deduced from the speed measurements taken from the upper and/or lower races. These were obtained from signals generated by a magnetic transducer in proximity with a 96 tooth ferrous gear wheel. The lower race speed was obtained from a gear wheel attached to the air bearing shaft as shown in figure 8. This was the only speed recorded for the stroboscope arrangement. The upper race speed was obtained from a gear wheel attached to the loading block as shown in fig.4. The generated signals were recorded as pulses per minute on a five-digit impulse counter. They can be displayed individually or as a ratio.

By counting over a period large enough to obtain at least three digits, an accuracy of 1% can easily be obtained.

The above system was found to be adequate for air bearing speeds above 20 rpm. The pulses generated below this speed were too low for a reliable count. The speed was then obtained by timing a number of revolutions of the upper race with a stop watch. A narrow white line painted on the loading block was used as a counting reference. By counting over a period of at least 15 seconds a satisfactory speed measurement was obtained.

For the counterrotation arrangement the slip between the upper and lower races was measured by observing their speed ratio on the counter. Over the loads and speeds encountered in typical tests the slip was always found to be less than 2%. This slip is the summation of that occurring between the balls and both the upper and lower races. It is not known how the slip is divided between the ball/steel race interface and the ball/glass interface. The amount of slip is small enough so that only the measurement of one race speed was necessary for the determination of the contact velocities.

Chapter 3 THE OPTICAL INTERFEROMETRIC SYSTEM

3.1 Introduction

Interferometry has already provided significant contributions to lubrication research, and it will no doubt play an important role in the future. One of the major difficulties in its application is the determination of the relevant interferometric principles and how they can be used to satisfy the design requirements of a specific investigation. Herein lies the purpose of this chapter.

The fundamentals of interferometry are quite well established as a vast number of textbooks(6-10) and papers on the subject will attest. Its importance has certainly been due to its sensitivity and versatility as a tool for making physical measurements. It has not left the field of lubrication untouched. Indeed, interferometric measurements of surface roughness and bearing curvatures are very common.

In 1919 Hardy(11) used interference colours to measure lubricating films on glass. The extension of its use to the measurement of ehd lubricating films is a relatively new inovation. Initial work by Kirk(12) in 1962, Archard and Kirk(13) in 1963, and Blok and Koens(14) in 1965 certainly indicated the feasibility and usefulness of the method. Since then the use of interferometry has been mostly exploited under the direction of Dr. A. Cameron at Imperial College. The classic work was that of Cameron and Gohar(15) and later Foord, Hammann and Cameron(16).

The beauty of the application of interferometry to the measurement of ehd films lies in its compatibility as a measuring tool with that which is to be measured. The

wavelength of light (λ) is a convenient unit of measure for the ehd film thicknesses usually found in practice. Also, an interferograph shows a much greater sensitivity in the direction of depth in comparison with the direction of extent. So, too, is the ehd film thickness much smaller in comparison with the contact length.

3.2 Requirements of the Interferometric System

The primary application of interferometry in this report is the measurement of film thickness under lubricant starvation conditions. These film thickness measurements are obtained from high-speed photomicrographs which record the condition of lubrication at an instant in time. This particular application demands certain requirements of the interferometric system. Firstly, the fringe order must be determined directly from the photomicrographs. Secondly, the fringes should be closely spaced and form a continuous pattern which will enable the thickness at intermediate fringe positions to be estimated. Thirdly, to avoid damage under severely starved conditions, the reflecting surface of the transparent race must have sufficient durability.

The objective was to design an interferometric system which would fulfill these special requirements and at the same time satisfy the necessary conditions for obtaining clear and well-defined fringes. To obtain clear interference fringes certain interferometric principles must be considered. These are presented and discussed in conjunction with the choice of an interferometric system.

3.3 Basic Interferometric System

The basic interferometric system is shown in fig.9. The interference fringes formed by this system are based on the " division of amplitude" of the incident wave. The incident light wave from the source is divided by the partial reflection of the semireflecting layer on the bottom surface of the glass. The resulting wave fronts maintain their original width but have reduced amplitudes.

Consider two incident rays 1 and 2 coming from a point S on the source. For near normal incidence the two rays are very nearly parallel, and the angles of incidence (i) and refraction (r) are essentially the same for both rays. The refraction from glass to oil is neglected since the refractive index is nearly the same for both media. The first ray is transmitted through the oil film and reflected off the ball surface along the path SIJKL. The second ray is reflected off the semireflecting layer along the path SMKN. The two rays which meet at point K interfere. The resulting intensity at this point depends on the phase difference between the rays. If they are nearly in phase they "constructively" interfere giving rise to an intensity which is greater than that of the sum of the two rays acting separately. Conversely, if the rays are nearly out of phase, they "destructively" interfere, giving rise to an intensity which is less than the sum of the two rays acting separately.

The phase difference depends on the "optical" path difference between the two rays which for normal incidence is 2nh where n is the refractive index of the medium and h is the physical gap thickness. There is often a phase change in reflection at either one or both of the reflecting surfaces. A correction ϕ must then be added to the path difference. For

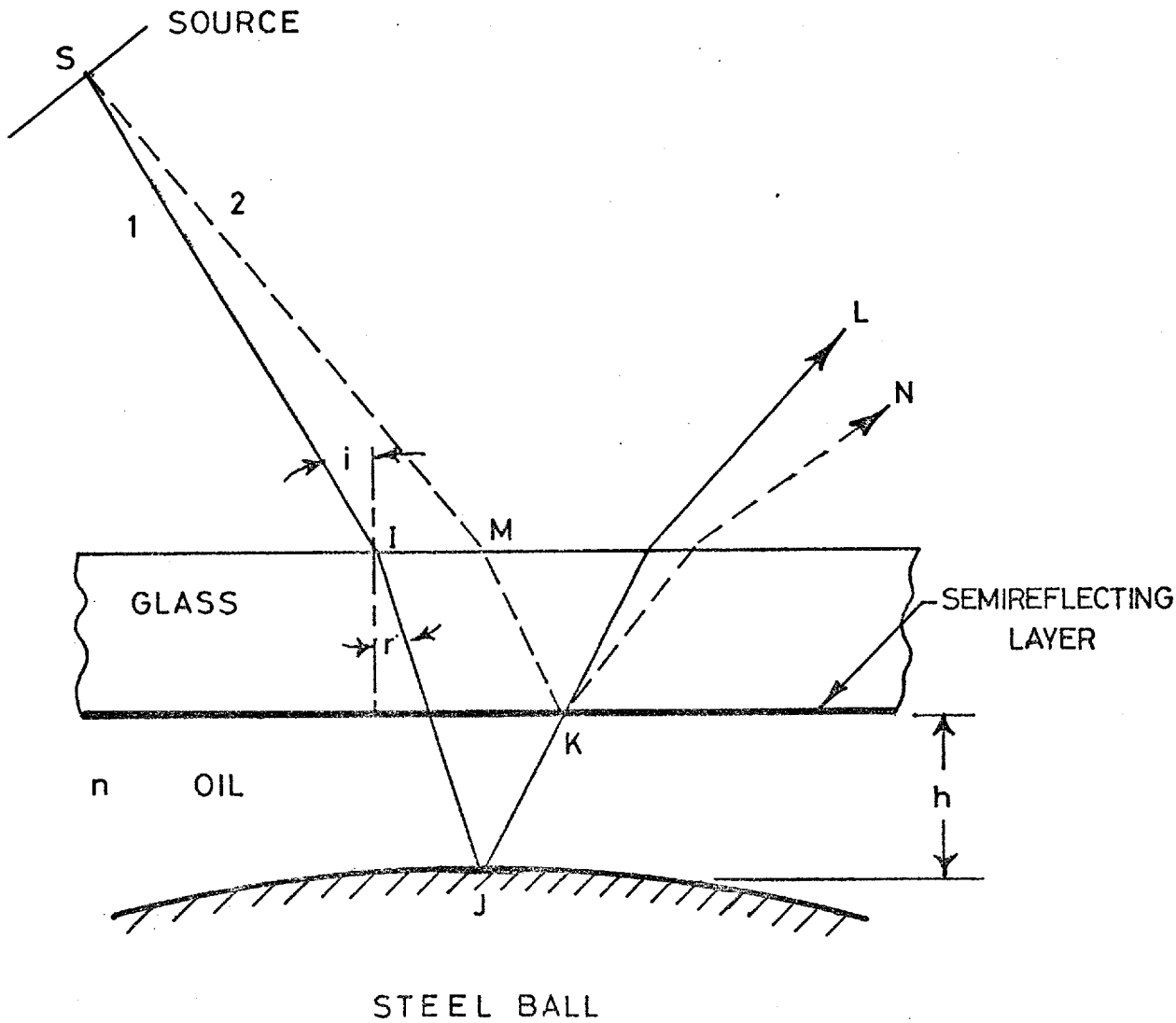


Fig.9 Basic interferometric system

oblique incidence the path difference (δ) can be shown to depend on the cos of the angle r. Thus, the path difference becomes

$$\delta = 2nh \cos r + \phi$$

Constructive interference occurs when the path difference is equal to an integral number of wavelengths, $K\lambda$. Therefore, when

$$K\lambda = 2nh \cos r + \phi$$

bright fringes are produced. K is called the "fringe order" and can take on integer values of 0, 1, 2,.... Destructive interference occurs when $\delta = (K + \frac{1}{2})\lambda$. Therefore, when

$$(K + \tfrac{1}{2})\lambda = 2nh \cos r + \phi$$

dark fringes are produced.

If λ , cos r, and ϕ are constant, the fringe order will depend on the optical path difference 2nh. These fringes are called "fringes of equal thickness" which according to the rays drawn in fig.9 are localized on the lower glass surface.

3.4 Fringe Visibility

a. Introduction

The clarity of fringes is usually described by their "visibility". Michelson(17) defined visibility as

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} and I_{min} are respectively the maximum and minimum intensities of the fringes. When $I_{min}=0$ the visibility equals a maximum of one. When $I_{max}=I_{min}$ the visibility is zero and no fringes are seen at all.

To obtain clear fringes (i.e. fringes of high

visibility) the interferometer must comply with certain interferometric criteria. These are now presented.

b. Intensity of interfering rays

If the two rays emitted by the source in fig.9 are "coherent" (i.e. in phase and of the same frequency), the resulting intensity at point K will depend on the individual intensities (I_1 and I_2) of the two rays and their path difference (δ), according to Fresnel's classical formula

$$I = I_1 + I_2 + 2\sqrt{I_1}\sqrt{I_2}\cos\frac{2\pi\delta}{\lambda}$$

Maximum intensity occurs when $\cos\frac{2\pi\delta}{\lambda} = 1$, and minimum intensity occurs when $\cos\frac{2\pi\delta}{\lambda} = -1$. Thus,

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1}\sqrt{I_2}$$

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1}\sqrt{I_2}$$

The visibility is

$$V = \frac{2\sqrt{I_1}\sqrt{I_2}}{I_1 + I_2} \qquad \ldots.(3.1)$$

which has a maximum of one when the intensities are equal ($I_1 = I_2$). For this case the intensity distribution follows a $\cos^2$ law according to $I = \cos^2\frac{\pi\delta}{\lambda}$. Equal intensities are achieved by controlling the reflectivity of the surfaces. This is dealt with in section 3.5.

c. Coherence and bandwidth

If visible interference fringes are to be obtained from the source of fig.9, the interfering rays must be mutually coherent. That is, at point K their phase difference must be constant. This condition is satisfied if the two rays emitted by the source are in phase, have the same frequency, and are continuous over the path difference. In thermal sources the

atoms emit waves that are of finite length and which can vary in frequency. The coherence of two beams which may interfere is linked with the length (coherence length) and consequently the duration (coherence time) of the emitted wavetrains. Very good monochromatic thermal sources have a coherence time on the order of 10^{-18} sec. Lasers can have much longer coherence times which may be on the order of 10^{-2} sec. If, for example, the wavetrains emitted along ray 1 of fig.9 are shorter than the path difference over which interference is to take place, then there will be no fixed phase relationship between the two rays at point K. The two rays at point L are then noncoherent and no visual interference will take place. The wavetrains emitted by a source vary in length; hence, the interfering rays may be "partially" coherent.

If $\gamma_{1,2}$ is defined as the degree of coherence, then Fresnel's formulae for I_{max} and I_{min} can be written as

$$I_{max,\ min} = I_1 + I_2 \pm 2\sqrt{I_1}\sqrt{I_2}\gamma_{1,2}$$

For mutually coherent beams $\gamma_{1,2}$ reaches a maximum of one, and for noncoherent beams $\gamma_{1,2}$ falls to a minimum of zero. The visibility of the fringes then becomes a function of the degree of coherence,

$$V = \frac{2\sqrt{I_1}\ \sqrt{I_2}}{I_1 + I_2}\gamma_{1,2}$$

which can only equal one if the two beams have equal intensities and are mutually coherent. As the path difference (δ) increases, the degree of coherence ($\gamma_{1,2}$), and hence the visibility, (V), decreases. Therefore, interference fringes can only be seen over a finite thickness.

A physical source is composed of many atoms each of which emits wavetrains which are not necessarily in phase with those of other atoms. Thus, the interfering rays can

only be coherent if they originate from the same atom in the source. Each atom produces its own visual interference effect and the resulting intensity at point K is the collective contribution of all the atoms in the source.

There is an alternative way of explaining the decrease in visibility with path difference which is compatible with the theory of partial coherence. The length of the wavetrains emitted by a source is related to the inverse of the bandwidth. All sources have a finite spectral bandwidth, i.e. the emitted light waves are composed of a mixture of wavelengths. In this respect no source is truely "monochromatic". If the source emits various wavelengths, the fringes formed by the shorter wavelengths will be closer together than those of longer wavelengths. At zero path difference the fringes formed by the various wavelengths coincide. As the path difference is increased the dark fringes of the shorter wavelengths begin to overlap the light fringes of the longer wavelengths (and vice versa). This confusion of fringe position causes a decrease in visibility. The rate at which the visibility decreases as the thickness h increases thus depends on the bandwidth of the source. Clear or visible fringes can only be obtained over large thicknesses if the bandwidth is sufficiently narrow. However, very wide bandwidths can be tolerated if the thicknesses to be measured are small. The realization of this is important in that it allows a greater choice of light sources and spectral distributions to be used.

It is possible to have a spectral distribution which consists of a group of shorter wavelengths and a group of longer wavelengths. At zero path difference the dark and

light fringes from both groups will coincide. At a particular thickness the dark fringes of one group overlap the light fringes of the other. As the thickness is increased further, the dark and light fringes of both groups will again coincide. The resulting visibility rises and falls in a periodic manner. The use of two distinct wave groups can be very useful for determining the fringe order. This is discussed in section 3.6.

d. Source or aperture size

If the reflecting surfaces of the interferometer are nearly parallel and illuminated normally with parallel rays, the thickness for a given fringe order is

$$h_N = \frac{K\lambda}{2n} - \phi$$

where h_N is the thickness for rays of normal incidence. The production of parallel rays implies a point source and a perfect lens. A very small source is often impractical because of intensity considerations. A source of some finite size must, therefore, be tolerated which leads to a range of angles of incidence. The thickness for a given fringe order is then given by

$$h_r = \frac{K\lambda}{2n \cos r} - \phi$$

where h_r is the thickness for rays of incidence at angle r (see fig.10). Thus, as some of the incident rays depart from normal, a given fringe order is shifted to a position of greater thickness. The change in thickness Δh which accompanies this shift is given by

$$\Delta h = h_N - h_r = \frac{K\lambda}{2n}\left(\frac{1}{\cos r} - 1\right) \qquad \ldots(3.2)$$

Because the finite size of a source produces a range of angles

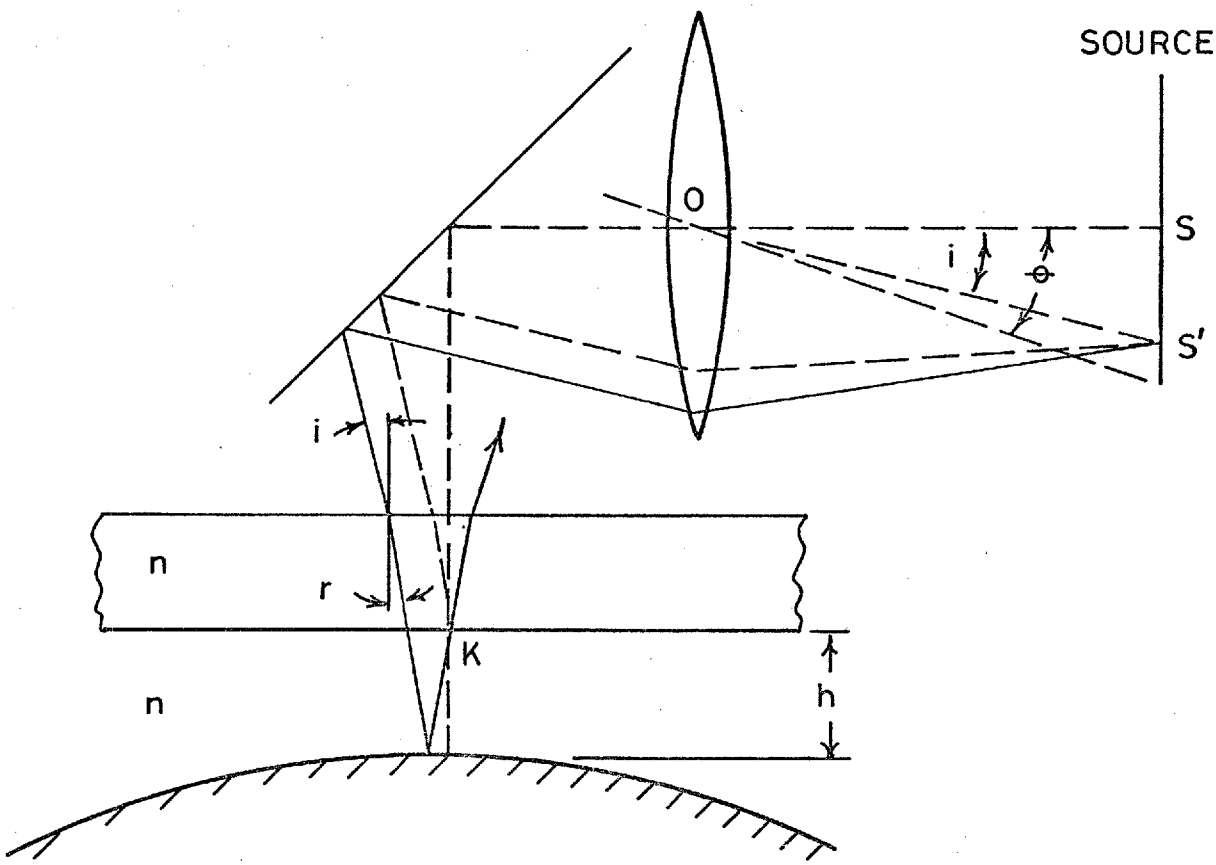


Fig.10 Illuminating system showing effect of source size on collimation

of incidence the resulting effect on the fringe position is an integrated one.

It will be shown later that the error introduced by Δh is small for the experimental conditions used here, and that the whole argument of fringe shift is irrelevant if the fringes are independently calibrated. Then a more important consequence of the source size is in connection with fringe visibility which is most clearly presented by Francon(9).

In fig.10 the interferometer is illuminated by a collimator which consists of an extended circular source placed at the focal plane of a lens. The interference at point K is first considered in connection with two points on the source, S and S'.

The path difference at point K between two rays from the point S of the source is

$$\delta = 2nh + \phi$$

The path difference at point K between two rays from point S' of the source is

$$\delta_{S'} = 2nh \cos r + \phi$$

For near normal incidence $\delta_{S'}$ can be written as

$$\delta_{S'} = 2nh(1-\frac{r^2}{2}) + \phi = \delta - nhr^2$$

Replacing the angle of refraction r with the angle of incidence i we get

$$\delta_{S'} = \delta - \frac{hi^2}{n}$$

If the two interfering rays are of equal intensity the resulting intensity at point K will, according to Fresnel's formula, be proportional to $1+\cos \frac{2\pi\delta_{S'}}{\lambda}$.

or $$I_K \propto 1 + \cos \frac{2\pi\delta_{S'}}{\lambda} = 1 + \cos \frac{2\pi}{\lambda}(\delta - \frac{hi^2}{n})$$

The intensity at K due to a small elementary ring seen from

point O under the angle di is

$$I_K = 2\pi \left[1 + \cos \frac{2\pi}{\lambda} (\delta - \frac{hi^2}{n}) \right] i \, di$$

If θ is the angle by which the entire source can be seen from point O, then the intensity at K due to the entire source is

$$I_K = 2\pi \int_0^{\theta} \left[1 + \cos \frac{2\pi}{\lambda} (\delta - \frac{hi^2}{n}) \right] i \, di$$

which gives

$$I_K = \pi\theta^2 + \frac{n\lambda}{h} \cos \frac{2\pi\delta}{\lambda} \sin \frac{\pi h\theta^2}{n\lambda}$$

The bright and dark fringes are given by

$$\cos \frac{2\pi\delta}{\lambda} = \pm 1$$

If $m = \frac{\pi h\theta^2}{n\lambda}$ the fringe visibility is

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{\sin m}{m}$$

The plot of V vs m is shown in fig.11. To obtain good fringe visibility the value of m must be small. For a fixed medium (n) and wavelength (λ) this either requires a small thickness (h) or a small source size (θ). If a minimun fringe contrast of 0.98 is chosen, then the value of m must not exceed 0.33. The limiting source size is, therefore,

$$\theta = 0.324 \left(\frac{n\lambda}{h}\right)^{\frac{1}{2}} \qquad \ldots(3.3)$$

This equation is used in section 3.8d in connection with the illumination system. It is significant that the limiting source size is proportional to $(\frac{1}{h})^{\frac{1}{2}}$. For very thin films a reasonably large source can be tolerated.

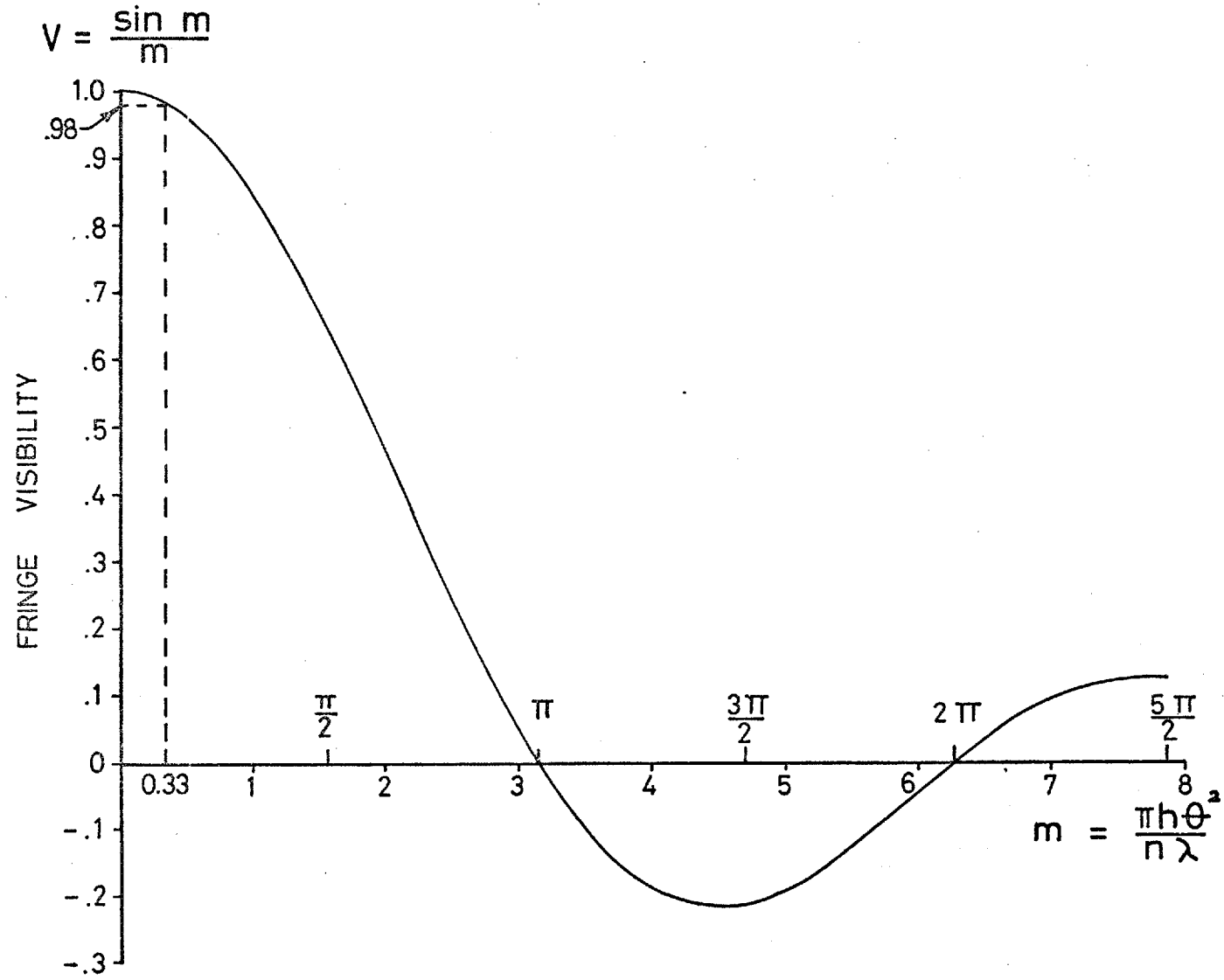


Fig.11 Fringe contrast as a function of m

3.5 Reflecting Surfaces

a. Introduction

The quality of the fringes and the accuracy of the measurements taken with the interferometer depend primarily on three conditions of the reflecting surfaces.

The first is surface finish. The intensity of a fringe will change from a maximum to a minimum (or vice versa) if the thickness changes by $\frac{\lambda}{4n}$ (approximately 3.5 micro-inches). It is, therefore, important that the surfaces conform locally to a mean elevation which is less than $\frac{\lambda}{4n}$. Any deviation from the mean leads to a confusion of fringe order and hence a decrease in fringe visibility. It has been found from practice that the reduction in visibility due to surface roughness can be made small or negligible if the surfaces have a roughness value which is less than one micro-inch c.l.a.

The second condition concerns that of geometry. If meaningful measurements of elastic deformation are to be obtained, it is essential that the deviation from a perfect surface geometry is small in comparison with the deformation due to pressure. The manufacture of rolling elements with both good surface finish and geometrical shape is not an easy matter. These two requirements for some geometrical shapes are not always compatible. It is primarily for this reason that the experimentation presented here is limited to point contact. Some initial work on line contact is presented in section 7.3.

The third condition is that of reflectivity. For two-beam interferometry, fringes of high visibility are obtained if the intensities of the two beams are equal (i.e. $I_1 = I_2$).

Equal intensities can be achieved by controlling the reflectivity of the surfaces. It is interesting that high reflectivity is not a necessary requirement for two-beam interferometry. It is possible to obtain fringes of good visibility with very low reflectivities (and hence high absorption and/or transmission) provided that $I_1 = I_2$. This is not, however, a very efficient use of the total light input. For multiple-beam interferometry, which is discussed in section 3.7, high reflectivities and low absorption are essential for sharpened fringes of good visibility.

b. Steel ball

The steel balls used in these experiments were specially made by Hoffmann Bearing Co. They are 1 inch in diameter, and they have a steel specification of EN 31 (1% C, 1.4% Cr). The spherocity is within 4 micro-inches and the surface finish is much better than 1 micro-inch c.l.a. This excellent surface finish was obtained by final lapping in a cast iron race. Surface defects are almost undetectable under the magnifications used here. The excellent finish has indeed contributed a great deal to the quality of the fringes.

The reflectivity of the steel balls was measured with a photocell and found to be 78% in air and about 58% in oil. If it is necessary, an increase in reflectivity of about 10 or 15% can be obtained by coating the ball with aluminium by thermal evaporation. However, the adhesion between aluminium and steel does not provide a very durable surface which may be necessary under severe lubrication conditions.

c. Glass race

The upper race of the thrust bearing is made of an

optical crown glass disc. It adequately satisfies the above requirements concerning surface finish and geometry.

In connection with the geometry of the mating surfaces (i.e. the ball and glass race), it is fortunate that a slight out-of-roundness of the ball and unevenness of the glass race can be made to produce a negligible effect on the film thickness between the surfaces of the interferometer. Film thickness is determined solely by the generated ehd film which will remain constant so long as the ehd conditions remain constant. The only significant condition that could be altered by geometrical irregulatities is that of load, particularly if the bearing is a statically indeterminate system. This is easily overcome by using a thrust bearing composed of only three balls and loaded centrally with a point load. In this way all redundant members are eliminated since each supporting element is completely necessary for equilibrium. Steady load and film thickness conditions are, therefore, maintained even though the bearing surfaces as a whole may fluctuate in the vertical direction by more than the film thickness itself. When observing the fringes through the microscope these fluctuations are not noticed since their magnitude is much smaller than the depth of focus.

To obtain fringes of high visibility it is most important that the reflectivity of the transparent race is compatible with that of the steel ball in making the intensities of the two interfering rays equal ($I_1 = I_2$). Gohar and Cameron (15), (18) achieved various reflectivities by using transparent materials which have refractive indices different from that of oil. For normal incidence the reflectivity, R, between two

dielectric materials (e.g. glass and oil) is

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \qquad \ldots(3.4)$$

where n_1 and n_2 are the refractive indices of the two media. Gohar and Cameron used three different materials: high refractive index glass (n=1.93), white synthetic sapphire (n = 1.77), and diamond (n = 2.20). The resulting reflectivities, using the refractive index of oil under pressure as 1.55, were respectively: 1.2%, 0.437%, and 3%. Since the reflectivity of the steel ball is 58% in oil, it is possible by using equation (3.1) to calculate the corresponding visibilities. These are respectively: 0.28, 0.17, and 0.44. The higher reflecting material of diamond (R = 3%) gives the best fringes, but with only a moderate visibility (V = 0.44). The reflectivity of the transparent materials were much too low in comparison with the reflectivity of the steel ball so that the establishment of equal intensities between the interfering rays was not achieved.

Higher reflectivities can be obtained by coating the transparent material with thin metal or dielectric films. Metal films such as silver and aluminium are very commonly used in interferometry(6); however, they do not provide a very durable surface for these lubrication experiments. Westlake (19 - 22) has used some dielectric films which have good durability. These are discussed in section 3.7c in connection with multiple-beam interferometry.

A metal reflecting layer of chromium was used for the experiments presented in this report. This material was used previously by Foord(2, 16). It has some peculiarities in connection with the intensity distribution and the phase

change on reflection. However, a number of other characteristics make it a very suitable choice for the particular application used here. First, a very durable coating can be obtained if the chromium is evaporated on a hot (140^{o} C) glass substrate. Second, the chromium layer establishes two-beam interference fringes and can be used with chromatic sources. This enables a continuous pattern of closely spaced fringes to be obtained from which the film thickness at imtermediate fringe positions can be estimated. Third, the reflectivity can be specified to provide interfering rays (I_1 and I_2) of equal intensity. This establishes fringes of high visibility. Fourth, chromium is chemically similar to commonly used bearing materials.

The intensity of the two interfering rays, which depends on the reflectivity and absorption of the reflecting surfaces, is shown schematically in fig.12. The values of reflectivity and absorption were obtained from measurements taken with the experimental light source and a photocell. The reflectivity of the steel ball was found to be 78% in air and about 58% in oil. The glass race, which was one left over from the experiments of Foord(2), was specified to have a chromium layer of 20% reflectivity. According to the manufacturer(23), this reflectivity was measured in air from the air/chromium side. The required reflectivity, however, is from the glass/chromium side. This was measured and found to be 14% in air. When oil is placed on the back surface, the reflectivity increases slightly to about 15%. This increase in reflectivity, though unexpected, is also predicted by the electromagnetic theory for reflection from an absorbing material (24).

Knowing the reflectivity of the chromium layer

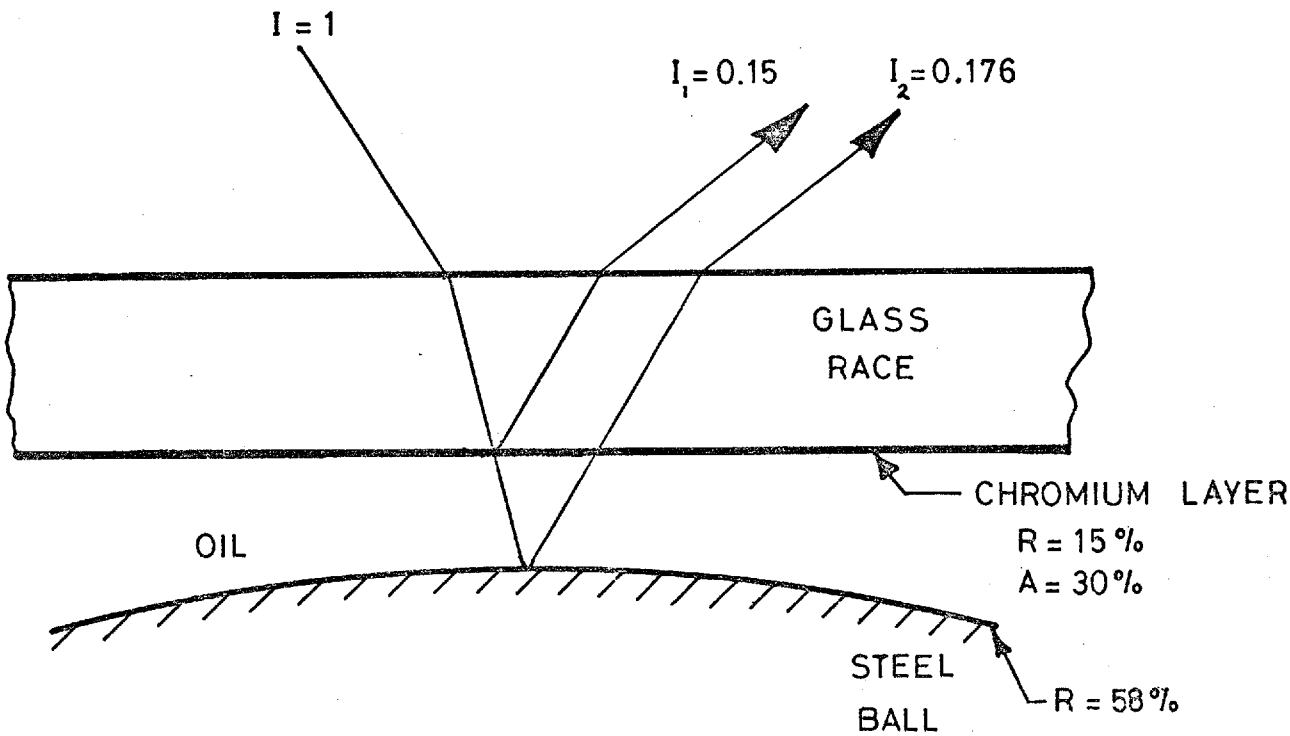


Fig.12 Reflectivity of surfaces and resulting intensities of interfering rays

it is possible to obtain its approximate thickness from the electromagnetic theory. This was found to be on the order of 170 Å which is much smaller than the usual ehd deformations. Therefore, its influence on the elastic properties of the glass substrate can be neglected. Further support of this conclusion is given by Westlake(25).

The absorption of the chromium layer can be deduced from the measurements of transmission and reflectivity of the glass race. The free electrons in the metal film cause a rather high absorption of about 30%. The high absorption is not detrimental to the fringe visibility for two-beam interferometry, provided that $I_1 = I_2$, but it does reduce the utilization of the total light input.

The intensities of the two interfering rays shown in fig.12 are nearly equal. They give a very good visibility of 0.996.

d. Non-reflecting films

The reflection from an air/glass interface as calculated from equation (3.4) is about 4%. If the upper surface of the glass race and the components of the microscope were not coated with non-reflecting films, a considerable amount of light would be lost by reflection. Moreover, the reflected light which may reach the image plane dilutes the image and therefore reduces the fringe visibility.

The upper surface of the glass race is coated with magnesium fluoride (n = 1.38) at a thickness of $\frac{\lambda}{4n}$ where λ corresponds to a wavelength approximately in the middle of the visible spectrum (usually green light, λ = 5500 Å).

Incident light reflects off the air/MgF_2 interface and the MgF_2/glass interface. Both reflections are from a "rare-to-dense" medium since the index of refraction of MgF_2 (n = 1.38) is less than glass (n = 1.52). This causes a phase change of π (or 180°) to occur for both reflecting interfaces. The rays reflecting from the MgF_2/glass interface travel through the film twice, thus making the path difference between the two reflecting interfaces equal to one-half wavelength which is the condition for destructive interference. The non-reflecting film does not destroy the light but redistributes it. Therefore, the decrease in reflectivity accompanies a corresponding increase in transmission.

The index of refraction of MgF_2 and glass are not exactly correct to establish interfering rays of equal intensity; therefore, the reflection is not totally suppressed. Also, the film is less operative for those wavelengths which are not near the middle of the visible spectrum. The film is, however, effective in reducing the reflection to about 1%. The optical components of the microscope are coated with similar non-reflecting films. This is normally done in microscope equipment.

e. Phase effects on reflection

It was stated in the previous section that reflection from a "rare" to "dense" medium causes a phase change of π, and that no phase change accompanies a reflection from a "dense" to "rare" medium. This appears to be the rule for dielectric reflections, but for metalic reflections the phase change is more complex. There are two metalic reflections in the interferometer. One from the opaque surface of the steel ball and the other from the thin film of chromium.

A simple experiment shows the approximate magnitude

of the phase change associated with the steel ball. A glass slide was placed on a glass lens in a medium of air. The "Newton's rings" in reflection were observed with a microscope. Since both media are dielectric, no phase change will occur at the glass slide/air interface; but a phase change of π will occur at the air/lens interface. Hence, destructive interference will occur at zero thickness. The central fringe corresponding to the contact area is, therefore, dark. If the lens is replaced by a steel ball the central fringe is again dark. Thus, the phase change of the air/steel interface must also be π . It is possible that the phase change is not exactly π as the eye is not too sensitive to slight intensity changes. Indeed, Barakat and Mokhtar(26) have found, for example, that opaque interfaces of glass/copper and glass/indium have phase changes which are slightly different from π .

A calculation of the phase change has been made using an equation based on the electromagnetic theory(24); however, the results are unreliable due to the uncertainty of the material constants of refractive index and absorption coefficient, which may vary with chemical and mechanical treatment.

The phase change associated with the chromium film is also uncertain. It is neither 0 nor π. The central fringe in figs. 14a, c and e are therefore neither light nor dark but are somewhere in between (light grey).

Also associated with the thin chromium film is an unsymmetrical intensity distribution. This is shown in fig.13 which is a microphotometer plot of the intensity distribution of the 5th light fringe of fig.14a. It shows that in the direction of increasing fringe order the intensity rises to

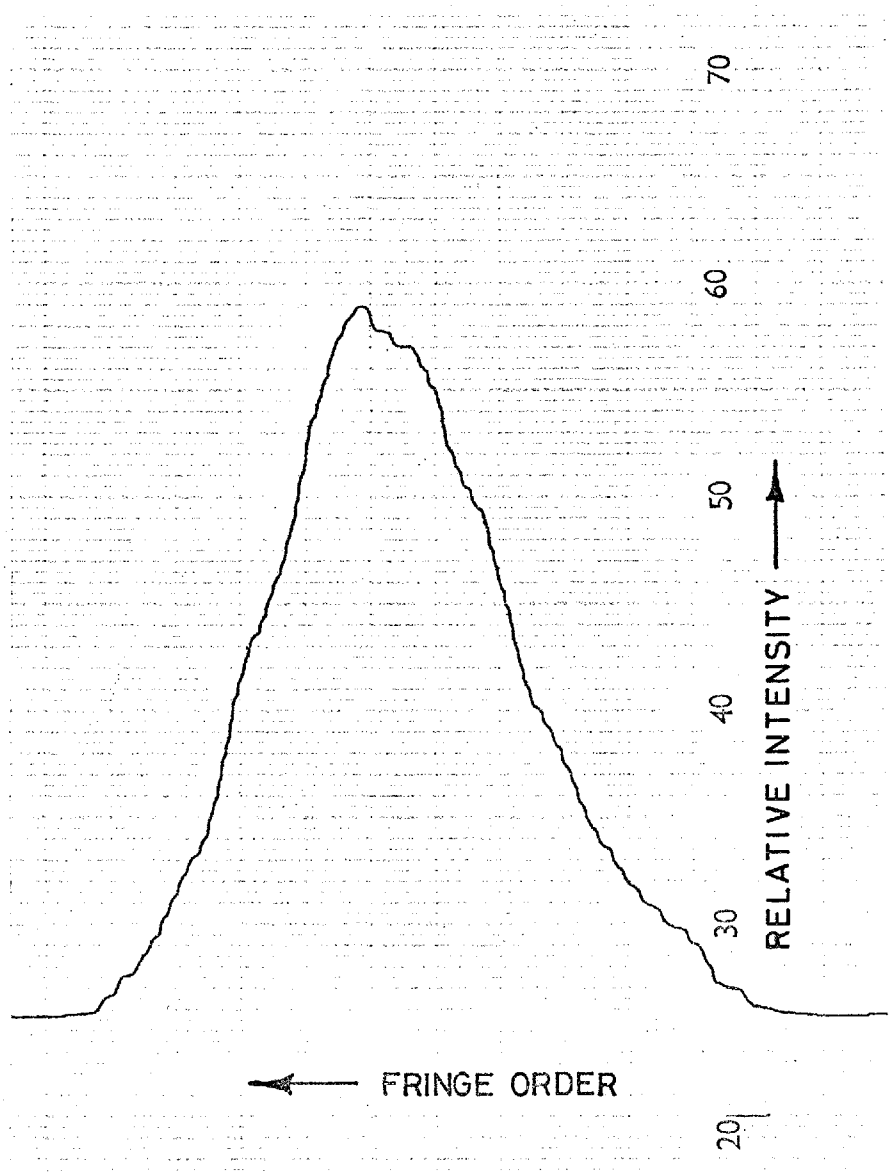


Fig.13 Intensity distribution of 5th green fringe of fig.14a

a maximum at a slower rate than it falls to a minimum. This is slightly exaggerated in the figure because the fringes were obtained from a curved surface. The unsymmetrical intensity distribution is due to the fact that the phase change of the chromium film does not effect the reflected and transmitted rays equally. According to Tolansky(27) the unsymmetrical distribution occurs with thin metal films which are less than 500 $\mathring{A}$ thick, and the position of maximum intensity varies with the thickness of the metal film. A detailed explanation of this effect is given by Holden(28).

The effects of phase change alter the position of the fringes. This depends not on the absolute phase change of each reflecting surface, but on the relative phase change between both reflecting surfaces. By suitable calibration the effects of phase change on fringe position can be fully accounted for. The calibrated fringes can then be used to obtain the numerical value of the phase change. This is discussed in section 3.9c.

3.6 Duochromatic System

a. Introduction

Gohar(1) and Foord(2) obtained interference fringes with both monochromatic and white light sources. Sodium and mercury discharge lamps were used for the former, and tungsten filament and quartz-iodine lamps were used for the latter. The relatively narrow bandwidth of a monochromatic source forms equally spaced dark and light fringes. The spacing between corresponding dark and light fringes is a function of the wavelength and is equal to $\frac{\lambda}{4n}$. Coloured fringes are formed

by the continuous spectrum of white light. They are the result of the summation of the individual fringes produced by each component wavelength of the visible spectrum. That is, each wavelength forms a set of individual fringes. The fringes of the shorter wavelengths are more closely spaced than the longer wavelengths. These fringes overlap one another. The resulting colour at each point is determined by the intensity of the various wavelengths present at that point. In this way a definite sequence of closely spaced coloured fringes are formed. There are certain advantages and limitations inherent in both monochromatic and white light fringes. These are discussed in the light of four considerations.

The first is fringe visibility. The thickness over which good fringe visibility is maintained is a function of the bandwidth as discussed in section 3.4c. The relatively narrow bandwidth of a monochromatic source generally provides ample visibility over the usual ehd film thickness range. This, however, is not true for the continuous spectrum of white light. As the film thickness increases the interference colours become increasingly more impure or unsaturated. In particular, after about 8 or 10 fringes so many colours are present at one point that the resultant hue is practically white, even though interference is still taking place. It is for this reason that clear fringes can only be obtained for oil film thicknesses up to about 40 micro-inches.

The second consideration is that of fringe spacing. The closely spaced fringes of white light provide good detail in film shape as shown by Foord et al(16). The fringe spacing of $\frac{\lambda}{4n}$ for a monochromatic source, however, is sometimes a relatively coarse unit of measure, particularly for the longer wavelengths.

The third consideration is that of fringe sensitivity. In a monochromatic system the fringes are determined by the variation of intensity. In a white light system, however, the fringes are also determined by the variation of colour. Since the eye is more sensitive to changes in colour than changes in intensity, it is possible to detect smaller variations in thickness with the white light system.

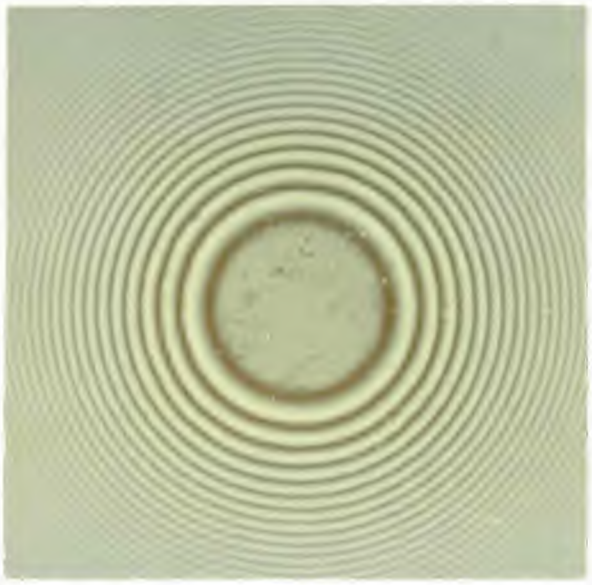
The last consideration is that of fringe order. A monochromatic source forms a sequence of identical dark and light fringes. If the wavelength is known, the fringe spacing is automatically determined, but it is necessary to count fringes to determine the fringe order. The coloured fringes of a white light source are not equally spaced. Their position must be determined by calibration. However, the fixed sequence of coloured fringes provide a convenient method for determining the fringe order. It should also be noted that the sequence of coloured fringes not only provides the fringe order but also the direction in which the fringe order is changing. This is often misleading in a monochromatic system.

b. The duochromatic green-red system

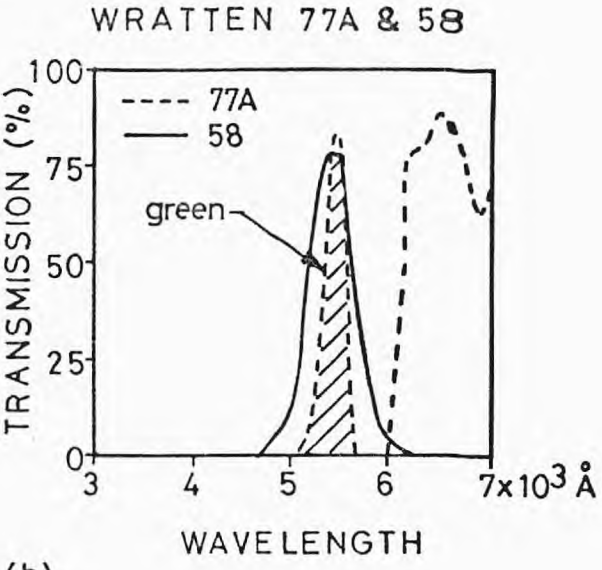
In light of the above discussion most of the advantages appear to lie in the white light system. The main failing of this system, however, is the limited film thickness range over which fringes of good visibility are observed (40 micro-inches). This problem is overcome by making a compromise between the monochromatic and white light systems. By limiting the spectral distribution to two discrete wavebands, it is possible to obtain good fringe visibility over a satisfactory film thickness range while still maintaining most of the conveniences of the white light system.

Two discrete wavebands, one green and one red, were obtained by using a Wratten 77A filter in conjunction with a white light source (xenon discharge lamp). This filter was chosen, firstly, because green and red fringes provide good colour contrast; and secondly, it was possible to obtain two more filters which would pass either the green or red bands individually when superimposed. The Newton's rings produced by the individual green and red bands and their combination are shown in fig.14a, c and e. These fringes were obtained by using the corresponding filters represented in figs.14b, d and f. The geometry is that of a steel ball loaded (0.5 lbf.) against the glass race in a medium of oil (n = 1.4809).

The origin of the coloured fringes produced by the duochromatic system can be obtained by considering the summation of the individual monochromatic fringes. This is shown in fig.15 which was obtained from the calibration data to be described in section 3.9b. The first two columns of the figure show the thickness (in air) of the dark and light fringes of the individual green and red bands. The green fringes are more closely spaced than red fringes because of the shorter wavelength of green light. When both wavebands are used simultaneously the chromatic fringes, as shown in the third column, are obtained. The colour of the chromatic fringes is determined by the intensity of the green and red wavebands at each point. At zero thickness both green and red fringes are in phase relative to each other. At about 45 micro-inches the green band has gained a half wavelength on the red. Hence, a light fringe of one falls on a dark fringe of the other. The corresponding chromatic fringes are, therefore, green and red. At about 90 micro-inches the green band has gained a whole wavelength on the red. Thus,the light and dark

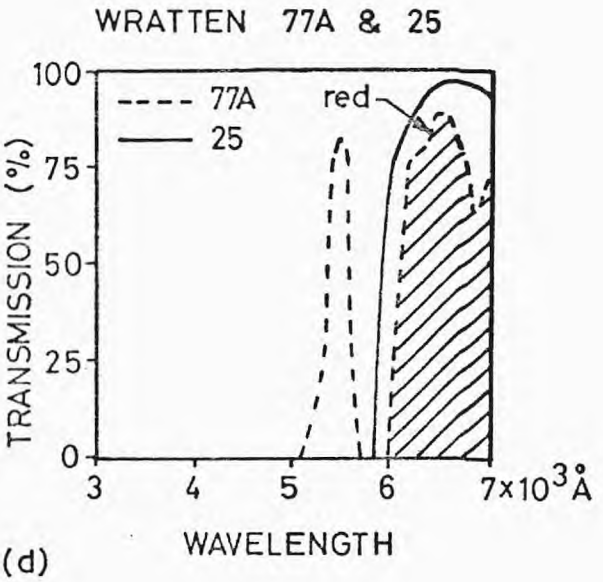


(a)

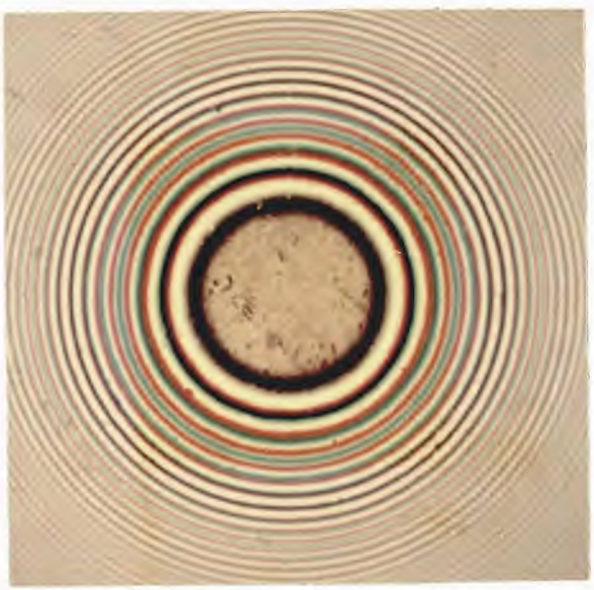


(b)

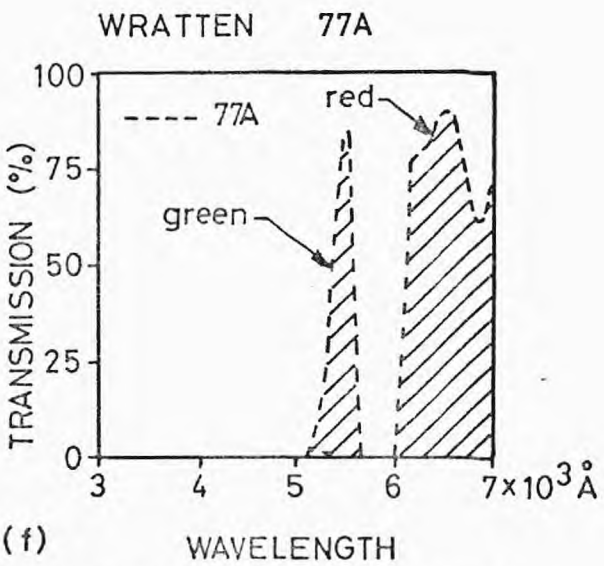
(c)



(d)



(e)



(f)

Fig.14 Monochromatic and duochromatic fringes corresponding to various filters

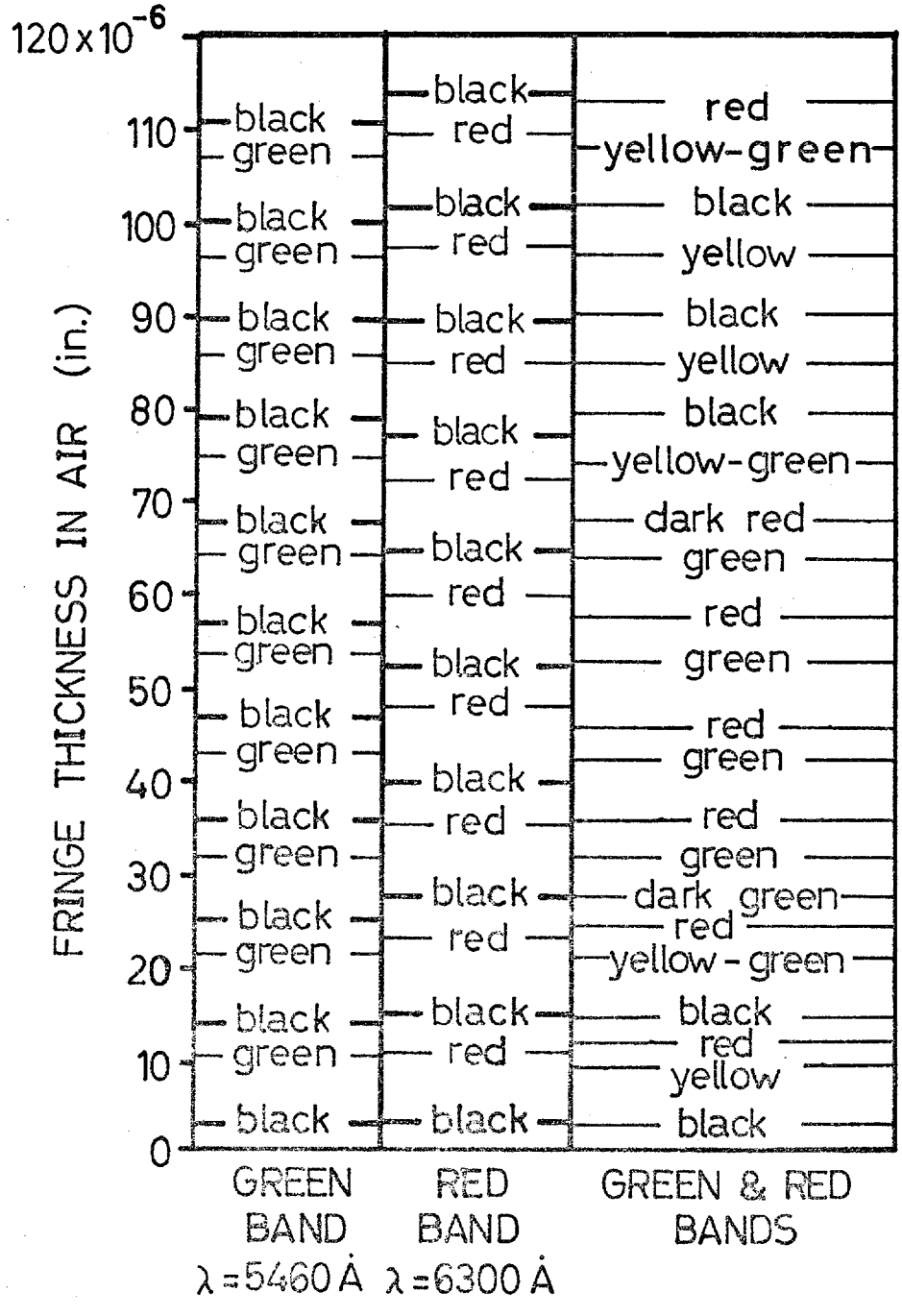


Fig.15 Fringe thicknesses corresponding to the monochromatic and duochromatic systems

fringes of each waveband are again in phase. The corresponding chromatic fringes then become yellow and black since yellow is the result of the addition of green and red.

3.7 Multiple-beam Interferometry

a. Effect of reflectivity

When the surfaces of the interferometer are coated with a highly reflecting material, multiple reflections occur as shown in fig.16. The reflected beams are related in phase and combine in such a manner as to alter the intensity distribution of the fringes. For a reflection interferometer the intensity distribution as given by Holden(28) is

$$I = R_1 + \frac{T_1^2 R_2 + 2T_1\sqrt{R_1R_2}\cos(\Delta - F) - 2T_1R_1R_2\cos F}{1 + R_1R_2 - 2\sqrt{R_1R_2}\cos\Delta}$$

where $F = 2\gamma_1 - \beta_1 - \phi_1$

$\Delta = h + \beta_1 + \beta_2$

β_1 & β_2 = phase change on reflection at surfaces 1 and 2 within the medium of oil

ϕ_1 = phase change on reflection of surface 1 within the medium of glass

γ_1 = phase change on transmission of surface 1

h = physical separation between surfaces

R_1, R_2 = reflectivities of surfaces

T_1 = transmission of surface 1

As the effective reflectivity ($\sqrt{R_1R_2}$) of the surfaces increases, the dark fringes become highly sharpened. This is shown in fig.17 for reflectivities of $R_1 = R_2 = 27\%$ and $R_1 = R_2 = 87\%$.

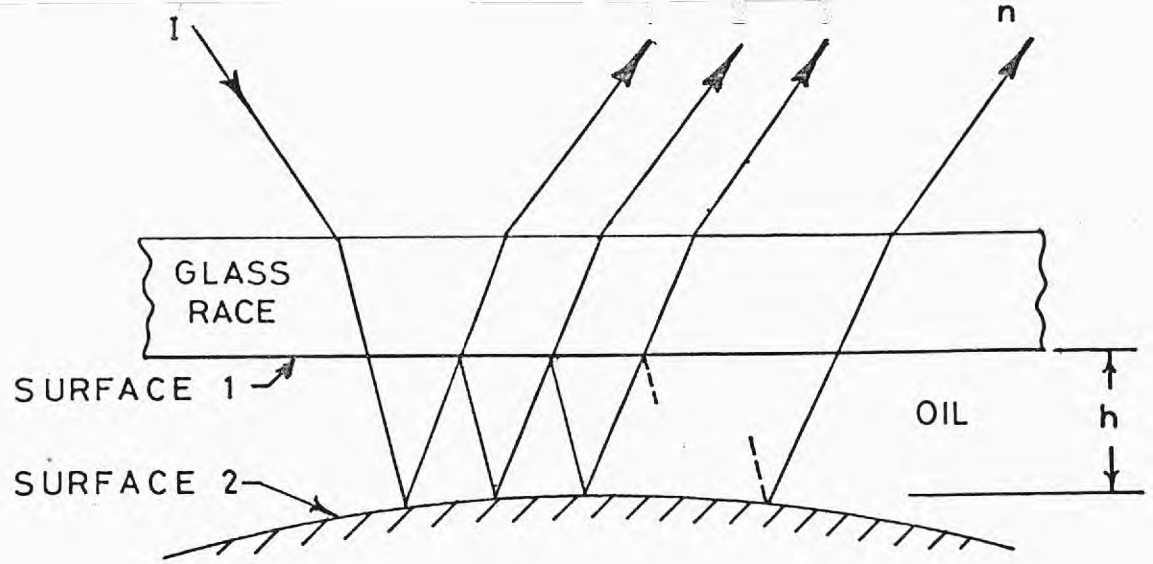


Fig.16 Multiple reflections due to highly reflecting surfaces

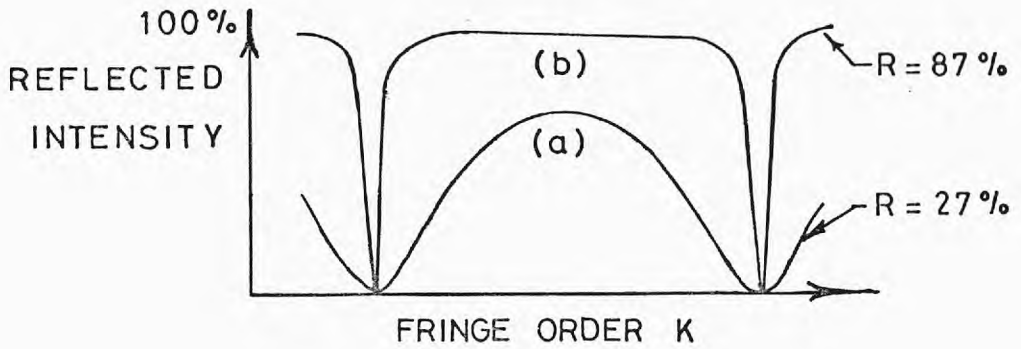
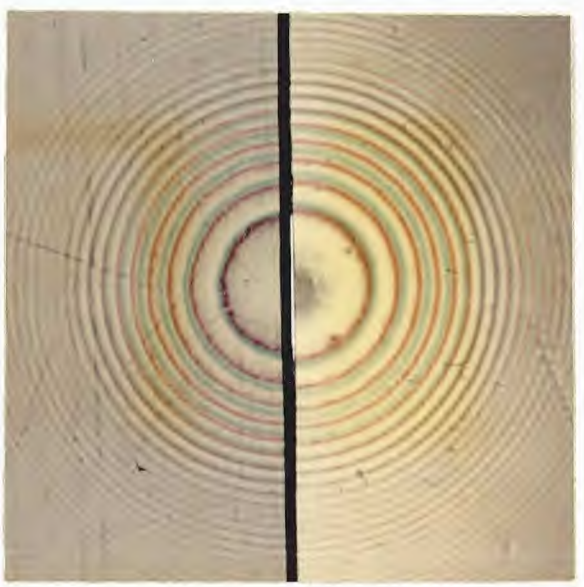


Fig.17 Effect of reflectivity on intensity distribution

Al. on glass,
Al. coated
ball

(in oil)



TiO_2 on glass,
steel ball

(in oil)

Fig.18 Duochromatic multiple beam fringes

b. Effect of absorption

The intensity distribution given by the above equation is a function of the transmission (T_1) of the first surface and the reflectivity of both surfaces (R_1 and R_2). If the reflecting surfaces absorb some of the light, the reflectivity of the surfaces is reduced. This will increase the width of the dark fringes. More importantly, however, is the effect of absorption (particularly of the first surface) on the transmission (T_1). Absorption can seriously reduce the transmission in a highly reflective film. The effect of this absorption prevents the minimum intensity from reaching zero. This can seriously effect the fringe visibility in a multiple-beam interferometer.

c. Establishment and uses of multiple-beam fringes

The establishment of highly sharpened multiple-beam fringes of good visibility is primarily a function of the choice of reflecting surfaces. Highly reflecting films of silver are often used for multiple-beam interferometry; however, the durability of silver films for lubrication experiments is lacking. Chromium has good durability, but its absorption is too high to obtain multiple-beam fringes of both good visibility and sharpness. Greater success appears to lie in dielectric films which can have the combination of good durability, high reflectivity, and low absorption. Westlake(20, 25) has successfully used TiO_2 and various multilayer stacks. The multilayer stacks are designed to give high reflectivity for a selected wavelength. They consist of layers of dielectric materials each having the same optical thickness of one-quarter wavelength but different refractive

indicies. Their design is similar to the non-reflecting films discussed in section 3.5d but produce an opposite effect, namely, increased reflection. Multilayer stacks give high reflectivity for only a narrow range of wavelengths thus limiting their use to monochromatic sources.

The multiple-beam fringes produced with a monochromatic source consist of narrow dark fringes separated by broad light fringes. Accurate measurements can be obtained with the dark fringes since their fringe position can be determined with very good precision. However, the light fringes are unfortunately too broad for useful measurement. This results in a rather large fringe spacing which prevents the estimation of film thickness at intermediate fringe positions and hinders the determination of film shape.

A greater number of multiple-beam fringes can be obtained by using a chromatic source. This requires a reflecting layer with low absorption and high reflectivity over a large range of wavelengths. Fig.18 shows two attempts using the duochromatic green and red system. The fringes on the left-hand side were obtained by coating both the glass and ball surfaces with a metal layer of aluminium. The fringes on the right-hand side were obtained with a single quarter-wavelength dielectric layer of TiO_2 on glass. The aluminium coating has a higher reflectivity than TiO_2 and therefore gives narrower fringes. However, the durability of the TiO_2 layer is superior to aluminium and is thus probably more suited for lubrication experiments.

3.8 Optical Equipment

a. Microscope

The interference fringes were observed with a microscope, adjustable along three axes. A magnification of 50X was found to be convenient for observing both the Hertzian region and the inlet lubricant boundary for the lubricant starvation experiments. A 5X objective (N.A. 0.12) was used in conjunction with a 10X wide field eyepiece. The wide field eyepiece was useful for locating the fringes when setting up the equipment; however, a certain amount of distortion is associated with the edges of the field of view. The magnification of 50X made it possible to obtain all the measurements within the central region of view where distortion is negligible.

b. Light source

(1) Xenon discharge lamp

The illuminating source is a xenon-filled high voltage discharge lamp (Ferranti type CC10) which has a colour temperature of about 6000^{o}K and a continuous spectrum similar to daylight. The lamp, which is shown in fig.19, is enclosed within a fibre cylinder to provide adequate insulating protection. A Wratten 77A filter attached to the cylinder establishes the duochromatic green and red system discussed in section 3.6 . The lamp is cooled by compressed air which is throttled through two 0.020 inch diameter holes on one side of the cylinder and is exhausted on the other side.

(2) Electric flash unit

The circuit diagram of the electronic flash unit

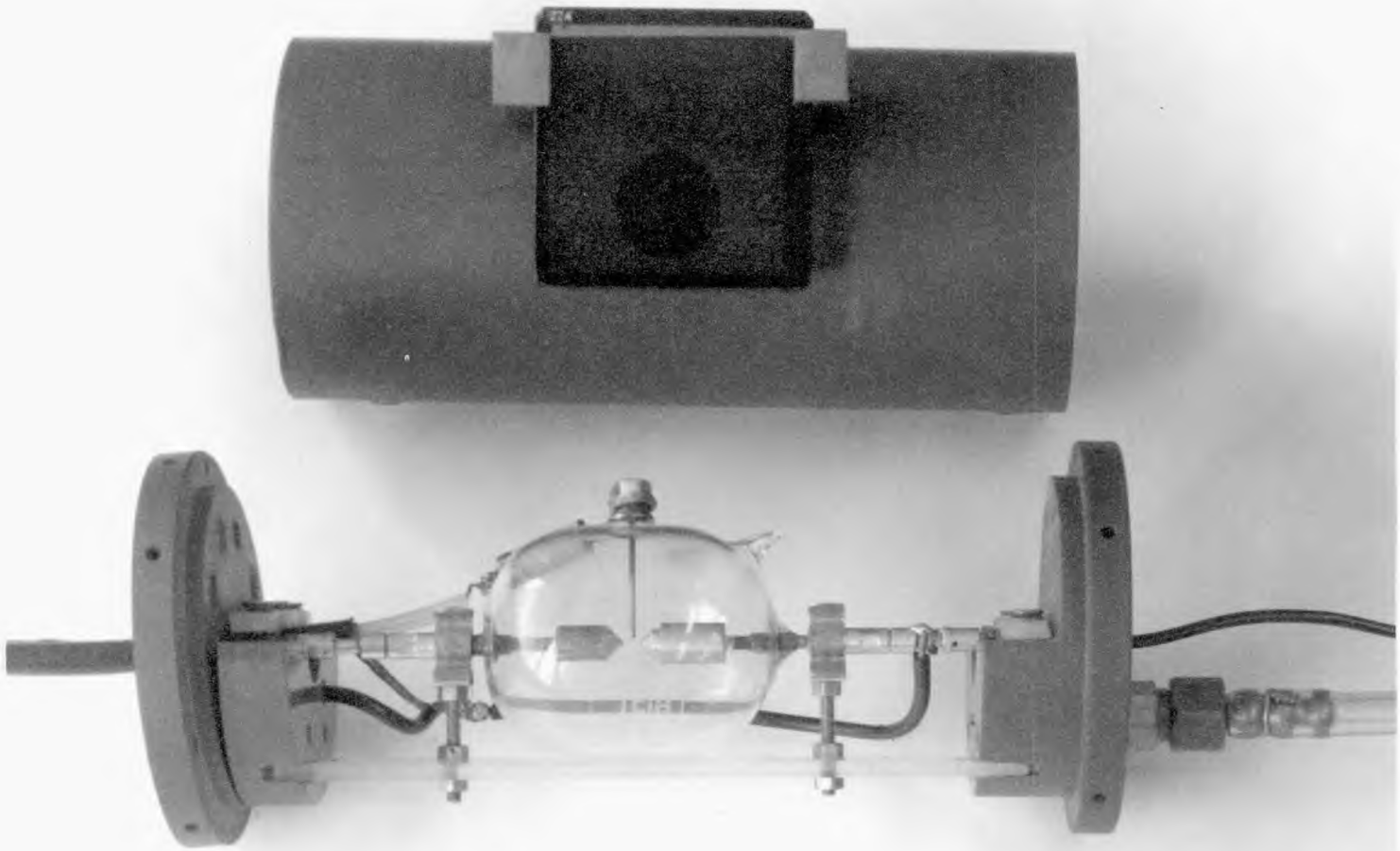


Fig.19 Xenon discharge lamp, housing and filter

is shown in fig.20. The flash unit design is somewhat hybrid in nature since the CC10 flash lamp had only recently been developed at the time of purchase, and only a basic circuit diagram of the power supply was available from the manufacturer. The electronic flash unit provides the required electrical energy across the electrodes of the xenon lamp. Various capacitors (0.5 to 198 micro-farads) are charged to a potential of 1000 volts derived from a DC voltage source. The discharge is initiated by a trigger electrode which provides a very high voltage impulse (8 KV) obtained from the secondary side of a firing transformer (PT 60). The primary side of the transformer is connected with a capacitor (1 micro-farad) in conjunction with a high-peak current relay valve (EN 10). An input pulse to the relay valve commences the discharge operation.

(3) Triggering

The input pulse is derived from either an internal or external signal which provides two basic modes of operation.

The internal signal comes from a variable frequency multivibrator (ECC 81). Its repetitive signal discharges the lamp continually. The frequency can be adjusted to a sufficient rate so as to make the light source appear continuous to the eye. This mode of operation was used for the visual observation in the counterrotation arrangement where the balls rotate in a fixed position in space.

The external signal is produced by a Dawe Phototransistor Pickup. This mode of operation was used for stroboscopic observation and for high-speed photomicrography. To isolate the contact area of a particular ball while the bearing is in operation, the xenon lamp is synchronized with the cage rotation.

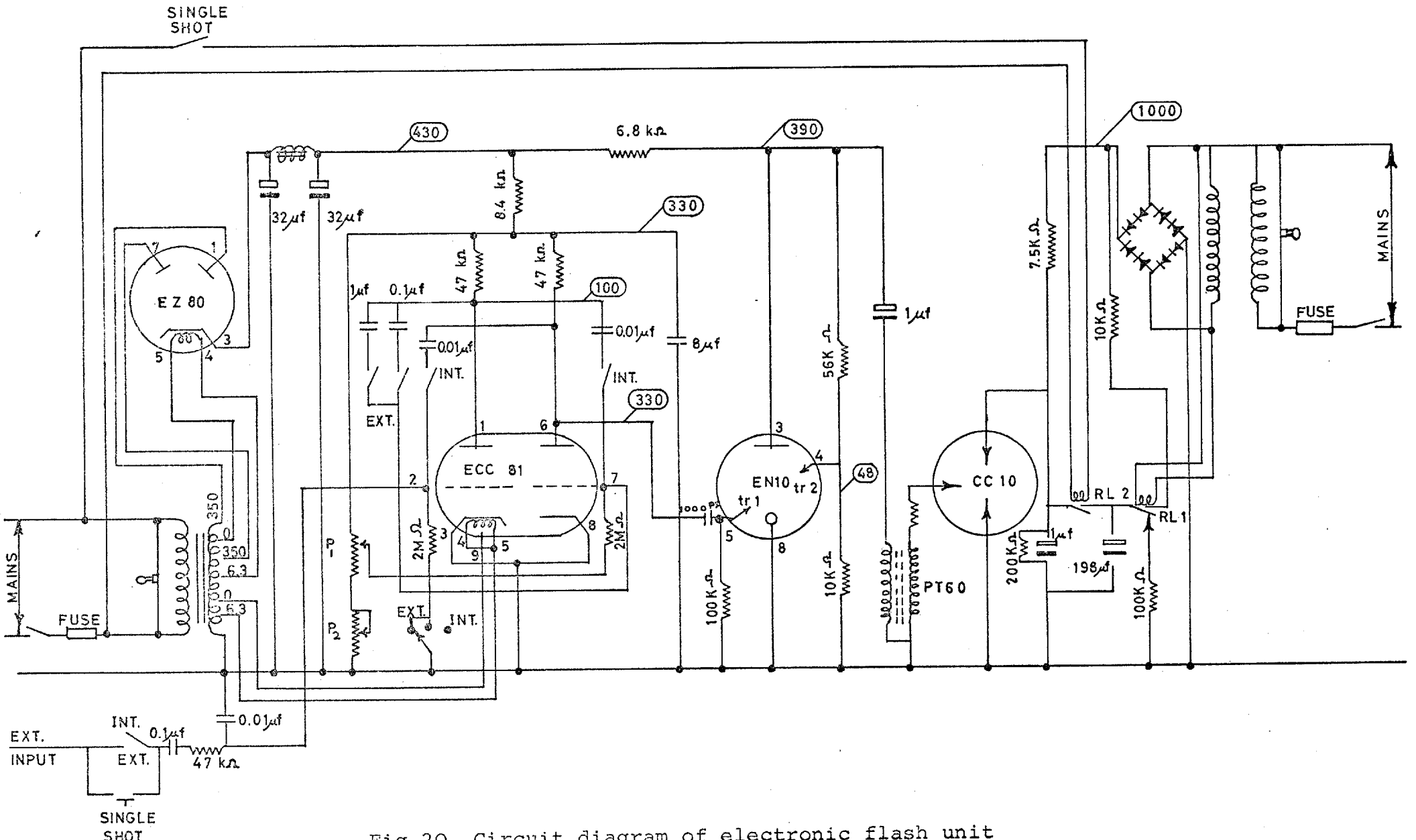


Fig.20 Circuit diagram of electronic flash unit

The cage is furnished with a small mirror which reflects light from a lamp and lens system onto a phototransistor when the ball passes beneath the point of observation. This is shown in fig.2 and diagrammatically in fig.1. The signal from the energised phototransistor is applied to a two-transistor bi-stable circuit. An inductance contained within the switching circuit provides a pulse of 30-40 volts. In this way the light source is synchronized with the cage rotation so that the interference fringes of a particular ball can be observed stroboscopically.

c. High-speed photomicrography

High-speed photomicrography provides a recording of the condition of lubrication at an instant in time. In this way the film thickness under lubricant starvation conditions was obtained as a function of a transient inlet lubricant boundary. Also, a high-speed photomicrograph enabled the oil film rupture in the exit region to be clearly seen.

The exposure of a high-speed photomicrograph was obtained by using a single discharge from the xenon lamp. For the stroboscopic arrangement this discharge was synchronized with the cage rotation by using the phototransistor in the manner discussed in the previous section. The phototransistor also commences the discharge for the counterrotation arrangement. In this case, however, the signal is produced by manually interrupting the supply of light to the phototransistor. The photomicrographs were recorded with a 35mm camera attached to the microscope. The microscope objective and eyepiece were used to form the image on the camera. Various magnifications were achieved by adjusting the camera distance

from the eyepiece by means of extension rings.

The exposure was controlled solely by the electrical energy supplied to the xenon lamp. Camera aperture, shutter speed and its synchronization to the event are, therefore, not involved. The flash energy, E, supplied to the lamp is given by

$$E = \tfrac{1}{2}CV^2$$

where C is the capacitance and V the applied voltage. The required flash energy, which depends on the magnification, the type of illuminator and the speed of the film, was determined by trial exposures. Kodak High Speed Ektachrome (daylight type) was used for colour photography. For normal conditions a flash energy of 100 joules (V = 1000 volts, C = 200 micro-farads) was found to be compatable with its film speed of ASA 160. Ilford HP4 was used for black and white photography. It has an ASA speed of 650 when developed with a high-speed developer (Microphen). A flash energy of 33.5 joules (V = 1000 volts, C = 67 micro-farads) was normally used for this film.

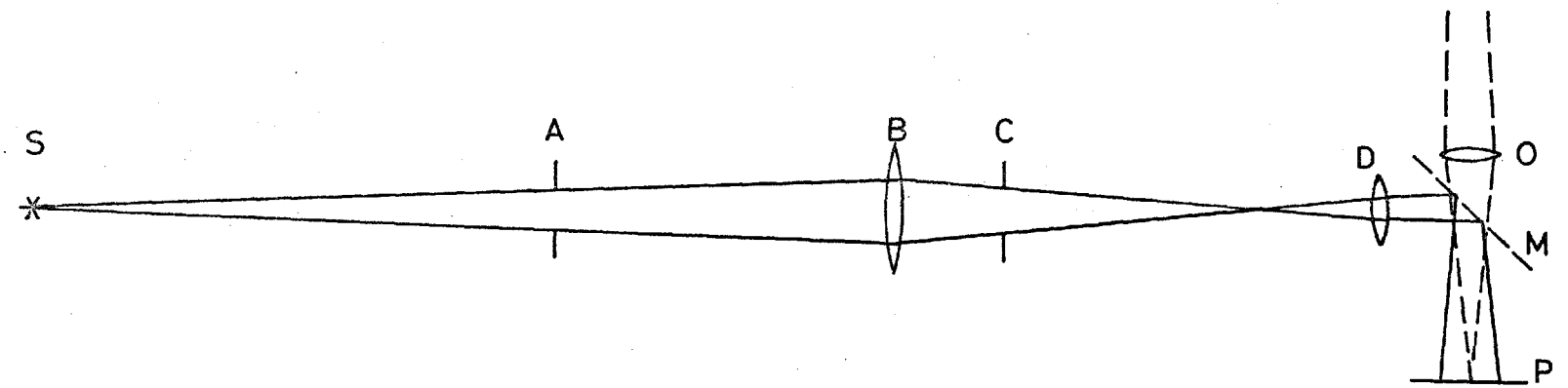
High surface speeds, where both a short flash duration and sufficient illumination were required, were the most difficult to photograph. The flash duration, which is defined as the interval of time the flash intensity is above one third the maximum, is primarily a function of the discharge capacitance. A capacitance of 200 micro-farads gives a flash duration of about 200 micro-seconds, whereas a capacitance of 1 micro-farad has a flash duration of approximately 10 micro-seconds. Short flash durations imply low capacitance and, therefore, low flash energies. Very low flash energies cannot be tolerated because a large percent of light (90%) is lost within the microscope illuminator and the inter-

ferometer. However, by 'overdeveloping' the high-speed black and white film it is possible to obtain satisfactory exposures with a capacitance of 1 micro-farad. This enables surface speeds on the order of 100 in./sec. to be photographed.

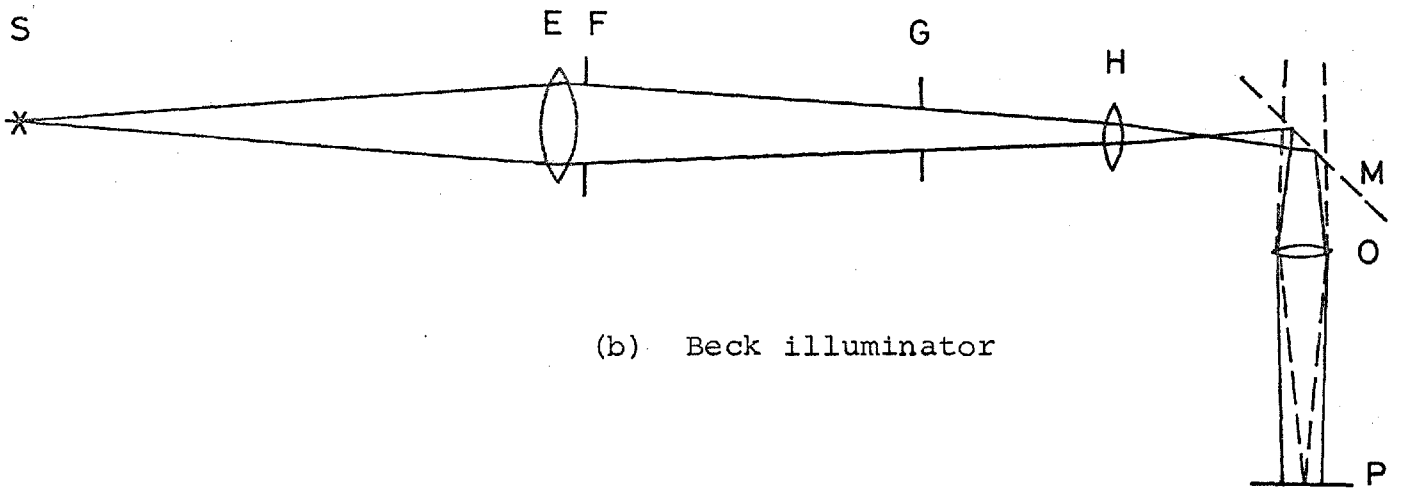
d. Vertical illuminators

The basic purpose of a vertical illuminator is to illuminate the interferometer in a direction normal to the reflecting surfaces. For satisfactory visual observation and photomicrography it is important that the vertical illuminator provides adequate means for controlling the intensity and reducing glare. For purposes of good fringe visibility the vertical illuminator must also provide normal illumination within a certain range of angles of incidence. The effect of non-normal illumination (i.e. the source or aperture size) was discussed in section 3.4d and a limiting criteria was derived.

A vertical illuminator (or collimator) designed by Professor Hopkins(29) for the Imperial College Lubrication Laboratory is shown in fig.21a. Light from the source S is reflected normally onto the object plane P by a semi-reflecting glass M, which is located below the objective O. A is a field stop which is focused on the object plane P. The field stop diameter reduces glare by limiting the area illuminated. B is a field lens which illuminates an aperture stop C. The aperture stop is located at the focal plane of an aperture lens D. This arrangement is similar to the collimator of fig. 10 where the source S is placed at the focal plane of a lens. The aperture stop controls the intensity



(a) Hopkins illuminator



(b) Beck illuminator

Fig.21 Vertical illuminators

of illumination. It also controls the range of angles of incidence (i.e. degree of collimation) in the same manner as the source size in fig.10. This duo role of the aperture stop implies a compromise between the intensity of illumination and the degree of collimation.

This illuminator was used by Foord et al(16), and a similar one with adjustable field and aperture stops was used by the author for some preliminary work. This work pointed out two limitations. First, it was found that the illuminator provided very good collimation, but at the expense of intensity which was greatly needed for high-speed photomicrography. Second, it was found that having the illuminator below the objective severely restricted the working distance for many applications.

These two limitations led to a consideration of the compromise between the intensity of illumination and the degree of collimation. For a good fringe visibility of 0.98, equation 3.3 gives the limiting aperture size θ . θ is the angle by which the entire aperture C can be seen from centre of the aperture lens D. It is also the maximum angle of deviation from normal for any ray coming from the illuminator. To evaluate the allowable aperture size, a refractive index (n) of 1.5 was used for the media of glass and oil. This has a beneficial effect in that it causes the light to be refracted toward the normal as it passes from air to glass (see fig.10). The wavelength of the green band (5460 $\mathring{A}$ or 21.5 micro-inches) was used for λ,and the largest calibrated fringe thickness (in oil) of 80 micro-inches was used for h. Equation 3.3 gives the limiting aperture size of $\theta = 11.8^{o}$. This is considerably larger than the maximum aperture size obtainable in the illuminator (4^{o}).

Thus, it appears that over the required film thickness range of 80 micro-inches the restriction on collimation is not very severe. Moreover, under these conditions it is perfectly acceptable to use an ordinary commercial incident illuminator, which provides greater intensity and also a longer working distance by having the illuminator located above the objective. Most of the experiments were conducted with a Beck Vertiphase illuminator shown in fig.21b. It was not designed to produce rays of parallel light, but the maximum deviation from normal (8^{o}) is well within the limit stated above. The objective O and the semireflecting glass M are exactly the same componants used in the Hopkin's illuminator. E is an optical glass aperture lens and F is an aperture stop. The latter controls the intensity and to some extent the degree of collimation. G is a field control iris diaphragm which limits the area illuminated, and H is an objective field lens.

In section 3.4d it was shown that when the interferometer is illuminated with rays which depart from normal, the fringes are shifted to a position of greater thickness. The increase in thickness Δh associated with this shift is given by equation 3.2, and it is proportional to

$$\frac{1}{\cos r} - 1$$

where r is the refracted angle of incidence (see fig.10). If a ray coming from the illuminator has a maximum angle of deviation from normal of 8^{o}, then the refracted angle of incidence r according to Snell's law is 5.32^{o}. This gives an increase in thickness of 0.43%. Since the illuminator gives a range of angles of incidence between O and 8^{o} the effect is an integrated one which will make the effective increase in thickness less than 0.43%. Over the maximum film thickness

range of 80 micro-inches this change in thickness is almost too small to detect. Moreover, it becomes irrelevant if the fringes are independently calibrated.

3.9 Calibration

a. Introduction

The absolute thickness corresponding to each fringe of the systems shown in fig.14 must be determined by calibration. This is necessary because of the anomalous phase changes associated with the metalic reflections of the steel ball and the chromium layer (see section 3.5e). Furthermore, some of the coloured fringes of the duochromatic system are not saturated, spectral colours. Their position is somewhat subjective to the individual and cannot be accurately determined from the knowledge of the two basic wavelengths involved.

The calibration procedure used by Cameron and Gohar(15) was to measure the diameters ($2r_f$) of the fringes produced by the reflecting surfaces of a steel ball of known radius R; the ball was allowed to 'touch lightly' a flat glass plate. Foord et al(16) raised a chromed spherical lens having a large radius of curvature until it 'just touched' the glass plate. This provided larger fringe diameters which could be more easily measured; however, the opaque chromium surface has a different phase change on reflection than the steel balls used in the experiments. If the fringe radius r_f is much smaller than the radius of the sphere, then the thickness h can be calculated from the "known" geometry by the relation

$$h = \frac{r_f^2}{2nR} \qquad \ldots(3.4)$$

Because of the very small thicknesses that h may acquire, and because h is very sensitive to the measured values of r_f, ($h \propto r_f^2$) two significant errors may occur. First, if the sphere and glass plate are allowed to contact each other as shown in fig.22, a small but finite load that may be present will cuase the surfaces to deform, giving rise to a Hertzian contact area of radius a. This will tend to make the measured values of r_f too large and thus cause the calculated values of h for each fringe to becomes greater than the actual value. Secondly, if there is nearly zero load applied to the system, the mean surface profiles may become separated as the result of surface roughness or dust particles between the surfaces. This has an opposite effect on the measured values of r_f which will cause the calculated values of h to become smaller than the actual values. These two errors became apparent during the calibration experiment described below.

b. Calibration experiment

A one-inch diameter steel ball which was used in the experimental bearing was loaded against the glass race with a balance sensitive to less than 1/10 gram. The fringe diameters ($2r_f$) were measured with a filar eyepiece for a series of loads between 0.5 and 500 grams. At small loads the system was very sensitive to vibrations produced within the college building and from traffic moving down Exhibition Road. To overcome these external vibrations, a viscous damper was employed in the system and the readings were taken during the night. The measured values of the fringe radii r_f

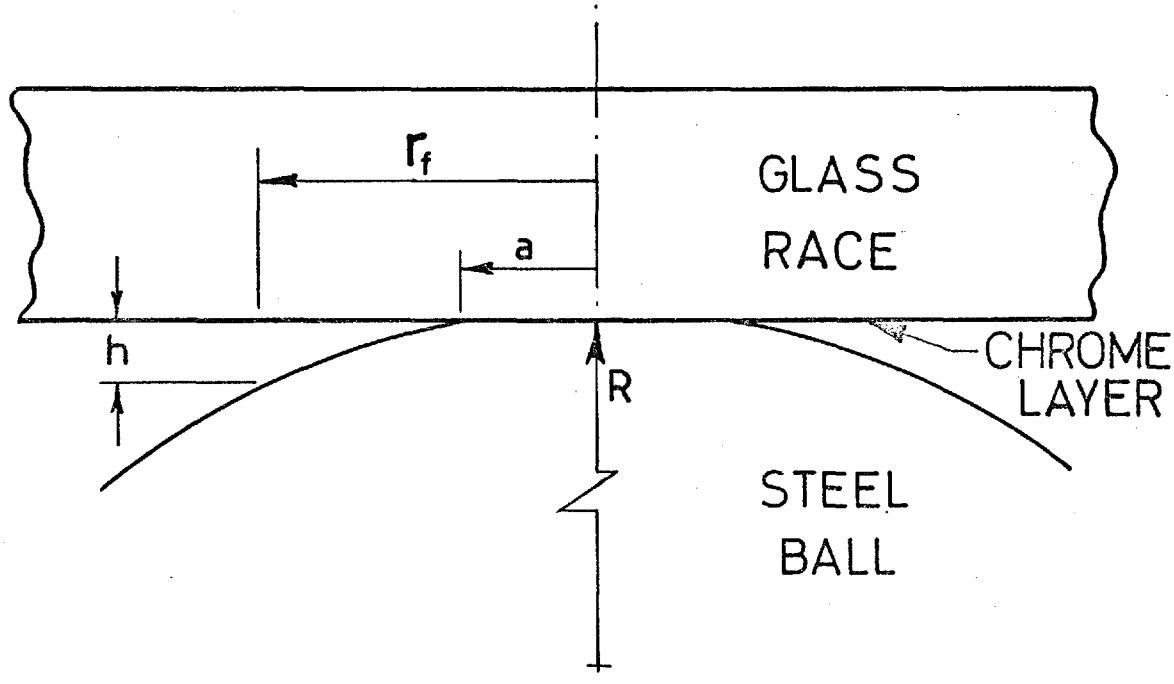


Fig.22 Ball and race contact for fringe calibration

were plotted against the calculated Hertzian radius (a) for each load. The Hertzian radius (a) was obtained from the following relation(30)

$$a = \left(\frac{3}{2}\frac{wR}{E'}\right)^{1/3} \qquad \ldots(3.5)$$

where w = applied load

R = effective radius, $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$, (R = 0.5in.)

E' = effective elastic modulus of materials

$\frac{1}{E'} = \frac{1}{2}\left(\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right)$, $E' = 16.92 \times 10^6$ psi

E_1 = elastic modulus for glass (10.76×10^6 psi)

E_2 = elastic modulus for steel (30×10^6 psi)

σ_1 = Poisson's ratio for glass (0.223)

σ_2 = Poisson's ratio for steel (0.30)

The plot of r_f vs a for the first seven dark fringes of the green band (λ = 5460 Å) is shown in fig.23. The shape of each curve can be deduced by considering the influence of load on the fringe radius r_f. As the load approaches zero, the Hertzian radius should become a negligible proportion of r_f. Thus, the curve of a vs r_f should approach the asymptote r_f = constant. The experimental data deviated from this condition for loads below 10 grams. Also, it was noted during the experiment that as the load was reduced from 10 grams, the central fringe corresponding to the Hertzian contact area changed from light grey to black. This suggested that the surfaces were being separated, possibly as a result of surface roughness or dust particles.

To eliminate the errors due to the imperfect contact between the two surfaces the curves were extrapolated to zero load. This was accomplished by drawing curves through the data points of the higher loads only and judiciously extending the curves to zero load so that they approached an asymptote

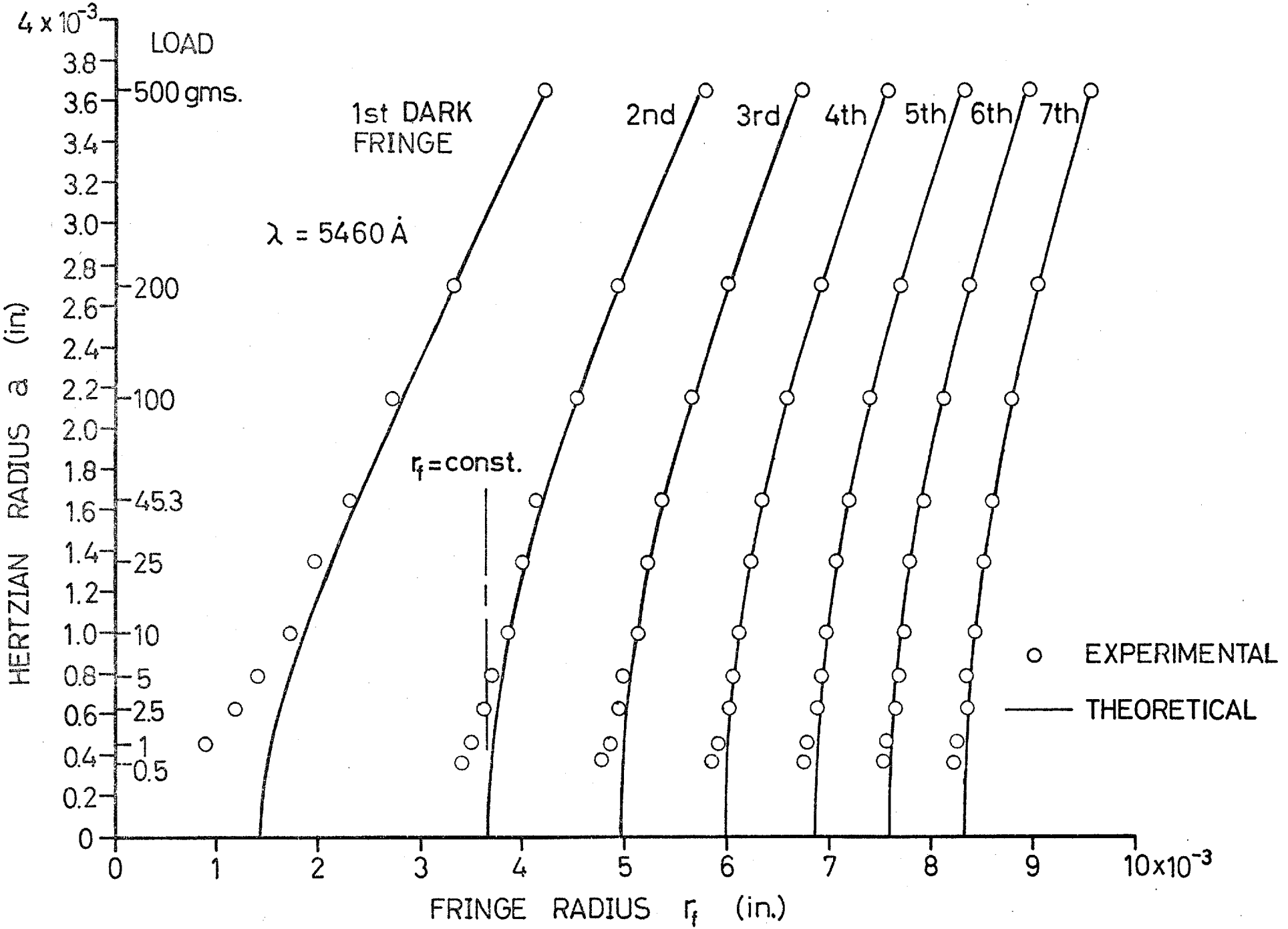


Fig.23 Effect of contact deformation on fringe radius

r_f = constant. For each curve this constant is the required fringe radius (r_f) for the calculation of the fringe thickness h in equation 3.4. All the fringes were calibrated in this way. The results for the duochromatic system are shown in fig.24.

To verify the shape of the extrapolated curves the theoretical shapes of the curves a vs r_f were determined from the Hertzian equation for the gap outside the main contact region(30).

$$h = \frac{a\, p_{max}}{E'} \left[-\left(2-\frac{r_f^2}{a^2}\right) \cos^{-1} \frac{a}{r} + \left(\frac{r_f^2}{a^2} - 1\right)^{\frac{1}{2}} \right] \quad \ldots(3.6)$$

where $p_{max} = \frac{3}{2} \frac{w}{\pi a^2}$

This equation was solved for r_f as a function of the Hertzian radius by using a trial and error procedure. The fringe thickness h, which must be constant, was first determined from the measured values of r_f at a reference load of 200 grams. The solid curves shown in fig.23 were obtained in this manner. They verify the assumption that the curve of a vs r_f approaches an asymptote of r_f = constant as the load approaches zero.

The fringes can also be calibrated by using the fringe radii r_f determined from the theoretical curves at zero load. These were also used in equation 3.4 to calculate the fringe thicknesses. They agree with the results of fig.24 to within 4%.

It must be noted that the theoretical curves provide only the theoretical <u>shape</u>, whereas their <u>position</u> is determined by the data of the reference load. It should also be appreciated that the latter method of calibration is fundamentally equivalent to calculating the fringe thicknesses directly from equation 3.6 using the measured values of r_f at a load of 200 grams.

FRINGE COLOUR	THICKNESS IN AIR (INCHES)
light grey	0.0×10^{-6}
black	3.20
yellow	9.55
red	12.6
black	14.8
yellow-green	21.2
red	24.5
dark green	28.0
green	32.0
red	36.1
green	42.7
red	46.5
green	53.2
red	58.0
green	64.0
dark red	68.2
yellow-green	74.6
black	79.5
yellow	85.0
black	90.8
yellow	96.0
black	102.0
yellow-green	108.0
red	113.5

Fig.24 Results of fringe calibration

This now results in two useful methods of calibration which can be summarized as follows. The first method is to obtain the fringe radii at zero load by extrapolation. This can be accomplished by measuring the fringe radii for a series of loads approaching zero and extrapolating the curves in a manner consistent with the condition that r_f = constant at zero load. This eliminates the errors due to the imperfect contact between the surfaces. The extrapolated values of r_f at zero load are then used in equation 3.4 to calculate the corresponding fringe thicknesses. The second method is to obtain the fringe radii for a system which has been elastically deformed under a given load. The measured fringe radii are then used in equation 3.6 to calculate the corresponding fringe thicknesses. This should be done for several loads from which an average is taken. The second method is experimentally easier to carry out; however, it relies on a knowledge of the elastic properties of the materials to a much greater extent than the first method.

The fringes calibrated in the above experiment were obtained in air which has a refractive index very close to one (1.0003). To obtain the fringe thickness for a medium other than air the calibrated fringes must be divided by the refractive index of the medium used. It is possible that a change in refractive index from air to oil ($n \simeq 1.5$) alters the phase change on reflection and, therefore, upsets the fringe calibration. This possibility was checked by calibrating the fringes of the duochromatic system in both media and comparing the results. The ball was applied to the glass race with a load of 100 grams and the fringe thicknesses were cal-

culated from equation 3.6 using the measured fringe radii. The fringes which were calibrated in air were divided by the refractive index of the oil (n = 1.4938). The results, which correlated to within ± 2%, showed no consistant error which could be attributed to an alteration of the phase change on reflection resulting from the two media. It is assumed, therefore, that the calibrated fringes provide the correct optical thickness to within about 4% accuracy. The actual thickness requires a knowledge of the refractive index of the medium. This is the subject of section 5.2.

c. Measurement of phase change on reflection

The phase change correction ϕ can be obtained from the calibrated fringes of the two monochromatic systems. This is shown in fig.25 where the fringe thickness is plotted against the fringe order. The fringe order is given in terms of the light fringes so that a light fringe occurs at zero order. This is used as a reference point since a light fringe will occur for the condition of no relative phase change between the reflecting surfaces (i.e. $\phi = 0$). At zero thickness the fringe order is displaced from zero by an amount equal to $\frac{3\pi}{5}$. This is the numerical value of the phase correction ϕ.

The phase correction ϕ is the relative phase change between the reflecting surfaces. It is determined by the absolute phase change of the steel surface (ϕ_{steel}) and the chromium surface ($\phi_{cr.}$). That is

$$\phi = \phi_{steel} + \phi_{cr.}$$

The experiment discussed in section 3.5e indicated that $\phi_{steel} = \pi$ (or very nearly so) and that $\phi_{cr.}$ was neither 0 nor π.

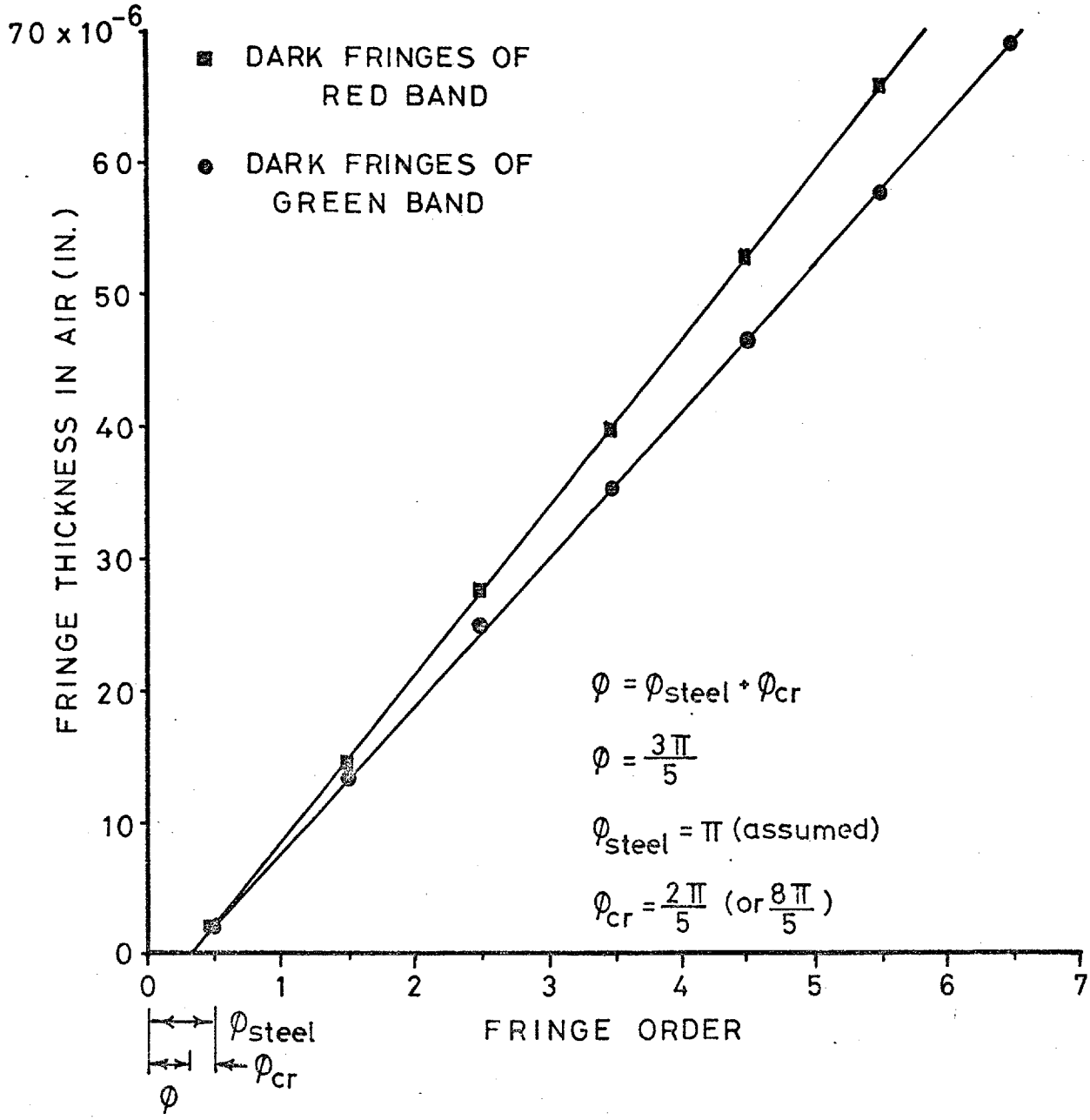


Fig.25 Fringe thickness vs fringe order showing effect of phase change on reflection

$\phi_{cr.}$ can, therefore, have two possible values: $-\frac{2\pi}{5}$ or $\frac{8\pi}{5}$. The former is probably the more likely. These two values can be deduced by remembering that ϕ_{steel} causes the zero order fringe to becomes dark, and that $\phi_{cr.}$ alters this, but in a direction which makes the first predominant fringe dark.

The phase change correction of $\frac{3\pi}{5}$ was also obtained from the microphotometer plot of the fringe intensities described in section 3.5e. By knowing the intensity of the central fringe it is possible to calculate the phase change correction by assuming that the intensity distribution follows the $\cos^2$ law. It should also be noted that since the eye is not very sensitive to small changes in intensity, this method can be used to detect small changes in film thickness and to establish the film thickness at intermediate fringe positions with considerable accuracy. This provides feasible measurements below the first black fringe which occurs at about 2.1 micro-inches (540 Å) in oil.

3.10 Conclusions on Interferometry

a. The requirements for good fringe visibility can be used to design an interferometric system which is compatible with the experimental measurements to be taken. These requirements are concerned with the characteristics of the reflecting surfaces and the type and method of illumination.

b. The reflectivity of the surfaces must provide interfering beams of equal intensity (i.e. $I_1 = I_2$). A partially reflecting layer of chromium satisfies this requirement and also provides a very durable surface.

c. Fringe intensity changes from maximum to minimum (or vice versa) when the thickness changes by $\frac{\lambda}{4n}$. Therefore, to obtain fringes of high visibility, the local irregularities of the surfaces in the vertical direction must be much smaller than $\frac{\lambda}{4n}$. This requires surface roughness values which are on the order of 1 micro-inch c.l.a. The glass race and the specially made steel balls amply satisfy this condition.

d. The requirements of the type and method of illumination in connection with fringe visibility is a function of the experimental film thickness range required.

e. A sufficient "degree of coherence" between the interfering beams must be maintained over the required film thickness range. This can be thought of in terms of the length of the emitted wave trains or the width of emitted waveband(s). A duochromatic system which is composed of green and red wavebands provides fringes of good visibility over a satisfactory film thickness range of 80 micro-inches. The fringes are closely spaced and have good colour contrast. The definite sequence of coloured fringes provides a convenient means for determining the fringe order.

f. For intensity considerations a source or aperture of some finite size must be tolerated. This leads to a range of angles of incidence which confuses the fringe position and, therefore, decreases the visibility. To obtain a good fringe visibility of 0.98 the limiting aperture size is given by

$$\theta = 0.324 \left(\frac{n\lambda}{h} \right)^{\frac{1}{2}}$$

For a maximum film thickness of 80 micro-inches the restriction on the aperture size is not too severe ($\theta = 11.8^{o}$). This allows the use of a standard incident illuminator which provides adequate intensity and control of illumination as well as a long working distance.

g. The method of calibrating fringes by applying together two surfaces of known geometry until they are 'just touching' can lead to two significant errors. These errors result from the inability to obtain a perfect contact between the two surfaces.

h. Two methods of calibration were found suitable. The first method attempts to obtain the fringe radii (r_f) for a condition of perfect contact by extrapolation. The fringe thicknesses are then calculated from the equation for undeformed surfaces. The second method deforms the surfaces under a given load. The fringe thicknesses are then calculated using the Hertzian equation for the gap outside the contact region. The second method is easier to carry out; but it relies on a knowledge of the elastic properties of the materials.

i. The anomalous phase changes associated with metalic reflections can be elucidated by simple experiments. The numerical value of the phase correction ϕ was obtained from the calibrated fringes of the two monochromatic systems. This phase correction is the relative phase change between the two reflecting surfaces. It causes the fringes to shift by an amount equal to $\frac{3\pi}{5}$.

Chapter 4 THE STATE OF ELASTOHYDRODYNAMICS

4.1 Introduction

The fundamental aspects and general features of ehd lubrication in conjunction with previous literature are presented in this chapter. The purpose of this chapter is to establish those aspects of the subject which are relevant to the explanation and development of the results presented in later chapters.

The development of ehd has mainly occurred over the last twenty years. It has reached a state where the fundamentals have become quite well established and are now being exploited in practical applications. The relatively rapid development has been aided by the coupling of experimental and theoretical investigations. These have elucidated the effects of load, speed, material properties, and geometry in connection with film shape, film thickness, and pressure distribution.

4.2 Reynolds Equation

The formulation of hydrodynamic lubrication is based on the Reynolds equation which describes the generation of pressure in lubricating films. The equation in two dimensions for an incompressible fluid is

$$\frac{\partial}{\partial x}\left(\frac{h^3}{\eta}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{h^3}{\eta}\frac{\partial p}{\partial y}\right) = 12\ u\left(\frac{\partial h}{\partial x}\right) \qquad \ldots(4.1)$$

where u is one-half the sum of the surface speeds in the x-direction, and p is the pressure generated in the film of thickness h having a viscosity η. The derivation of this

equation from the Navier-Stokes equations or from first principles involves the usual simplifying assumptions; see e.g. Cameron(30) and Pinkus and Sternlicht(31). These are: (a) no body forces, (b) pressure and viscosity constant through the thickness of the film, (c) curvature of the film ignored, (d) Newtonian lubricant, (e) laminar flow, no vortex flow, (f) no fluid inertia, (g) no slip at the bearing surfaces, and (h) only the velocity gradients with respect to the z direction considered.

If side leakage is neglected Reynolds equation can be integrated with respect to x to give

$$\frac{dp}{dx} = 12\ u\eta\left(\frac{h-\bar{h}}{h^3}\right) \qquad \ldots(4.2)$$

where $\bar{h}$ is the local film thickness at maximum pressure. This equation provides the important qualitative description: that pressure generation, which depends on the speed and viscosity, can only be obtained when the surfaces are converging (i.e. when h is larger than $\bar{h}$).

4.3 Classical Hydrodynamic Theory

In 1916 Martin(32) solved the Reynolds equation in an attempt to explain the absence of wear in gear teeth by the presence of a hydrodynamic oil film. Assuming rigid cylinders and an isoviscous fluid, he obtained the following film thickness expression

$$\frac{h_o}{R} = 4.896\ \frac{u\eta_o}{w}$$

where h_o is the film thickness on the line of centres. This expression predicts film thickness values which are orders of

magnitude smaller than the actual values encountered in highly loaded contacts.

An extension of the classical hydrodynamic theory to include the effects of elastic distortion was considered by Peppler (1936), 1938) and Meldahl in 1941; see Dowson and Higginson(33). Later Gatcombe (1945) considered the classical hydrodynamic theory with the added effects of the influence of pressure-viscosity characteristics of the lubricant. Neither the individual effects of elastic distortion nor pressure-viscosity characteristics could substantially improve the classical hydrodynamic theory to explain the existance of an oil film which was larger than the normally encountered surface roughness values. To predict a film thickness with the correct order of magnitude the combined effects of elastic distortion and pressure-viscosity must be considered.

4.4 Approximate Solution to the Elastohydrodynamic Problem

The first significant solution which considered the simultaneous effects of elastic distortion and pressure-viscosity characteristics was proposed by A.M. Ertel- see (30), p.203 and (34)- and published in 1949 by Grubin(35). This ingenious analysis, known as Grubin's theory, brought out the fundamental physical mechanisms associated with the lubrication of highly loaded contacts. It also provided the first significant film thickness equation for the elastohydrodynamic line contact problem. The basis of this analysis is outlined below and used later in section 6.11 for the development of a lubricant starvation theory.

The variation of viscosity with pressure is usually

given by the exponential relation

$$\eta = \eta_o e^{\alpha p} \qquad \ldots(4.3)$$

where η_o is the atmospheric viscosity, α is the pressure-viscosity coefficient, and η is the viscosity at the pressure p. This relation describes the pressure-viscosity characteristics of most lubricants with reasonable accuracy. A power law obtained by Chu and Cameron(36) from the viscosity data of the 1953 ASME report(37) was shown to be more accurate for paraffinic oils, particularly at high temperatures, but is more cumbersome to work with.

When the exponential relation is employed in the integrated form of the Reynolds equation we get

$$e^{-\alpha p} \frac{dp}{dx} = 12\ u\ \eta_o \left(\frac{h-\bar{h}}{h^3}\right) \qquad \ldots(4.4)$$

As explained by Dowson and Higginson(33), if the right-hand side of equation 4.2 represents the pressure gradient of a variable viscosity fluid, then the right-hand side of equation 4.4 represents the pressure gradient of a constant viscosity fluid. If the pressure generated by a constant viscosity fluid is called the "reduced pressure", q, then

$$\frac{dq}{dx} = 12\ u\ \eta_o \left(\frac{h-\bar{h}}{h^3}\right) \qquad \ldots(4.5)$$

and

$$\frac{dq}{dx} = e^{-\alpha p} \cdot \frac{dp}{dx} \qquad \ldots(4.6)$$

Equation 4.5 is the integrated Reynolds equation in terms of the reduced pressure, and equation 4.6 can be integrated to give the reduced pressure in terms of the actual pressure, p, and the pressure-viscosity coefficient, α.

$$q = \frac{1}{\alpha}\left(1-e^{-\alpha p}\right) \qquad \ldots(4.7)$$

To account for the effects of elastic distortion, the Grubin theory assumes that the deformation of a highly loaded lubricated contact is the same as that for the Hertzian dry contact case. This implies a semi-elliptical Hertzian pressure distribution over a parallel film of thickness h_o; and consequently $h_o = \bar{h}$. According to equation 4.2, a hydrodynamic pressure can only be generated in the convergent inlet region. The geometry of this region is determined by the film thickness h_o and the gap outside the Hertzian contact zone. For point contact the latter is given by equation 3.6. This assumed geometry facilitates the solution of the integrated Reynolds equation in terms of the reduced pressure for the convergent inlet region.

The Grubin theory also assumes that the inlet pressure, p, must reach a high value at the edge of the Hertzian zone. If p in equation 4.7 is large, $\bar{e}^{\alpha p}$ must be very small so that $q \simeq \frac{1}{\alpha}$. This assumption facilitates the establishment of a film thickness formula for highly loaded contacts. For a line contact geometry with a load per unit width of w, the film thickness is given by

$$\frac{h_o}{R} = 1.95 \left(\frac{\alpha \eta_o u}{R} \right)^{0.73} \left(\frac{w}{E'R} \right)^{-0.091} \qquad \ldots (4.8)$$

Ertel went on to consider how certain physical conditions would alter the results connected with the initial assumption of a Hertzian dry contact shape. In this way he brought out many of the characteristic features of the highly loaded ehd problem. In the inlet region the generated hydrodynamic pressure should in some way approach the Hertzian pressure. Furthermore, this inlet pressure will cause an additional deformation which will increase the effective radius

of the converging region. This will be beneficial to the pressure generation.

To account for the falling pressure at the exit, a constriction in the film must occur near the end of the Hertzian zone such that h is less than h_o. To accomodate this constriction and to maintain a parallel film over the main portion of the Hertzian zone, he predicted that a pressure peak must occur just prior to the constriction. These characteristics were brought out more clearly in later solutions.

4.5 Full Solutions

A solution to the ehd problem which took a more accurate account of the elastic deformation than the Grubin solution was given by Dowson and Higginson(38) in 1959. They calculated the normal displacement v for a semi-infinite solid and added the displacement to the curved surface of an equivalent roller. Thus,

$$h = h_o + \frac{x^2}{2R} + v$$

where h_o is the film thickness at the centre of the undeformed roller. The normal displacement v due to a variable pressure p(s) between $x=s_1$ and $x=s_2$ was given by

$$v = -\frac{2}{\pi E'} \int_{s_1}^{s_2} p(s) \ln(x-s)^2 \, ds + \text{constant}$$

They developed a numerical method using the inverse solution of the Reynolds equation that found the pressure curve which satisfied both the elastic and hydrodynamic requirements of the problem. Their results showed the film shapes for different loads along with the corresponding pressure distributions. The latter did not show the characteristic

outlet pressure peak predicted by the Grubin solution(35). This was due to the relatively low modulus of elasticity and speed used in the calculation. In a later paper(39) by the same authors the pressure peak was always found to be present for more realistic materials, speeds, and loads.

The development of a rigorous iterative procedure was given by Archard, Gair, and Hirst(40) in 1961. They treated the iteration of the inlet and outlet regions separately by assuming that under high loads and low speeds the inlet and outlet pressures would not influence each other. Their results showed how the pressure distribution and film shapes were dependent on speed, load, pressure-viscosity, and elasticity.

The effects of lubricant compressibility was presented in 1962 by Dowson, Higginson, and Whitaker(41), and complete details of their solution were given later by Dowson and Whitaker(42). Some of the results for film shape and pressure distribution are shown in fig.26. The addition of compressibility reduces the central film thickness (h_o) and causes a slight increase in thickness just prior to the outlet constriction. The minimum film thickness (h_m) at the rear constriction remains essentially unchanged. The effect of compressibility on pressure distribution reduces the height of the pressure peak and displaces it downstream.

The generation and dissipation of heat within an ehd contact was considered by Archard(43) in 1959 and Crook(44) in 1962. They showed that conduction to the bearing surfaces was the most important heat dissipation mechanism. For pure rolling the film temperature rise is small and confined to

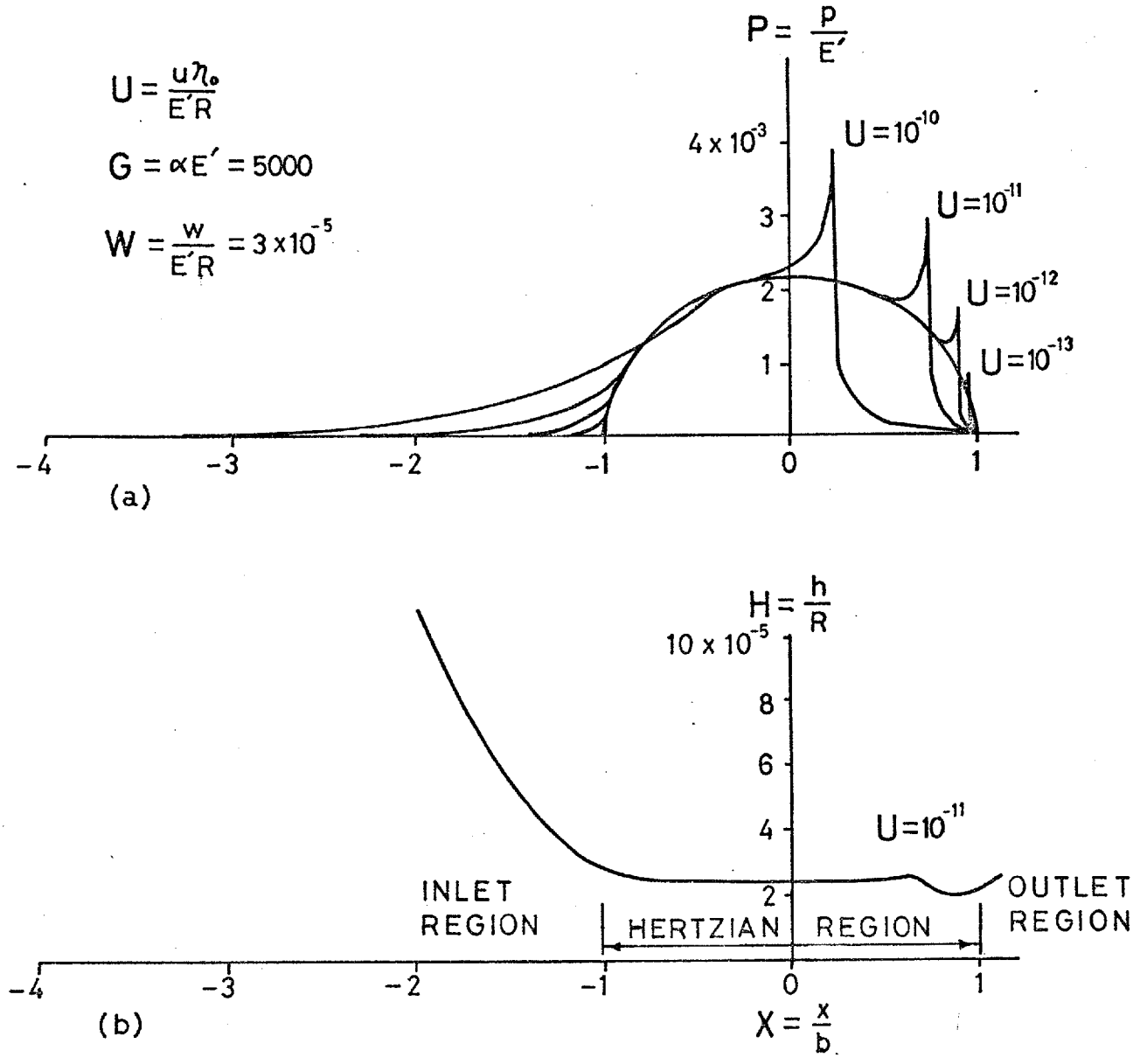


Fig.26 Theoretical solution showing: (a) pressure distribution, and (b) film shape for a compressible lubricant; (taken from Dowson, Higginson and Whitaker ref.41)

the inlet region. The introduction of sliding considerably increases the film temperature in the Hertzian zone where pressures, and hence viscosities, are high; however, the temperature rise in the inlet region remains essentially unchanged. Crook(45) found experimentally that there was very little reduction in film thickness with the introduction of sliding. This result underlines the view that film thickness is primarily determined in the inlet region where the temperature rise is small.

The inclusion of thermal effects in the ehd theory was presented by Cheng and Sternlicht(46) in 1964, and in 1965, Cheng(47) and Dowson and Whitaker(48). Their results supported earlier theoretical work and showed that the characteristic pressure spike still persisted when thermal effects were considered.

4.6 Film Thickness Formulae

In 1961 Dowson and Higginson(49) supplied a formula for minimum film thickness h_m (i.e. at the outlet constriction) which was determined from their previous isothermal incompressible solutions(38), (39) and (50). The formula in terms of their chosen dimensionless parameters is

$$H_m = 1.6 \frac{G^{0.6} U^{0.7}}{W^{0.13}} \qquad \ldots(4.9)$$

where

$H_m = \frac{h_m}{R}$ film thickness parameter

$G = \alpha E'$ materials parameter

$U = \frac{\eta_o u}{E'R}$ speed parameter

$W = \frac{w}{E'R}$ load parameter

This result agrees with the Grubin film thickness formula (equation 4.8) to within about 20% for most conditions.

There is some lack of agreement in connection with the number and choice of dimensionless parameters which most amply represent the isothermal ehd problem. The basic elastic and hydrodynamic equations require only three groups. Blok et al (51) presented a survey chart of theoretical solutions using three groups. Greenwood(52), recognizing the insensitivity of the materials modulus E', suggested another set of three groups.

The use of four groups provides a greater physical meaning. Dowson and Higginson's groups separate the effects of load and speed in their film thickness formula. Halling (53) suggests a film thickness formula with the following four groups:

$$\left(\frac{h_m}{R}\right), \left(\frac{\eta_o}{w}\right), \left(\frac{\alpha w}{R}\right) \quad \text{and} \quad \left(\frac{E'R}{w}\right)$$

The first two groups represent the classical Martin solution, and the addition of the last two groups account, respectively, for the effects of pressure-viscosity and elastic deformation. Archard(54) presents another set of dimensionless parameters which provide good physical meaning. These are based on the groups presented by Blok, but in terms of certain characteristic pressures and deformation which describe the character of the system.

It was pointed out that the indices of the Dowson and Higginson film thickness formula (equation 4.9) were not fully compatible with the requirements of the basic equations. A more appropriate formula suggested by Dowson(55) in 1967 took the form

$$H_m = 2.65 \frac{G^{0.54} U^{0.70}}{W^{0.13}} \qquad \ldots(4.10)$$

The relative influence of each variable on film thickness can be obtained by writing equations 4.9 and 4.10 in dimensional terms.

$$h_m = 1.6 \frac{\alpha^{.6} (\eta_o u)^{.7} (E')^{.03} R^{.43}}{w^{.13}} \qquad \ldots(4.11)$$

$$h_m = 2.65 \frac{\alpha^{.54} (\eta_o u)^{.7} (E')^{-.03} R^{.43}}{w^{.13}} \qquad \ldots(4.12)$$

It can be seen that the influence of load on film thickness is small, and that the exponent of the materials modulus is positive in equation 4.11 and negative in equation 4.12. The latter, which was pointed out by Halling(53), suggests that the film thickness is insensitive to changes in elastic constants. The important result, therefore, is that film thickness is primarily a function of α, η_o, u, and R.

4.7 Experimental Investigations

Many of the theoretical predictions presented in the previous section were verified experimentally. The coupling of theoretical and experimental results enhanced the confidence and understanding of the various fundamental aspects of ehd lubrication. Some of the significant experimental investigations are presented below.

a. Film thickness

Film thickness measurements which substantially agree with theory were presented by Crook(44, 45, 56-58) in an informative series of papers between 1958 and 1963. From his electrical capacitance measurements in disc machines he concluded that film thickness is primarily a function of the

rolling speed and the lubricant viscosity at the temperature of the discs. Moreover, the influence of load and sliding speed had little effect on film thickness.

In 1961 Sibley and Orcutt(59) presented film thickness results using the x-ray transmission technique. Their film thickness results compared favorably with the Grubin theory, and their general conclusions were in harmony with that of Crook.

Some of the best agreement between theoretical and experimental film thickness was obtained by Dyson, Naylor, and Wilson(60) in 1965. They used a very successful capacitance arrangement to measure film thickness in a disc machine lubricated with a variety of fluids. They confirmed the theoretical prediction that film thickness is primarily a function of α, η_o, and u (R was kept constant) and is insensitive to load and sliding. A summary of the theoretical and experimental findings of the dependance of the film thickness on α, η_o, u, R, and w was presented in 1965 by Archard(61).

b. Film shape

In 1961 Crook(62) converted his four-disc machine to measure film shape using a chromium electrode evaporated on a glass disc. The film shape, which was recorded on a cathode ray oscilloscope, showed the characteristic ehd shape predicted by theory, including the rear constriction and the slight rise in thickness just prior thereto. Further measurements, under moderate loads, were made by Cheng and Orcutt(63) in 1965 using a platinum strip transducer.

In 1967 Hamilton and Moore(64) presented a clever method of obtaining a more accurate film shape by recording

simultaneously the pressure and shape from a manganin transducer. The corrected dielectric constant, and hence shape, was obtained from the related pressure curve. The film shape presented also showed the slight increase in thickness just prior to the outlet constriction.

A complete contour map of the film shape for line contact was given by Gohar and Cameron(18) in 1966 for quite moderate loads using interferometry. The most important result was that the minimum film thickness occurred at the sides of the contact rather than at the rear constriction. They also showed that the severity of the side closure could be relieved by suitable blending of the roller end. These results have important practical implications and point out the need to modify the minimum film thickness equations presented so far.

c. Pressure distribution

Detailed pressure measurement in a highly loaded contact is an arduous task because of the large pressure changes which occur over very small periods in time and short distances in space. Pressure measurements generally fall into two basic approaches. The first approach is to increase the Hertzian contact width by the use of conforming surfaces or soft materials. Conventional pressure measurements are then obtained from the stationary member of a sliding contact. Higginson(65) reported pressure measurements between a bronze disc and a rubber block in 1962 while Dowson and Longfield(66, 67) used conforming bronze and steel surface combinations in 1963 and 1964. Longfield(68) presented further measurements, but under higher loads, in 1965. The

results of these experiments showed how the pressure distribution changed from the classical pressure profile predicted by the Martin theory to the more Hertzian profile as the load was increased. Notable departures from the Hertzian pressure profile for the highly loaded cases were a steepening of the outlet pressure curve and a slightly suggestive bump just prior thereto(68). The predicted outlet pressure peak was not observed nor was it likely to be present under the given experimental conditions.

Measurements of the pressure distributions along the axial as well as the circumferential direction were reported by Dowson and Longfield(67) and Niemann and Gartner(69). Their results indicated that the pressure remained relatively constant over the contact width and that the side pressure decay was concentrated very near the edges. The sustained pressure near the edges of the contact supported the view that side leakage is of little importance in ehd contacts.

A second approach to pressure measurements deals with more realistic contact sizes in rolling. In the contributions by Kannel, Bell, and Allen(70), Cheng and Orcutt(63) in 1965, and Hamilton and Moore(64) in 1967, direct pressure measurements were obtained from a small manganin transducer located on the surface of one of the rolling elements. The pressure profiles presented by these authors were obtained under conditions more likely to produce the pressure peak. Some of the results, notably those of Kannel(71) and Hamilton and Moore(64), show some evidence of the pressure peak. This appears to be the only aspect of the theoretical pressure curves which has not been fully verified by experiment. Moreover, it is suggested by Fein(72) that compressional visco-

elastic effects may reduce or eliminate the pressure peak.

d. Temperature distribution

Most of the heat generated within the lubricant film is dissipated through the bearing surfaces. Therefore, the measurement of surface temperatures should reflect the heat generation within the film. Dowson and Longfield(66, 67) recorded temperatures on the stationary member of a sliding contact with thermocouples. The surface temperature was found to rise in the inlet region and fall somewhat at the outlet. Also, the overall temperature increased with sliding speed and load.

In 1965 Orcutt(73) and Cheng and Orcutt(63) presented temperature measurements in a rolling contact using a platinum transducer attached to a glass disc. Other measurements are reported by Kannel et al(74) in 1966 using a titanium wire, and Hamilton and Moore(64) in 1967 using a nickel guage. For pure rolling the temperature profiles show a rise in the inlet region which reaches a maximum approximately at the Hertzian zone. The temperature falls within the Hertzian zone because of the low rate of shear and the large amount of heat conduction to the bearing surfaces; but it again rises near the outlet region due to viscous heating. Under sliding conditions the inlet temperature rise remains essentially unchanged; but it increases considerably within the Hertzian zone where high pressures and viscosities are present.

The most important temperature, as far as film thickness is concerned, is that which governs the viscosity in the inlet region. Because of the large amount of heat con-

duction, Crook(45) found the controlling temperature to be that of the bearing surfaces. The surface temperature results of Cheng and Orcutt(63) for this region are of significance here because of their glass/steel materials combination. They found that the maximum inlet temperature increased with load and speed, and their recorded temperature rises ranged from 1 to 15^{o}F.

4.8 Point Contact

a. Experimental

The possibility of an effective ehd film between highly loaded point or elliptical contacts was established experimentally by Archard and Kirk(75) in 1961. They obtained film thickness measurements between rotating cylinders at right angles using capacitance. Their results, taken over a wide range of conditions, indicated that the ehd lubrication of point contact has basic features in common with line contact. Also, using cylinder materials of different elastic modulii(13) they showed experimentally that film thickness was insensitive to small changes in elastic modulus. Large changes in elastic modulus influenced not so much the film thickness as the form of lubrication. That is, for low elastic materials pressure-viscosity effects play a far less important role than elastic effects.

Most of the experimental information on the ehd lubrication of point contact was brought about by the use of optical interferometry which provided detailed information on both film thickness and shape. Initial work by Kirk(12) and

Archard and Kirk(13) using white light interference fringes between crossed cylinders of perspex showed a variation of film thickness across the Hertzian region with a notable thinning near the lateral edges. Using more realistic bearing materials of glass and steel in a sliding contact, Cameron and Gohar(15) in 1966 showed clearly the characteristic horse-shoe shaped constriction along the rear and lateral edges of the Hertzian region. A wide range of other materials combinations such as sapphire/steel and diamond/steel used by Gohar and Cameron(18) and rubber/glass used by Roberts and Tabor(76) also showed various forms of the horse-shoe shaped constriction.

The influence of load and speed on film thickness and shape was presented by Gohar and Cameron(18) in sliding. Foord et al(2,16), using an improved interferometric system, presented more elegant results in rolling. The general features of these results are: firstly, the minimum film thickness occurs at the lateral edges of the Hertzian region except for very low loads and/or high speeds; and secondly, the film thickness at the lateral edges is much more sensitive to changes in load than at any other location within the Hertzian region.

From his experimental results, Foord has recently provided an empirical formula for the film thickness (h_o) at the centre of the Hertzian region(21).

$$\frac{h_o}{R} = .86 \left(\frac{\eta_o u}{E'R}\right)^{.667} (\alpha E')^{0.6} \left(\frac{W}{E'R^2}\right)^{-.05}$$

When this is written in dimensional terms

$$h_o = .86 \frac{\alpha^{.6} (\eta_o u)^{.667} (E')^{-.017} R^{.433}}{W^{0.5}} \quad \ldots(4.13)$$

we see that the dependence of film thickness (at least for the central region) on the physical operating variables is remarkably similar to the line contact theory of Dowson (see equation 4.12).

b. Theoretical

The fundamental development of the ehd lubrication of point contact differs from line contact in that theoretical solutions have not kept abreast of experimental investigations. The difficulty lies in the complications resulting from the additional geometric dimension in the point contact problem. Consequently, the existing theoretical solutions are approximate in nature and generally follow the basic assumptions inherent in the Grubin solution for the line contact problem. The solutions, therefore, assume Hertzian conditions and deal with the thickness of a parallel film over the Hertzian region.

In 1965 Archard and Cowking(77) presented a solution which treated the point contact as an assembly of elemental line contacts modified by a side-leakage factor derived from classical theory. Their solution applies to elliptical as well as point contacts and has recently been extended by Snidle and Archard(78) to cover velocity directions which are not along either axis of the ellipse.

Cameron and Gohar(15) approached the problem more directly in 1966 by employing the two-dimensional Reynolds equation and solving a Grubin-type analysis in the inlet region. Although their approach is tenable, their solution was somewhat unrefined. A much more thorough solution has recently been computed by Evans(79), and a similar solution

using the same approach for elliptical contacts was presented by Cheng(80) in 1969.

The above solutions for film thickness can be presented in the following form

$$\frac{h_o}{R} = K\left(\frac{\alpha\eta_o u}{R}\right)^a \left(\frac{w}{E'R^2}\right)^b .$$

where the values of K, a, and b as determined by the various theories are present in table 1.

TABLE 1

Constants for Various Film Thickness Theories

Theories	K	a	b
Archard & Cowking	1.40	0.74	-0.074
Cameron & Gohar	3	1.0	-0.333
Evans	1.73	0.714	-0.048
Cheng	1.69	0.725	-0.058

Writing in dimensional form, using, for example, the solution of Evans, we get

$$h_o = 1.73 \frac{(\alpha\eta_o u)^{.714} (E')^{.048} R^{.382}}{w^{.048}}$$

The general dependance of film thickness on the physical variables is similar to that found experimentally by Foord (see equation 4.13). The important result, therefore, is that the film thickness for ehd point contact, like line contact, is primarily a function of α, η_o, u and R.

c. Similarities between the ehd lubrication of line and point contact

The results of Archard and Cowking(77) and Cheng(80) for elliptical contacts facilitates the comparison of film thickness between line and point contacts under comparable

conditions of stress. The theory of Archard and Cowking(77) predicts a film thickness for point contact which is about 31% lower than line contact, while their experimental results suggest a slightly greater reduction of 38%. Cheng's theoretical results show a more optimistic reduction of only 18% for his chosen operating conditions.

The distinguishing feature of point contact is the existence of side-leakage in the inlet region; however, as seen from above, its effect on film thickness is smaller than might be expected. Furthermore, the dependence of the central film thickness on the physical operating variables for point contact is essentially the same for line contact. This emphasizes the similarities which must exist between the ehd features of line and point contact.

4.9 Other Aspects of Ehd

a. Normal approach

In 1962 Christensen(81) presented an ehd analysis of the normal approach of cylinders and the normal approach of spheres in a later analysis(82). Direct experimental evidence of the latter using optical interferometry was given by Dowson and Jones(83, 84), Foord, et al(16), and Westlake and Cameron(20, 22). Under suitable ehd conditions a pressure higher than the Hertzian pressure can be generated between the approaching bodies. The effect on elastic deformation is to reverse the curvature in the high pressure region causing a fluid entrapment when the bodies touch.

b. Friction

Rolling friction in ehd contacts is primarily

associated with the viscous shear in the convergent inlet region, whereas sliding friction is primarily the result of viscous shear in the high pressure Hertzian region. Crook (57) has shown that rolling friction is proportional to film thickness and is dominated by sliding friction at all but the smallest sliding speeds. Experimental work by Smith(85) and Crook(57) showed that sliding friction increases to a maximum with slip speed and then decreases with further increase in slip.

Friction measurements are more easily obtained than interpreted. Crook(44, 57) explains the general features of his results in terms of temperature and viscosity changes within the high pressure Hertzian region, while Smith(85) suggests that the lubricant behaves like a plastic solid which can shear across the centre of the film when a critical stress is reached. More recent experimental results by Johnson and R. Cameron(86) and Plint(87, 88) show that under high contact pressures the coefficient of friction reaches a limiting value. Under these conditions the authors suggest that the critical stress hypothesis of Smith might be more relevant.

c. Rheology

A significant feature of the above friction results as well as those of Bell, Kannel, and Allen(89) is that frictional behaviour cannot entirely be explained in terms of temperature effects on a Newtonian fluid, and that rheological effects which are different from Newtonian must also be included. In 1965 Dyson(90) considered the continuous shear results of Crook and Smith in comparison with the

oscillating shear results of Barlow and Lamb(91). He based his comparison on the theory of Oldroyd(92) and concluded that the non-Newtonian behaviour of mineral oils can be explained in terms of viscoelasticity.

Non-Newtonian behaviour of mineral oils appears to manifest itself more in terms of the 'effective' viscosity in sliding friction than in film thckness. For example, film thickness measurements by Dyson et al(60) over a wide range of operating conditions show excellent agreement with Newtonian theory except for a slight, but significant, departure at high speeds and viscosities where measurements fall below theoretical predictions.

Rheological behaviour is linked with fluid structure, and it is possible that more complicated structures than straight mineral oils will reflect non-Newtonian behaviour in film thickness measurements. Archard and Kirk(93) in 1964 and Dyson and Wilson(94) in 1966 found that silicone fluids give a much lower film thickness than predicted by theory. Furthermore, Hamilton and Robertson(95) in 1967 and Foord et al(16) in 1968 found that polymer-blended oils are less effective in generating an ehd film than expected for a Newtonian fluid.

4.10 Application of Ehd Fundamentals

According to Tallian(96) the lubrication of ehd contacts is undertaken for two reasons: to control friction and to minimize failure. Both friction and failure are intimately related to film thickness. For pure rolling contact applications, control of failure is more important than friction

except for certain instrument ball bearings where friction requirements are very stringent.

The significance of the ehd oil film in relation to practical applications was shown by Dawson (97, 98, 99) in his work on pitting failure in a disc machine. He found a definite correlation between the number of cycles to pitting and the dimensionless ratio D, defined as the total initial surface roughness divided by the oil film thickness. Furthermore, case studies on spur and helical gears by Dudley(100) showed that wear rate was definitely related to the calculated film thickness. These correlations imply that surface damage is related to the amount of intermetallic contact between the mating surfaces. This underlines the importance of a 'mixed' or 'partial' ehd regime where the load is shared between ehd forces and intermetallic contact of asperities. Experimental and theoretical aspects of this regime have been presented by Tallian et al (101, 102) and Christensen(103).

An interesting feature of one of Dawson's tests(97) using a calculated film thickness about seven times greater than the total surface roughness was that pitting failure was concentrated along the extreme edges of the track. This was also shown by other workers, e.g. Kelley and Lemanski(104), and it indicates the importance of the minimum film thickness which occurs at the lateral edges of the Hertzian zone as shown by Gohar and Cameron(18). Another test reported by Kelley and Lemanski, where the edges of the track were rounded and the calculated film thickness was much greater than the total surface roughness, showed no pitting after 10^7 revolutions. Thus, the establishment of a full ehd film is essential for reducing or eliminating pitting failure. An experiment

by Garnell and Higginson(105) using electrical resistance in a roller bearing showed that a full ehd film was established when the calculated film thickness was about three times the combined c.l.a. roughness of the rolling surfaces.

Ehd lubrication has developed to a state wherein the basic fundamentals are quite well established and are now being exploited in practical situations. Dowson and Higginson(106, 107), for example, have applied the ehd analysis to roller-bearing and gear lubrication. It remains that one of the most important contributions which ehd research can make to practical applications is the prediction of film thickness.

4.11 Importance of Inlet Region and Lubricant Supply

An important aspect inherent in the ehd problem which was brought out in the Grubin theory(35) is that film thickness is primarily developed in the inlet region just in advance of the flat Hertzian region. The experimental work of Crook(56) and the thermal solutions of Cheng and Sternlicht (46) and Dowson and Whitaker(48) provide further evidence of the fact that film thickness is governed by the conditions in the inlet region. Therefore, the conditions of speed, viscosity, and geometry, as far as film thickness is concerned, must be those relevant to the inlet region. These are the controlling variables in the film thickness equations.

A condition of the inlet region which is not represented in the film thickness equations is that of the supply of lubricant. Hydrodynamic considerations require the gap between the surfaces to be adequately filled to facilitate the

important pressure build-up. If the inlet region is adequately filled with lubricant, the inlet pressure build-up will develop a film thickness which is insensitive to further increases in lubricant supply. This condition of lubrication is defined as 'flooded'. If the inlet region is inadequately filled, the inlet pressure build-up will be delayed, and a film thickness which is dependent upon the available lubricant supply will result. The condition of lubrication is then defined as 'starved'. This starved condition of lubrication has received very little attention both theoretically and experimentally even though its influence on film thickness in many practical applications is overriding.

4.12 Previous Work Relating to Lubricant Starvation

a. Experimental

Experimental work on this subject can more appropriately be classified as interesting examples of, rather than investigations of, lubricant starvation. An example of lubricant starvation in gears was reported by Howell(108) in 1967. Using electrical resistance measurements as an indication of film thickness, he found that film resistance decreased when the supply rate was reduced below a critical value. In 1950 Jones and Wilcock(109) supplied a series of ball bearings with minute quantities of oil. They found that the running time to failure was a direct function of the amount of lubricant initially supplied, and that failure appeared to occur when the initial lubricant supply had depleted to a certain critical value. A continuation of this work was presented by Booser and Wilcock(110) in 1953. They

presented a quantitative study of the effects of load, speed, and lubricant properties on the amount of lubricant required. Their results indicated that bearing life is intimately related to the mechanism by which the oil is lost from the bearing. They therefore correlated their results in terms of lubricant creep properties, i.e. viscosity and surface tension, as well as bearing speed and size.

Experimental investigations on the performance of instrument bearings have provided some important physical aspects relating to lubricant starvation. These bearings are typically run with a minimum supply of lubricant. One reason for this involves the large torque required to expel excess lubricant from the bearing raceways. In a review on experimental observations of angular contact instrument ball bearings, Kingsbury(111) described an experiment on the effects of bearing preload and lubricant supply on the 'basic speed ratio' (i.e. the amount of slip). It was noted that under constant preload the amount of slip decreased as the lubricant supply was reduced. From this behaviour (which among other things is related to film thickness) it was suggested that the typical rolling contact in an instrument bearing is operating in a starved condition.

More substantial evidence of this starved condition and its importance was presented in 1963 by Horsch(112) in connection with the performance of gyro spin-axis ball bearings. From capacitance and inductance measurements of dimensional changes within the gyro bearing he estimated a film thickness on the order of 10 micro-inches for normal operating conditions and lubricant supply. When additional lubricant was added, however, he estimated a film thickness of at least 20 micro-inches. Having thus substantiated the 'normally starved'

operating condition, he was then able to explain some of the causes of gyro instabilitiies in terms of film thickness changes. These film thickness changes are caused by the centrifugal release of an oil droplet onto the bearing raceway which then presents an abrupt increase in film thickness followed by a slow exponential return to the original film thickness. Accompanying this behaviour is a positional instability of the gyro wheel called 'oil jag' or 'jog', which can seriously effect the navigational accuracy of the instrument. Reducing the oil supply to smaller amounts will lessen the probability of oil jag but increase the probability of bearing failure. Failure is not from fatigue, but is due to the transformation of the lubricant into a varnish. Horsch(112) has shown that the rate of varnish formation is a function of film thickness, and since the bearing is operating in a starved condition it is intimately related to oil supply.

According to Freeman(113) the positional stability of the gyro wheel must be within 0.2 micro-inches in order to maintain the required navigational accuracy. This requirement, which is only a fraction of the total ehd film thickness, indicates thd importance of a stable ehd film which must be maintained under starved conditions. There is very little experimental evidence relating to the thickness of a starved ehd film and how this film is effected by various operating conditions.

b. Theoretical

Theoretical investigations of lubricant starvation under ehd conditions are indeed lacking. Significant solu-

tions, however, have been obtained for hydrodynamic conditions, i.e. for rigid surfaces. In general, these solutions are obtained by shrinking the length of the load carrying film. This is done by reducing the inlet position where $p = 0$ and specifying the outlet boundary condition to be $p = \frac{dp}{dx} = 0$.

Early solutions are reported by Büche in 1934 and Peppler in 1938 (see ref. 114). Other solutions are given by Sassenfeld and Walther (see ref. 30, pp. 177-179), Floberg (see ref. 30, pp. 175-177), Dowson and Whitaker(42), and Boness(115). These solutions were mainly concerned with the calculation of load capacity for various inlet boundary positions. The results show that the effect of inlet position on load capacity is not very significant unless the inlet position is very close to the line of centres. Boness(116), (117) and Dowson and Whomes(118) also included the effect of side leakage. Their solutions show that for severely starved conditions, particularly for thin films, the effect of the inlet boundary position is more important than side leakage.

In 1967 Fein and Kreutz(119) considered a brief and idealized approach to the starvation of a single hemispherical asperity model under ehd conditions. Their asperity model is shown in fig.27 where it is assumed that the lubricant is transported to the asperity as a film of thickness d on the moving surface. An asperity film thickness (h_o) under steady state starved conditions was obtained by considering a material transport balance, i.e.

$$ud = u \frac{h_o}{2}$$

and, therefore, $h_o = 2d$. For this to be true it must be

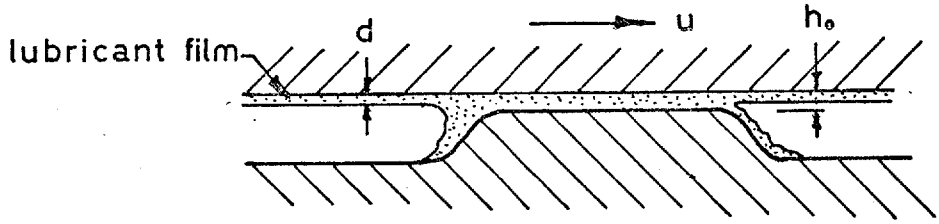


Fig.27 Starved lubrication in a hemispherical asperity model (taken from Fein and Kreutz, ref.119)

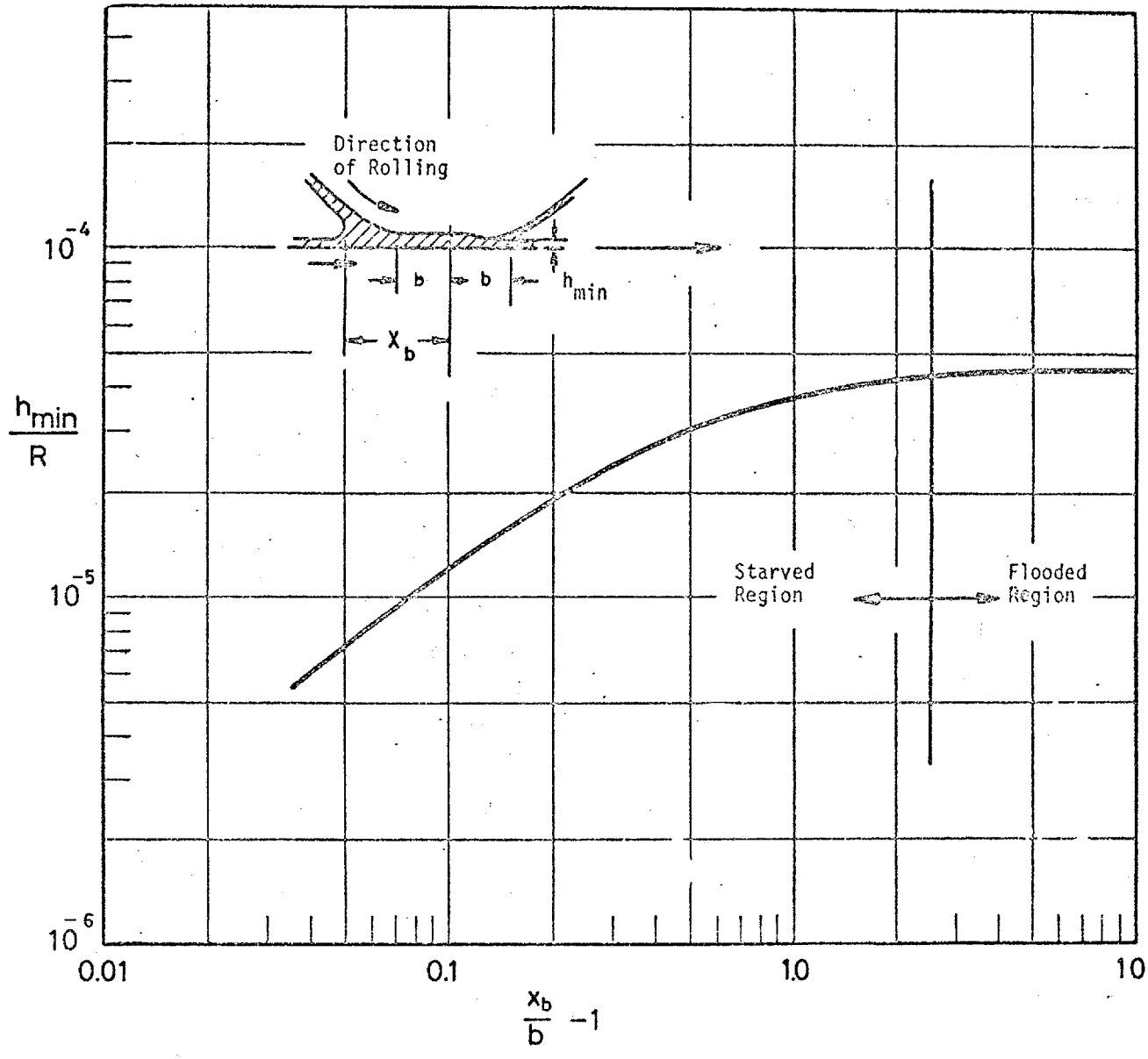


Fig.28 Effect of starved lubrication on film thickness for line contact (taken from Orcutt and Cheng, ref.120)

assumed that there is an initial pile-up of lubricant in the inlet region which remains as sliding is continued; i.e. there can be no side leakage. The most important feature of their analysis, however, is a statement suggesting that the vast majority of pressure buildup for an ehd contact occurs within about

$$\frac{3(Rh_o)^{\frac{2}{3}}}{a^{\frac{1}{3}}}$$

ahead of the Hertzian flat. Thus, to achieve a flooded condition, the gap between the surfaces must be filled to this distance. The derivation of this criteria was not given.

In 1965 Orcutt and Cheng(120) presented theoretically the effect of lubricant supply rate for the ehd line contact problem by specifying the location within the inlet region where the pressure begins to rise. This was done for only one speed(250 in./sec.) and Hertzian stress (120,000 psi) to illustrate the effect of the inlet lubricant boundary on film thickness. No other details of the solution are given. Their result is reproduced in fig.28, which shows that for the particular operating conditions the distance (x_b) from the line of centres to the inlet lubricant boundary must be at least 3.5 times the Hertzian half-width (b) to achieve a flooded condition. Presently, Dowson and Castle(121) are working on a similar theoretical investigation of starvation in the ehd line contact problem. No results are available at the time of writing.

* * * * *

The basic fundamentals of ehd lubrication have been presented in this chapter. It has been shown that ehd lubrication has now reached a state where these fundamentals are being applied to practical applications, and that the prediction of film thickness is a very important contribution. It has also been shown that for some applications the supply of lubricant in the inlet region plays a dominant role in determining film thickness, and that the performance of the particular application can be strongly dependant upon the absolute thickness of this film. Herein lies the importance of a knowledge of the ehd behaviour under these starved conditions.

Chapter 5 EXPERIMENTAL PRELIMINARIES

5.1 Description of Test Lubricants

The kinematic viscosity and specific gravity data of the test lubricants are shown in table 2.

TABLE 2

Viscosity and Specitic Gravity of Test Lubricants

LUBRICANT SYMBOL	MANUFACTURERS' REF. NO.	SPECIFIC GRAVITY+	VISCOSITY, CS. 75°F	100°F	140°F	210°F
SAE 10	BP L60/1591	.870	66.8	33.9	15.1	5.2*
SAE 30	BP L65/420	.883	261	113	39.0	11.43*
SAE 40	BP L65/557	.885	401	165	54.4	14.8*
TN 631	SHELL TN631/68	.901		137.5*		10.20*
LUB D	SHELL LUB D	.942	572	182	46.2	10.84*

+ MEASURED AT ABOUT 70°F

* MANUFACTURERS' DATA

The viscosities were measured with a suspended level viscometer and plotted on an ASTM standard viscosity-temperature chart, shown in fig.29. The lubricants designated SAE 10, SAE 30, and SAE 40 are paraffinic mineral oils, while TN 631 is a medium viscosity index naphthenic mineral oil. LUB D is a high viscosity, low viscosity index mineral oil (predominantly aromatic and naphthenic) which was used previously by Dyson et al(60) and Foord et al(16). It is used here as a reference fluid from which the pressure-viscosity coefficient of the other oils are obtained. This is the subject of section 5.4.

5.2 Determination of Refractive Index

To obtain the absolute thickness corresponding to each fringe the calibrated fringes must be divided by the

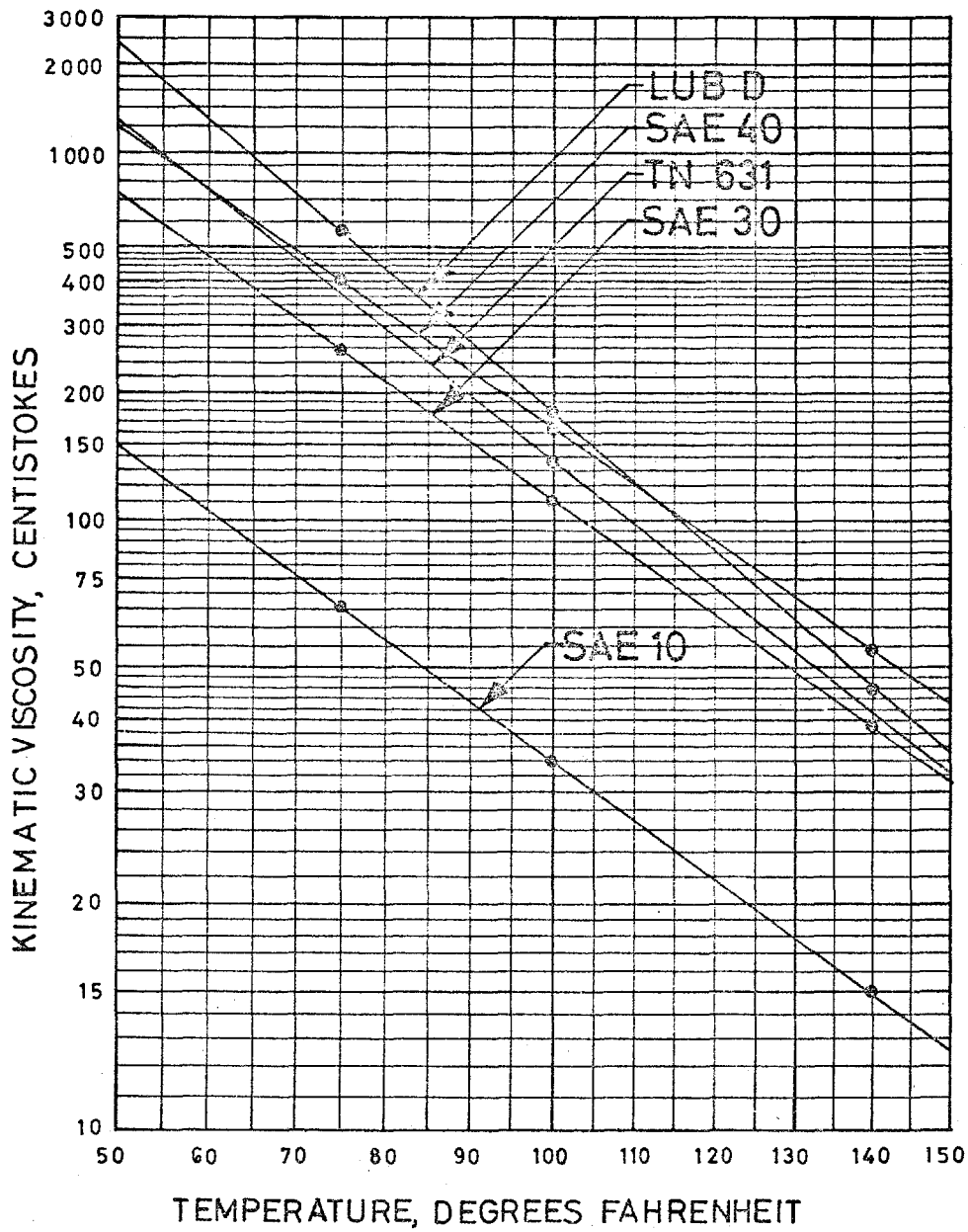


Fig.29 Viscosity-temperature characteristics of test lubricants

refractive index of the medium. The refractive index (n) changes with density (ρ) according to the Lorenz-Lorentz relation(122).

$$\left(\frac{n^2 - 1}{n^2 + 2}\right) \cdot \frac{1}{\rho} = \text{constant}$$

and can be calculated as follows:

$$n = \left(\frac{1+2A}{1-A}\right)^{\frac{1}{2}}, \text{ where } A = \frac{\rho}{\rho_o} \cdot \left(\frac{n_o^2 - 1}{n_o^2 - 2}\right)$$

and n_o and ρ_o are the refractive index and density at atmospheric pressure. Poulter, Ritchey and Benz(123) found that the Lorenz-Lorentz relation was correct to within about 0.6% for a paraffinic mineral oil subjected to pressures between 0 and 200,000 psi. This gives a refractive index error within about 1.5%. The density ratio ρ/ρ_o can be obtained from Hartung's empirical formula for Hydrocarbons, see Chu and Cameron(124)

$$\frac{\rho}{\rho_o} = 1 + \frac{42.8 \times 10^{-6}}{(\eta_{100})^{.0385}} \, p^{.75}$$

where η_{100} is the viscosity in centistokes at 100°F. This relation, which is applicable for pressures greater than 10,000 psi, was compared with the density measurements of a paraffinic mineral oil (32-G) described in the 1953 Pressure-Viscosity Report (37) and the density measurements of LUB D given by Galvin et al (125). The results, which are shown in fig.30, correlate to within 2%. A density error of this magnitude gives a refractive index error in the Lorenz-Lorentz relation of less than 1%.

The refractive index at atmospheric pressure (n_o) was measured for each test lubricant with an Abbé Refractometer. These are shown in fig.30 along with the variation of refractive index with the pressure as computed by the Hartung and Lorenz-Lorentz relations. For the highest test

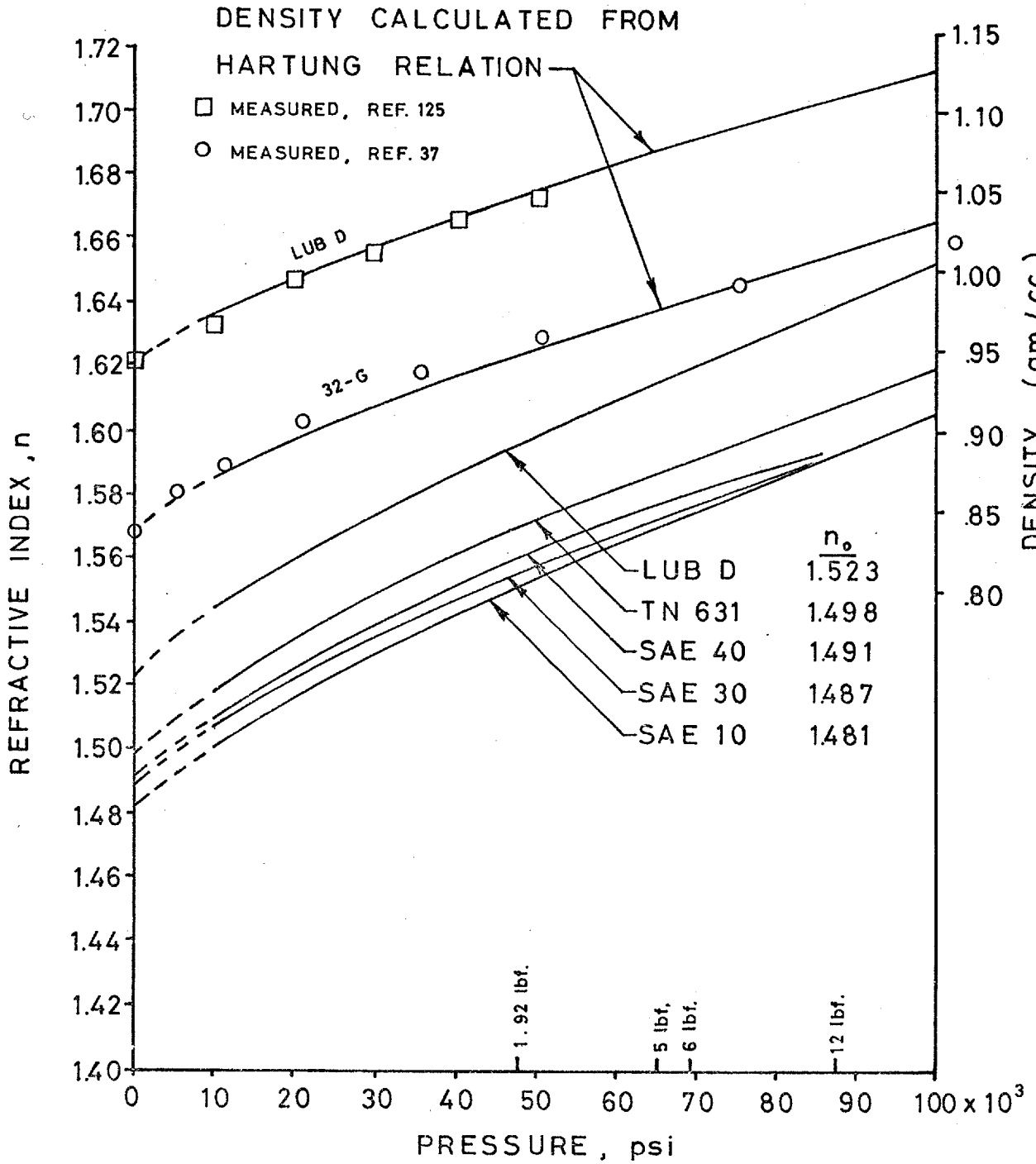


Fig.30 Variation of density and refractive index with pressure

pressure of 87,000 psi (12 lbf.) the density and refractive index increase on the order of 18% and 8% respectively.

To check the error due to refractive index dispersion (i.e. the variation of refractive index with wavelength) the refractive index of one of the test lubricants was measured using the green and red wavebands shown previously in figs. 14b and d as light sources for the Abbé Refractometer. The refractive index dispersion between these two wavebands was found to be less than 0.1%.

From the cumulative errors associated with the Hartung and Lorenz-Lorentz relations it is assumed that the refractive index can be determined to within about 2 or 3% for a given pressure. This gives a film thickness error of the same magnitude.

5.3 Film Shape and Pressure Distribution

A typical high-speed photomicrograph using the duochromatic interferometric system is shown in fig.31. The duochromatic system shows good detail in film shape, and it is clearly seen that for the given operating conditions, the minimum film thickness occurs at the lateral constrictions. In addition, the use of high-speed photography reveals the cavitation pattern in the outlet region as well as local surface defects within the Hertzian region.

a. Conditions along the centre line in the rolling direction

The film thickness along the centre line in the direction of rolling is plotted in fig.32. The measured shape is corrected for the variation of refractive index with pressure from an estimated pressure distribution.

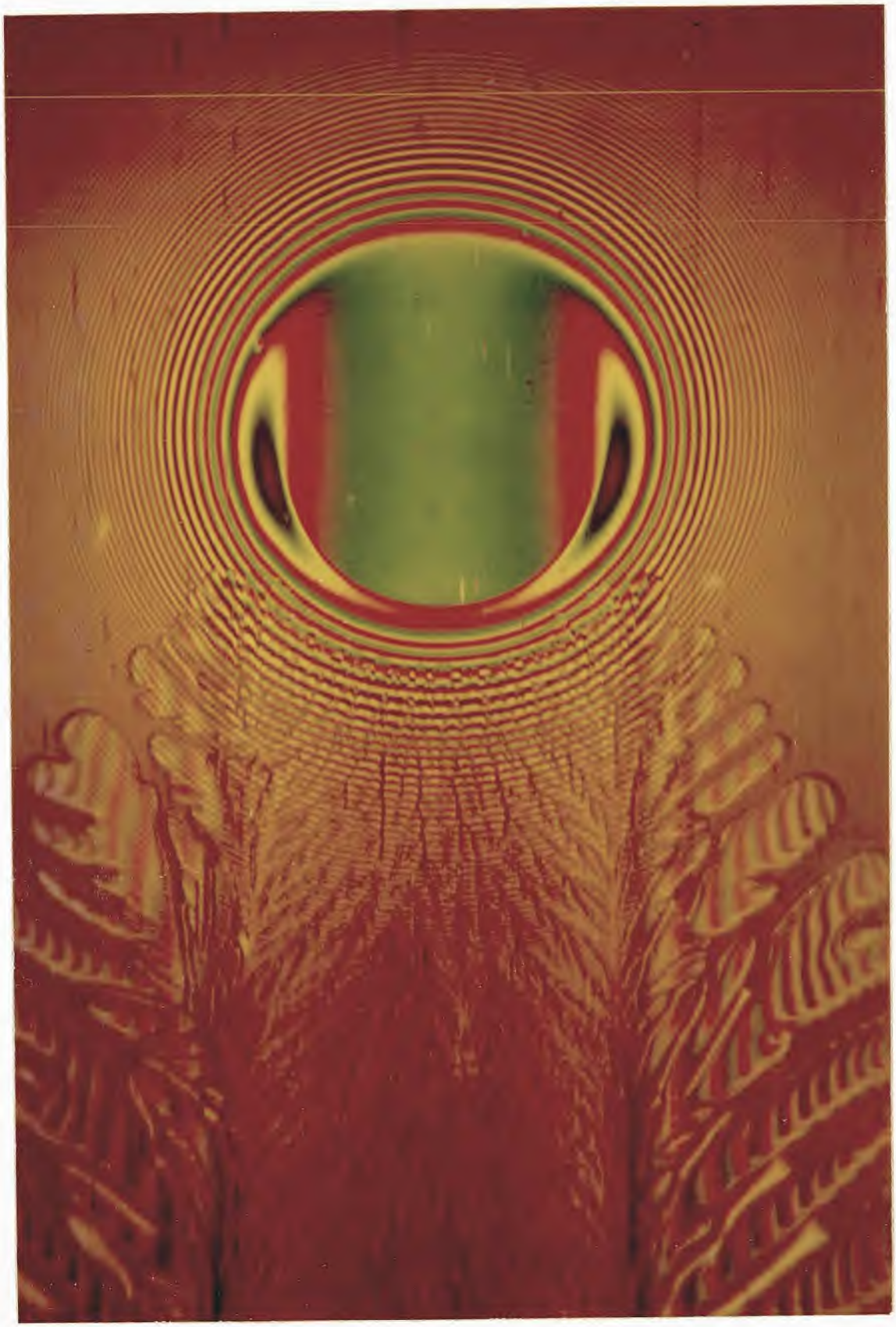


Fig.31 High-speed photomicrograph taken with the duochromatic interferometric system; u = 2.74 in./sec., R = 0.5 in., w = 12 lbf., p_{max} = 87,000 psi, LUB D, h_o = 18.3 micro-in., h_m = 8.5 micro-in.

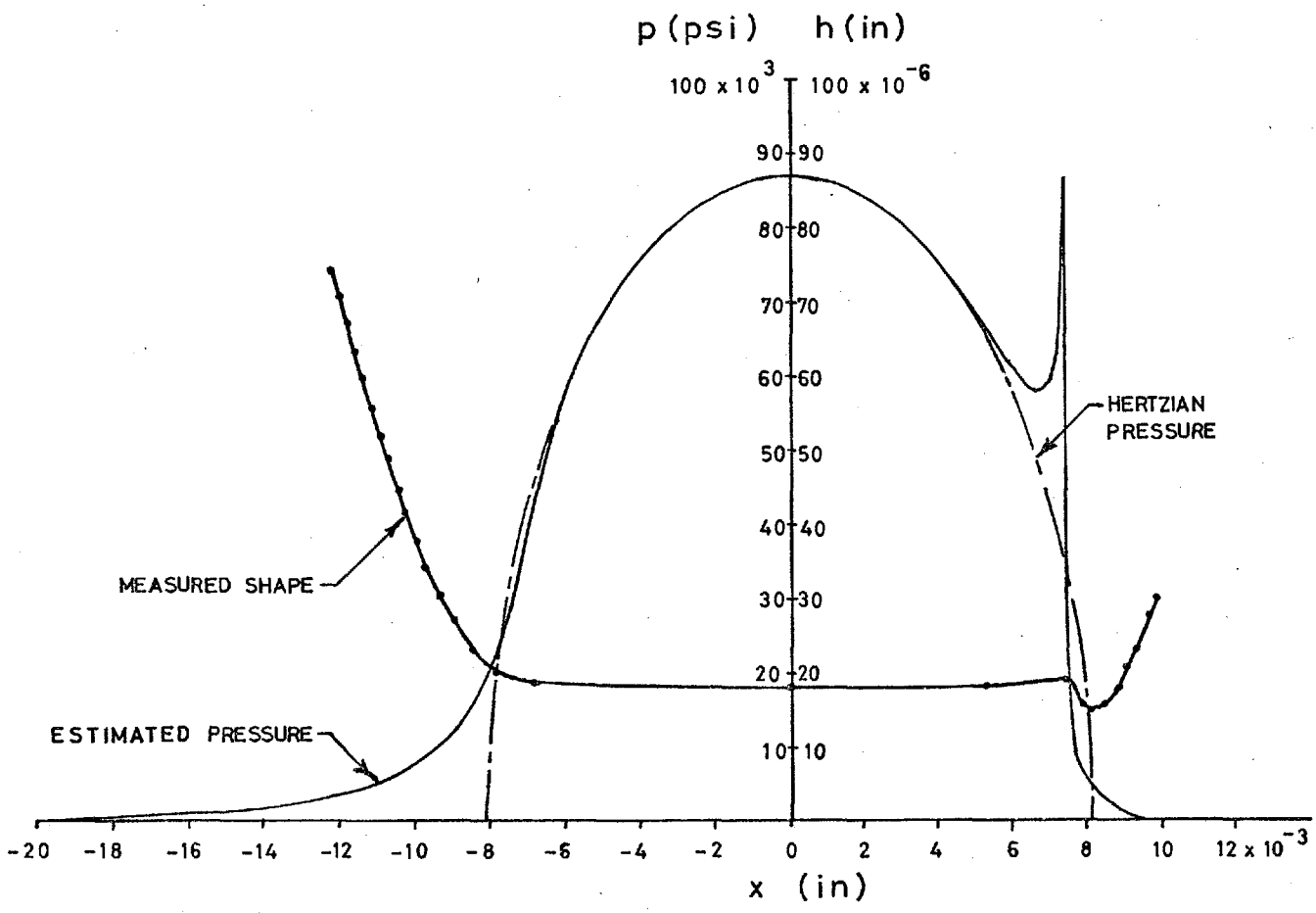


Fig.32 Measured film shape and estimated pressure distribution along the centre line in the direction of rolling

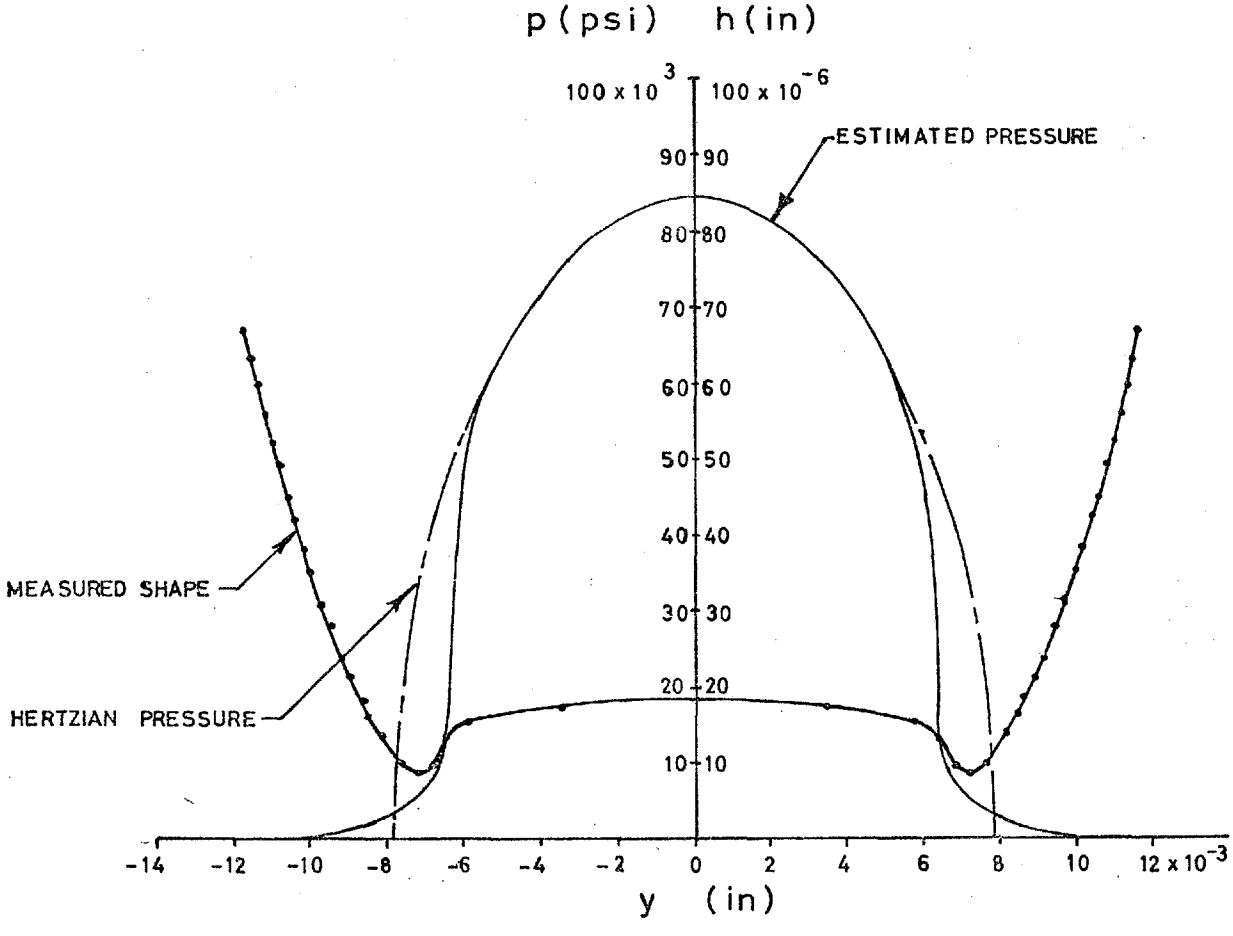


Fig.33 Measured film shape and estimated pressure distribution in the transverse direction at the location of minimum film thickness

The general features of this pressure distribution were deduced from the following considerations. The film thickness is very nearly constant over the centre of the Hertzian region; therefore, the pressure over this region must be very similar to the Hertzian pressure for static conditions. The termination of the pressure profile is clearly defined by the location of film rupture (where it is assumed that $p = \frac{dp}{dx} = 0$ in the outlet region; and the location within the inlet region where the pressure commences was determined from the experimental data on starvation (to be given in Chapter 6). This is the location, under the given operating conditions, at which the film thickness becomes sensitive to the inlet lubricant boundary. It can be shown that the deformation at the front edge of the Hertzian region is greater than the corresponding Hertzian deformation for static conditions. This is due to the ehd pressure generated in the inlet region. This pressure must blend in with the Hertzian pressure to establish a parallel film in the Hertzian region.

Near the back edge of the Hertzian region the film thickness drops very suddenly. This implies a very sharp pressure gradient. Furthermore, the relaxation of the surfaces into a constriction must mean that the pressure in this region is very low, much lower than the corresponding Hertzian pressure. Despite the low pressure at the rear constriction, the film thickness remains nearly constant in the region just prior to the constriction. Indeed, it can be seen from the photomicrograph that the green fringe is getting lighter in colour in this region indicating a slight increase in film

thickness. To maintain this near parallel film the pressure just in front of the constriction must be greater than the corresponding Hertzian pressure. It is this pressure that comprises the pressure peak predicted by theory.

An initial pressure distribution was drawn in light of the above considerations. Surface displacements were then computed for a line load on a semi-infinite flat solid using the pressure differences between the above pressure distribution and the Hertzian pressure distribution. The computed elastic displacements were then added to the Hertzian shape, thus giving a theoretical shape which could be compared with the measured shape. The computer programme was adapted from one used by Klemz(126) for line contact, and the method of calculating the surface displacements is described by Dowson and Higginson(38). It was assumed that the pressure distribution in the centre of the contact region is Hertzian. Thus, the problem consisted of modifying only the inlet and outlet edges of the pressure curve to make the experimental and theoretical shapes match. The inlet film shape was found to be far easier to match than the outlet film shape where the pressure gradients are very large. The resulting pressure distribution in the outlet region is, therefore, not expected to be very accurate.

The validity of calculating surface displacements from a line load on a semi-infinite flat solid and applying them to a point contact situation is clearly open to question. Therefore. the calculated displacements which are added to the Hertzian shape may not be numerically accurate. The computation was undertaken mainly to provide a qualitative indication of how the various changes in the pressure distribution would effect the elastic shape. The results of several trials

suggested the pressure distribution shown in fig.32. The inlet pressure distribution, which should be reasonably accurate, is used in later calculations.

b. Conditions along the transverse direction

The film shape along the transverse direction, taken at the location ($x = 2.06 \times 10^{-3}$ inch) of minimum film thickness, is plotted in fig.33. The measured shape has been corrected for the variation of refractive index with pressure from an estimated pressure distribution which is also shown in the figure.

The pressure distribution along the transverse direction is likely to be similar to the Hertzian pressure in the centre of the Hertzian region since the film thickness in this region is relatively constant. The pressure associated with the lateral constrictions, which are located within the Hertzian region, must be much lower than the corresponding Hertzian pressure. An important difference between the rear and lateral constrictions is that the slope of the surface leading up to the constriction increases for the former and decreases for the latter. Therefore, there is pobably no major increase in pressure above Hertzian associated with the region just inside of the lateral constrictions as is found with the rear constriction. Thus, the pressure peak (in whatever form it may take) is likely to be confined only in the region just upstream of the rear constriction. This pressure tends to boost up the downstream edge of the Hertzian region and appears to be one reason why the film thickness at the rear constriction is generally greater than the lateral constriction. The low film thickness associated with the lateral

constrictions appears to be the result of side leakage in the inlet region which is more severe at the sides of the inlet region than at the centre. This inhibits pressure generation and hence film thickness along the sides of the Hertzian region. For the given example, the film thickness at the lateral constrictions (h_m = 8.5 micro-inches) is 56% of the film thickness at the rear constriction (h = 15.2 micro-inches) and about 46% of the film thickness in the centre (h_o = 18.3 micro-inches).

5.4 Determination of Pressure-Viscosity Coefficients for Test Lubricants

The pressure-viscosity coefficient (α) of LUB D as measured by Galvin, Naylor and Wilson (125) is plotted as a function of temperature in fig.34. This lubricant was used as a reference to obtain relative α values of the other test lubricants. The method of determining α for each lubricant is described by Foord et al(16). It consists of comparing the dimensionless film thickness (h_o/R) of each lubricant with the reference lubricant, at a constant load (5 lbf.) and speed parameter ($\frac{\eta_o u}{E'R} = 2\times10^{-11}$).

The central film thickness of the test lubricants was measured as a function of speed using the counterrotation arrangement and plotted in dimensionless terms in fig.35. The measured values of the film thickness have been corrected for refractive index using the maximum Hertzian pressure (65,000 psi) which should be very similar to the actual pressure in the centre of the Hertzian region. The separation between the lines is, to a first order approximation, due to differences in α

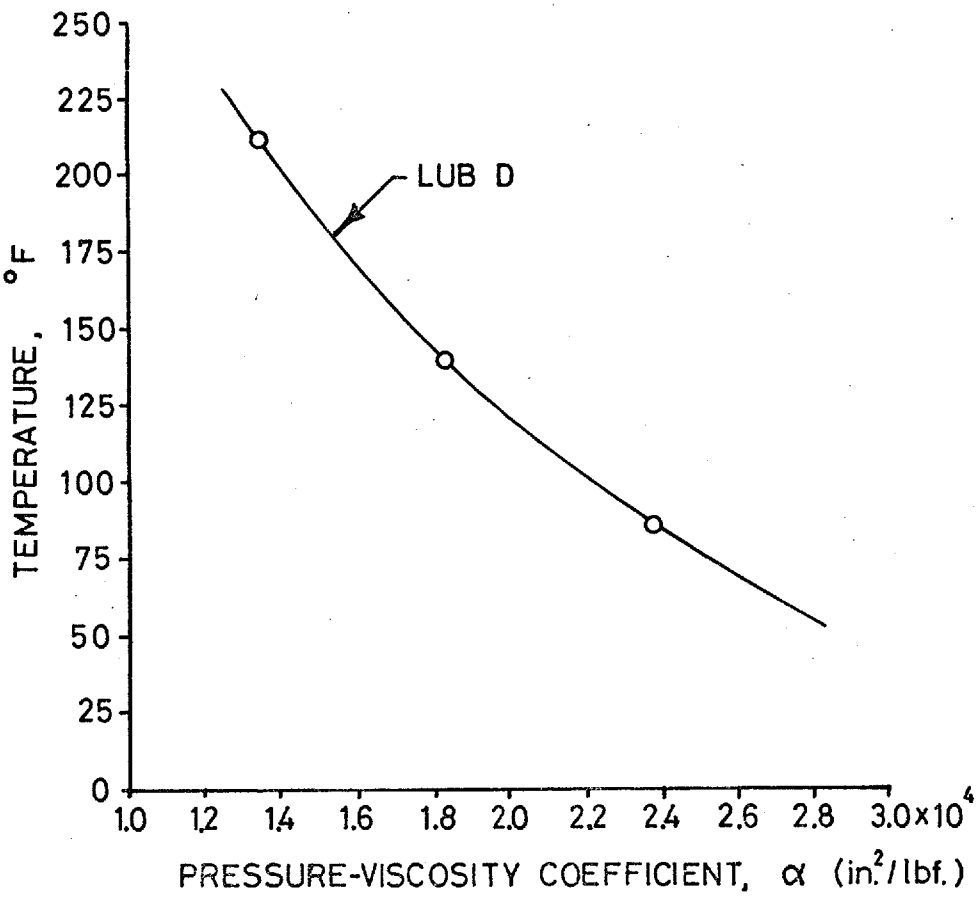


Fig.34 Pressure-viscosity coefficient vs temperature for LUB D between 0 and 5,000 psi. (taken from ref.125)

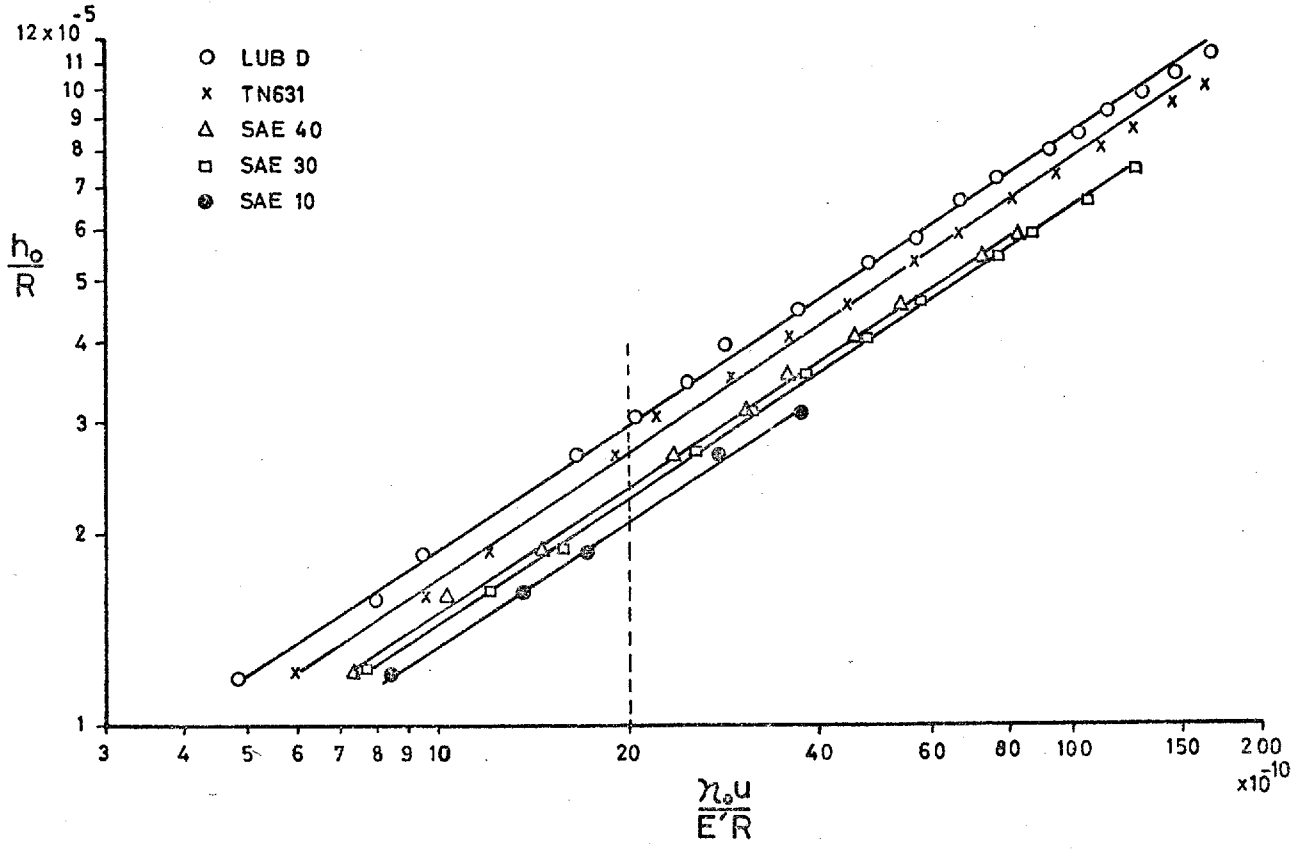


Fig.35 Dimensionless film thickness against speed parameter for each test lubricant

associated with each lubricant since, under constant load, α is the only ehd film thickness variable which is missing from the graph. The relative α for each lubricant was determined from the measured α of the reference lubricant ($\alpha_{LUB\ D}$) according to the relation

$$\alpha = \alpha_{LUB\ D} \left[\frac{(h_o/R)}{(h_o/R)_{LUB\ D}} \right]^{\frac{5}{3}}$$

It has been assumed here that h_o is proportional to α^6. This is consistant with the film thickness formula given by Dowson and Higginson(49) for line contact. The later formula suggested by Dowson(55) gives h_o proportional to α^{54}. More recently, Westlake(25) has found h_o proportional to α^{55} from his experimental point contact investigations. This lowers the calculated result by about 4%.

The relative α for each lubricant as measured by the above method is shown in table 3. These are compared with estimated α values which were obtained from temperature-viscosity

TABLE 3

Measured Pressure-Viscosity Coefficients

LUBRICANT SYMBOL	TEST TEMP.* ^oF	PRESSURE-VISCOSITY COEFF., $\frac{in^2}{lbf}$ MEASURED	ESTIMATED†
SAE 10	75.5	1.41×10^{-4}	1.28×10^{-4}
SAE 30	67.1	1.60	1.50
SAE 40	66.2	1.70	1.54
TN 631	73.4	2.14	1.75
LUB D	72.5	2.50+	1.89

* Taken as room temperature

\+ Measured by Galvin, Naylor and Wilson(125) and used as a reference

† Estimated using Fresco method(127)

characteristics using the correlation of Fresco(127). It should be noted that the estimated α value for the reference lubricant is 25% lower than the measured value given by Galvin et al(125). Therefore, a good correlation between the other lubricants can not be expected.

The accuracy of the measured values is subject to the errors associated with film thickness and speed measurements and the determination of the effective viscosity. The latter is likely to produce the greatest error since a difference in test temperature of $2^{o}F$ can lead to an error in α of 10%.

The results of Crook(56) show that the controlling viscosity is governed by the temperature of the bearing surfaces. Because of the difficulties in measuring surface temperatures of the rotating ball and/or the glass race, and the interpretation of thermocouple readings at positions around the contact region, the surface temperatures were assumed to be that of the ambient temperature. This appears to be a reasonable assumption since no change in film thickness, from an initial reading taken immediately after start-up, could be observed during running periods of over one hour. Thus, the cumulative heating of the surfaces must be very small, and the equilibrium surface temperature must be very close to the ambient temperature.

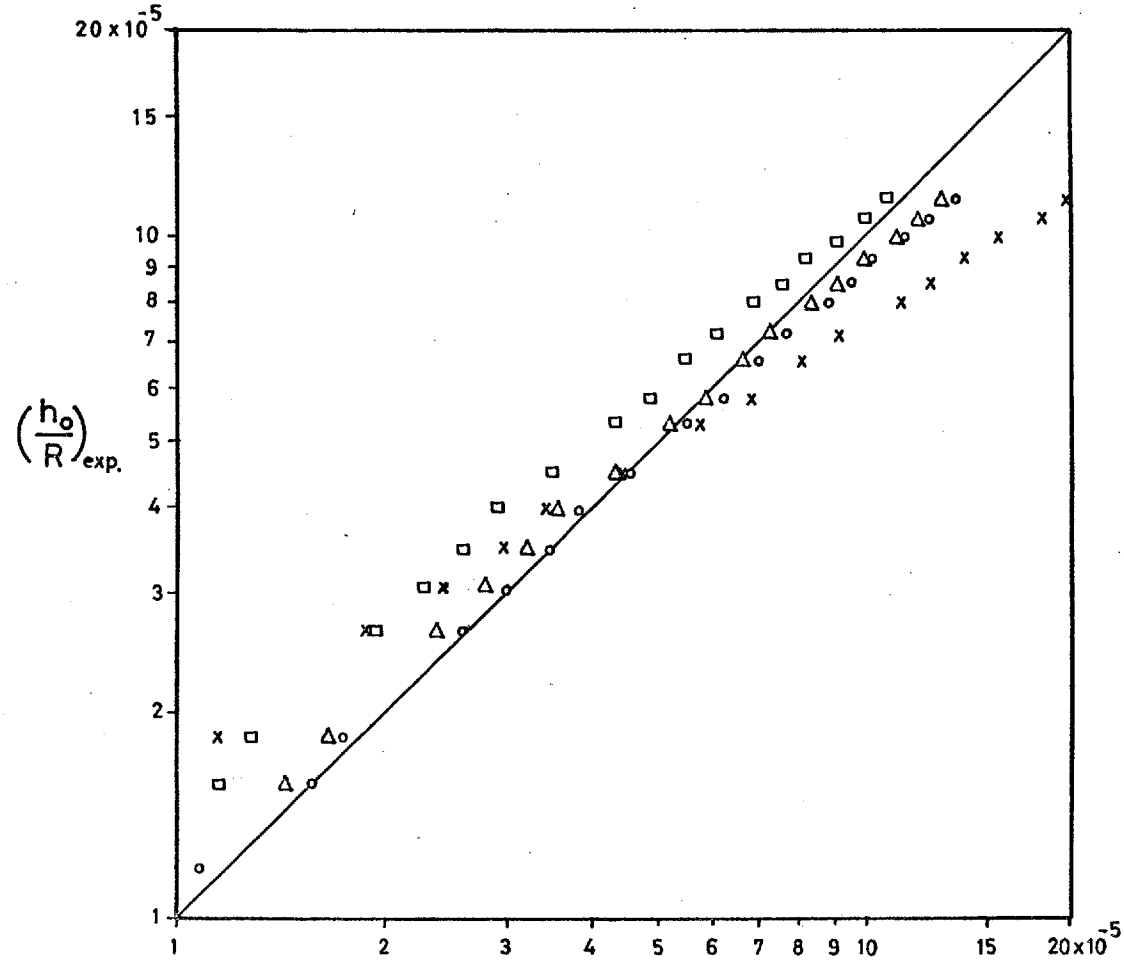
A local temperature rise will occur in the inlet region due to viscous heat generation. This may be of some significance here because the glass race has a thermal conductivity which is only 1/50 of steel. From the theoretical and experimental surface temperatures, presented by Cheng and Orcutt(63) for a glass disc, an estimated inlet temperature rise of 2 or $3^{o}F$ may occur for the highest rolling speed of

25 in./sec. used here. The effect of this temperature rise is unlikely to alter the film thickness a great deal. However, it may be one of the factors contributing to the slight drop in film thickness below the straight line shown in fig.35 for high values of the speed parameter.

5.5 Comparison of Theoretical and Experimental Film Thickness for Point Contacts

The experimental film thickness measurements of LUB D are plotted in fig.36 as a function of the theoretical film thickness predicted by the four film thickness formulae for point contact discussed in section 4.8b. There is very good agreement between the experimental measurements and the theoretical results of Evans(79) and Cheng(80). Their solutions were based on Grubin assumptions, and it now appears that these assumptions have enabled a reasonably accurate film thickness formula for point contact to be developed in the same manner as they did for line contact in the 1950's when a full solution had not yet been achieved.

The ability to predict the minimum film thickness at the lateral constriction is of practical importance. The central film thickness (h_o) and the minimum film thickness (h_m) as measured from random photomicrographs taken over a range of experimental loads between 1.92 and 12 lbf. are plotted in fig.37. The data for a materials conbination of sapphire/tungsten carbide are also included. These were obtained from a dimensionless plot of film thickness against speed given by Gohar(128). The plot of h_o against h_m shows a certain degree of scatter because h_m is more sensitive to load than h_o. To a first approximation



$$\left(\frac{h_o}{R}\right)_{theory} = K\left(\frac{\alpha \eta_o u}{R}\right)^a \left(\frac{w}{E'R^2}\right)^b$$

theoretical constants

SYMBOL	SOURCE	K	a	b
□	ARCHARD & COWKING	1.4	.74	-.074
x	CAMERON & GOHAR	3	1	-.33
Δ	CHENG	1.69	.725	-.058
o	EVANS	1.73	.714	-.048

experimental constants

LUBRICANT	LUB D
η_o	88.6×10^{-6} lbf.-sec./in.^{2}
α	2.5×10^{-4} in.^{2}/lbf.
w	5 lbf.
E′	16.9×10^6 psi.
R	0.5 in.

Fig. 36 Comparison of point contact theories and experiment

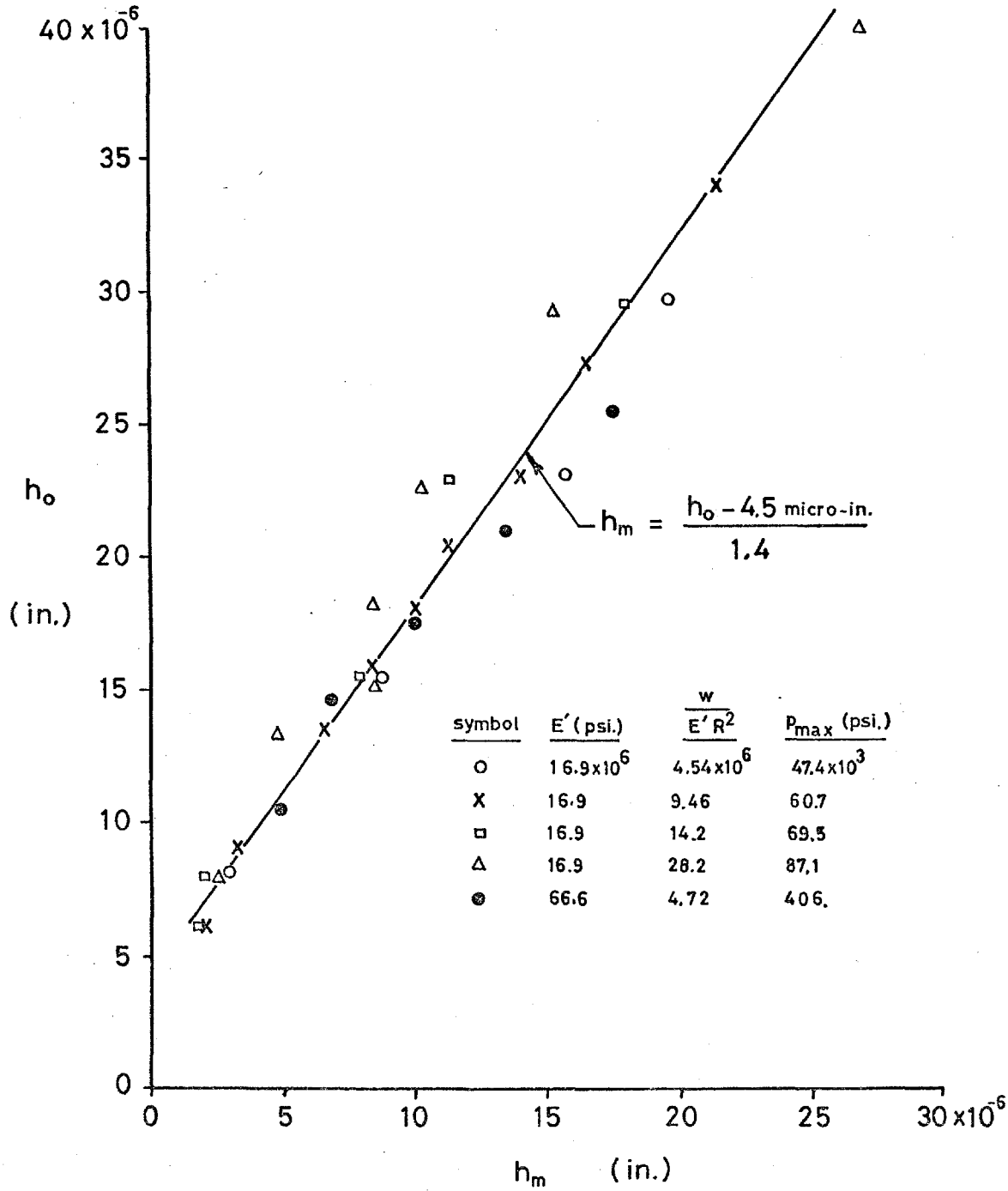


Fig.37 Plot of the central film thickness against the minimum film thickness for a range of experimental conditions

$$h_m \simeq \frac{h_o - 4.5 \text{ micro-inches}}{1.4}$$

An interesting feature of the results is that the load parameter $\frac{W}{E'R^2}$ for the sapphire/tungsten carbide data is nearly the same as the lowest load parameter for the glass/steel data. These two cases show very good correlation even though the reduced elastic modulus of the former (66.6×10^6 psi) is four times the latter, and the maximum Hertzian pressure of the former (405,000 psi) is eight and one-half times the latter.

At the time of writing, both Gohar(129) and Westlake (25) have obtained empirical film thickness formulae for point contact from optical measurements. Gohar(129) used several different material combinations covering a 140-fold range of Young's Modulus. He found that $h_o \propto (E')^{-.1}$, indicating that the central film thickness is not very sensitive to changes in the elastic properties of the materials. Westlake(25) used a glass/steel materials combination and obtained an expression for the minimum film thickness at the lateral constrictions as well as the central film thickness. An interesting feature of his results is that the central film thickness increases somewhat more rapidly with speed at higher loads.

Chapter 6 LUBRICANT STARVATION

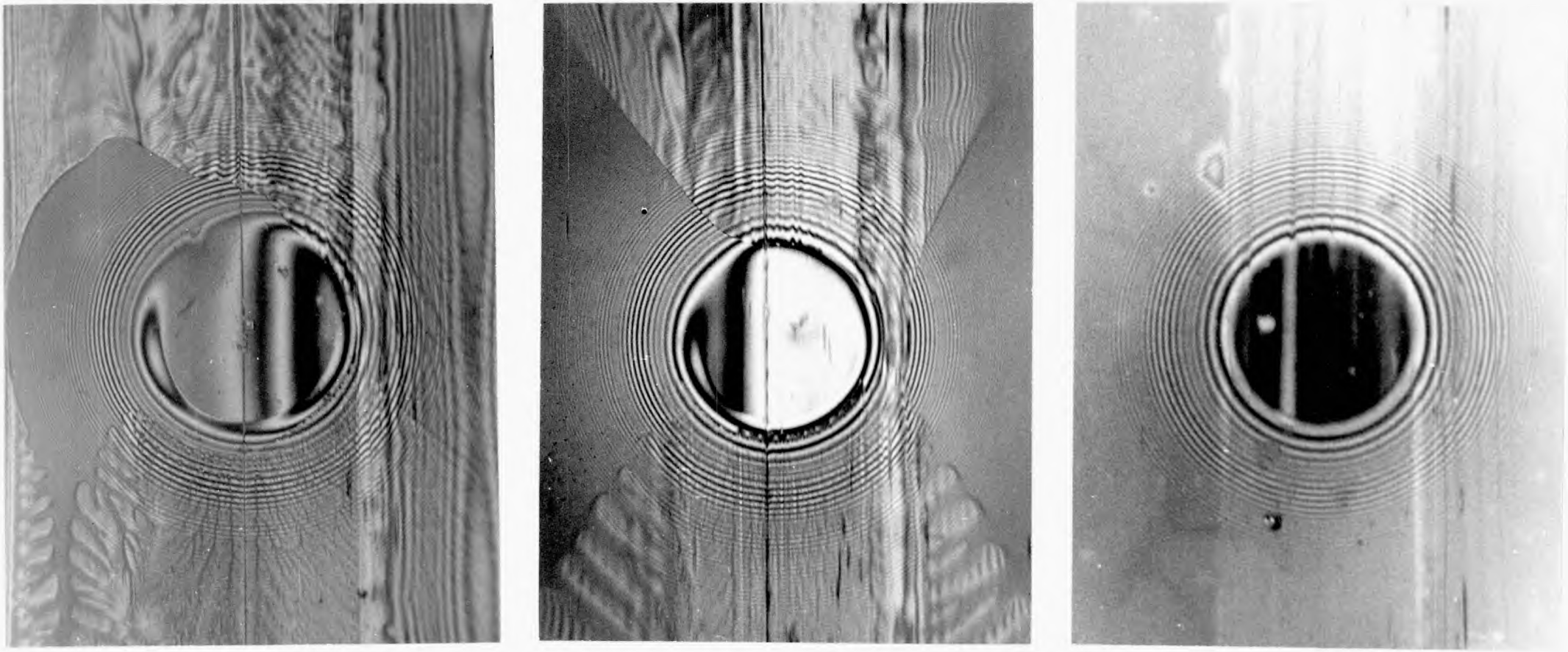
6.1 Introduction

In this chapter the effect of inlet lubricant supply conditions on film thickness and shape are shown experimentally. The general features of the results are discussed, and a theoretical approach to the starvation problem is developed. The experimental and theoretical results are then compared.

6.2 Inlet Lubricant Boundary and Film Shape

The experiments described here were conducted with the counterrotation arrangement described in section 2.3. Before each test the bearing components were cleaned with solvent (either benzine or acetone). Lubricant was applied to the ball and raceway surfaces during the reassembly of the bearing. Under static conditions surface tension causes the lubricant to form a circular boundary or meniscus around the contact region. A side view of the lubricant boundary can be seen in fig.4 of Chapter 2.

Under dynamic conditions the quantity and distribution of lubricant in the inlet region is reflected in the shape and location of the lubricant boundary in the inlet region. The photomicrographs in fig.38a and b show clearly defined areas where the gap between the bearing surfaces is completely filled with lubricant. Since the medium is continuous, these areas give very uniform interference fringes. Outside these areas the medium is a composite of air and oil, and the fringes which form between the ball and glass race are shifted to locations of greater thickness because the refractive index of air is less than oil. This is clearly seen in the



(a) (b) (c)

Fig.38 Effect of inlet lubricant boundary on film shape

inlet region of fig. 38b. The degree of fringe shift depends on the relative amounts of air and oil at each location. In this way the rippled fringes in the inlet region show quantitatively the distribution of lubricant on the bearing surfaces. A fringe which is shifted closer to the Hertzian region indicates a greater percent of oil.

The rippled fringes in the inlet region also show that the lubricant is distributed over the bearing surfaces in the form of small ribs which are generally orientated in the direction of motion. These ribs are developed in the outlet region where the lubricant film ruptures into striations of air and oil. This is most clearly seen in fig.38a. An interesting feature is that the oil film cavitates preferentially along the scratch which runs along the track of the glass race. This causes two parallel ribs (seen as narrow, dark lines with finger-like projections) to form adjacent to the scratch. These ribs, though somewhat flattened out, can also be seen in the inlet region. Thus, the amount and distribution of lubricant in the outlet region influences the inlet region.

Excess lubricant is frequently found on the sides of the track as shown in fig.38b. These areas act as reservoirs from which lubricant can flow into the inlet region by the action of surface tension. The rate of flow is probably a complicated function of surface tension, viscosity, gap thickness, and the shape of the lubricant boundary; and the amount of lubricant which is recaptured in the inlet region by surface tension will certainly depend on the time available, i.e. the speed of the bearing surfaces. The amount and distribution of lubricant which fills the gap at any given time is clearly determined by the inlet lubricant boundary.

The ehd films shown in fig.38 are clearly effected by a starved inlet region; and the film shape, which can be obtained by reading the fringes as lines of constant thickness, is strongly dependent upon the shape of the inlet lubricant boundary. The ehd film in fig.38c is very severely starved. The black fringe in the Hertzian region indicates a film thickness of only 2 micro-inches (500 Å). The ehd film in the Hertzian region consists of a series of channels. These are caused by the ribs of lubricant which can be seen in the inlet region.

These results not only show that the film thickness is primarily determined in the inlet region, but also that the local film thickness at each point within the Hertzian region is primarily a function of the inlet conditions immediately upstream of that point.

6.3 Test Conditions and Procedure

The purpose of the starvation experiments was to determine the effect of the inlet boundary on the central film thickness for a range of operating conditions. For a constant speed, load, and lubricant properties, the required information is the central film thickness (h_o) and the boundary distance (x_b) measured from the centre of the Hertzian region to the inlet boundary immediately upstream. This is shown in fig.39.

It was not always possible to obtain a uniform distribution of lubricant over the bearing surfaces. Under starvation conditions this makes the inlet boundary and the central film thickness a transient phenomenon, and it was found necessary to obtain the required information from high-speed photo-

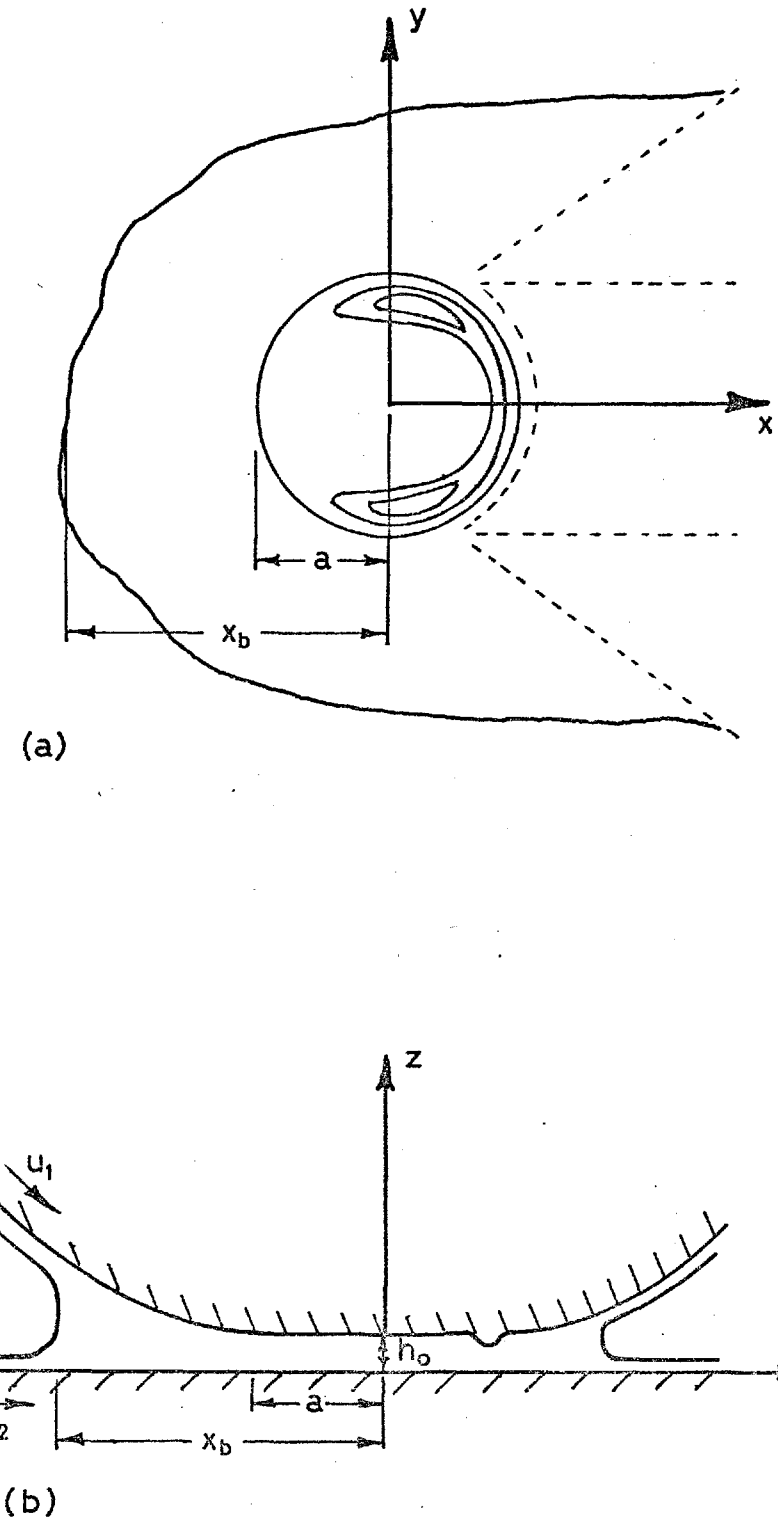


Fig.39 Geometry of the contact region showing required measurements for starvation; (a) plan view, (b) cross-sectional view along centre line

micrographs. These were taken with black and white film because of the high cost of producing the same in colour. To aid the recognition of fringes a series of photomicrographs were obtained under flooded conditions, each having a central film thickness corresponding to a given calibrated fringe. In this way the relative intensity corresponding to each coloured fringe could be determined. Wherever possible, pictures were taken during a test run when the inlet boundary gave a central film thickness corresponding to a particular fringe. This reduced the need for estimating film thicknesses at intermediate fringe positions.

The centre of the Hertzian region, which is the origin of the x and y-axes, was determined graphically from the fringes outside the Hertzian region in the transverse direction. The x-axis is parallel to the track and located at one-half the diameter of a given fringe. The y-axis was located from the perpendicular of a line drawn tangent to a fringe outside the Hertzian region and parallel to the x-axis. The boundary distance (x_b) was measured from the photomicrographs whose scale factor was obtained by photographing a rule with 0.01 inch graduations. Because the local film thickness in the Hertzian region is primarily a function of the inlet conditions immediately upstream, the measurement of x_b was always taken from the centre of the Hertzian region to the lubricant boundary immediately upstream, even though the lubricant boundary was sometimes irregular across the inlet region.

Starvation was achieved by gradually reducing the quantity of lubricant within the bearing system. A typical

test run consisted of the following:

(1) Thrust bearing dismantled and components cleaned with benzine or acetone
(2) Test lubricant applied to bearing surfaces during reassembly in sufficient amounts to provide a flooded inlet region for the predetermined operating conditions.
(3) Bearing run at predetermined load and central film thickness
(4) Temperature and speed recorded. Picture taken of flooded condition
(5) Bearing stopped. Glass race and/or test ball removed and cleaned
(6) Bearing reassembled and operated at the previously measured speed. Two or three pictures taken of starved condition
(7) Steps (5) and (6) repeated until lubricant almost completely removed from bearing surfaces

The chosen operating conditions consist of three loads on the test ball (1.92, 6, and 12 lbf.)and four initial (i.e. flooded) film thicknesses. The lowest load corresponds to zero load on the spring balance, and the highest load is near the breaking load of the glass race. The four initial film thicknesses correspond to the first four red fringes of the duochromatic system. To maintain a convenient operating speed these film thicknesses were obtained by varying the test lubricant as well as the bearing speed. All possible combinations of the chosen operating conditions were tested, giving twelve individual test runs. The relevant data for each test run is given in table 4.

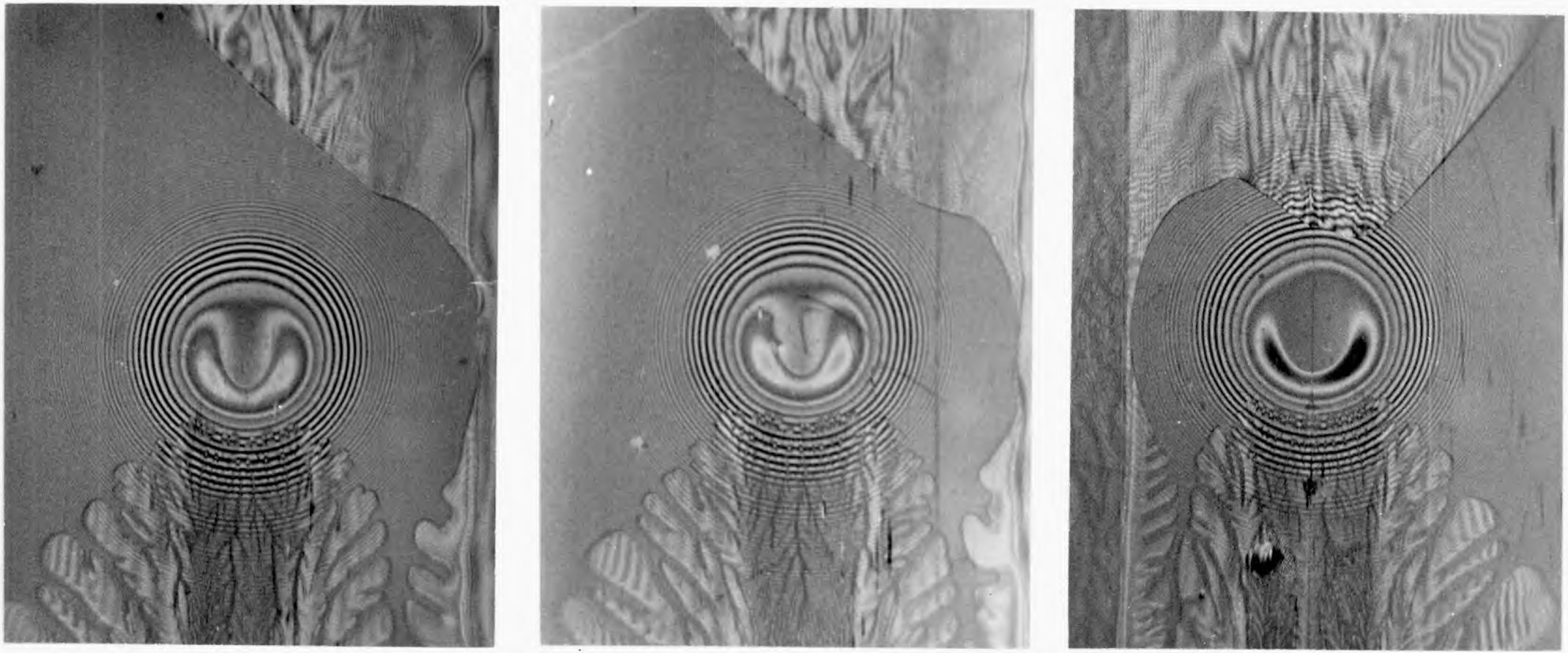
6.4 Starvation Results

Fig.40 shows a sample of photomicrographs taken from a test run. The test load is 1.92 lbf., and the initial film thickness corresponds to the fourth red fringe (h_o = 30 micro-

TABLE 4

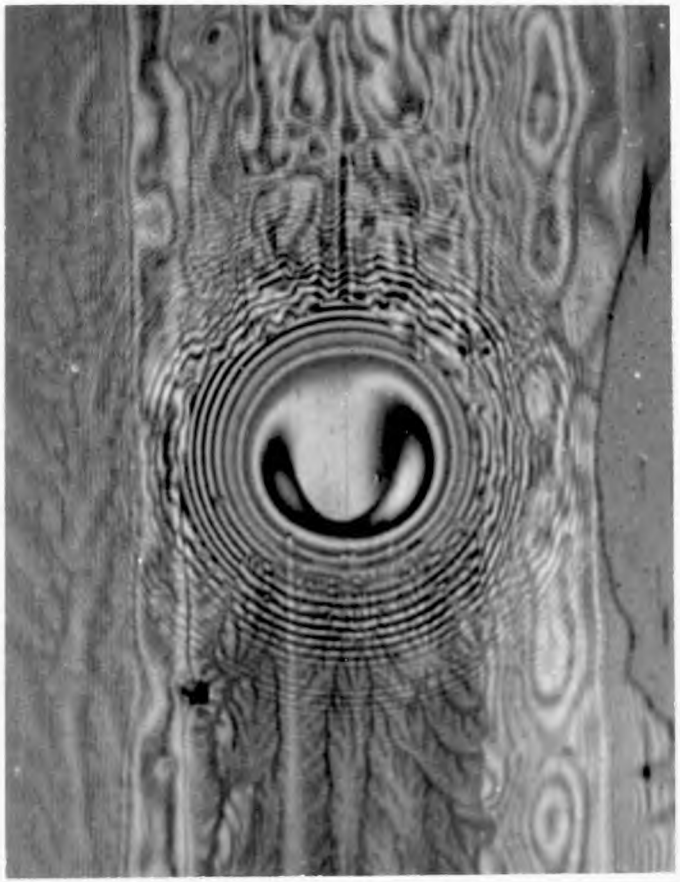
Relevant data for each starvation test run

test loading	w = 1.92 lbf. a = 4.39×10^{-3} inches p_{max} = 47,400 psi				w = 6 lbf. a = 6.42×10^{-3} inches p_{max} = 69,500 psi				w = 12 lbf. a = 8.1×10^{-3} inches p_{max} = 87,000 psi			
central fringe for flooded condition	1st red	2nd red	3rd red	4th red	1st red	2nd red	3rd red	4th red	1st red	2nd red	3rd red	4th red
h_o ($\times10^{-6}$ inches)	8.13	15.75	23.2	29.8	8.0	15.55	22.7	29.4	7.80	15.4	22.6	29.1
$u = \frac{1}{2}(u_1+u_2)\frac{\text{in.}}{\text{sec.}}$	11.7	5.7	10.75	9.35	11.9	6.34	6.7	9.24	12.45	6.29	14.3	9.81
test temp. ^{o}F	73.5	70	70.5	68	72.5	69.9	72.5	66.2	72.5	67.6	74.4	66.2
lubricant	SAE10	SAE30	SAE30	SAE40	SAE10	SAE30	TN631	SAE40	SAE10	SAE30	SAE30	SAE40
$\eta_o \times10^{-6}\ \frac{\text{lbf.-sec.}}{\text{in.}^2}$	8.7	40.3	39.6	68	8.96	40.3	55.0	74.5	8.96	43.5	34.0	74.5
α ($\times10^{-4}\text{in}^2/\text{lbf.}$)	1.41	1.60	1.60	1.70	1.41	1.60	2.14	1.70	1.41	1.60	1.60	1.70
$\frac{\alpha\eta_o u}{R}\times10^{-8}$	2.87	7.35	13.6	21.6	3.0	8.17	15.7	23.4	3.15	8.89	15.6	24.8

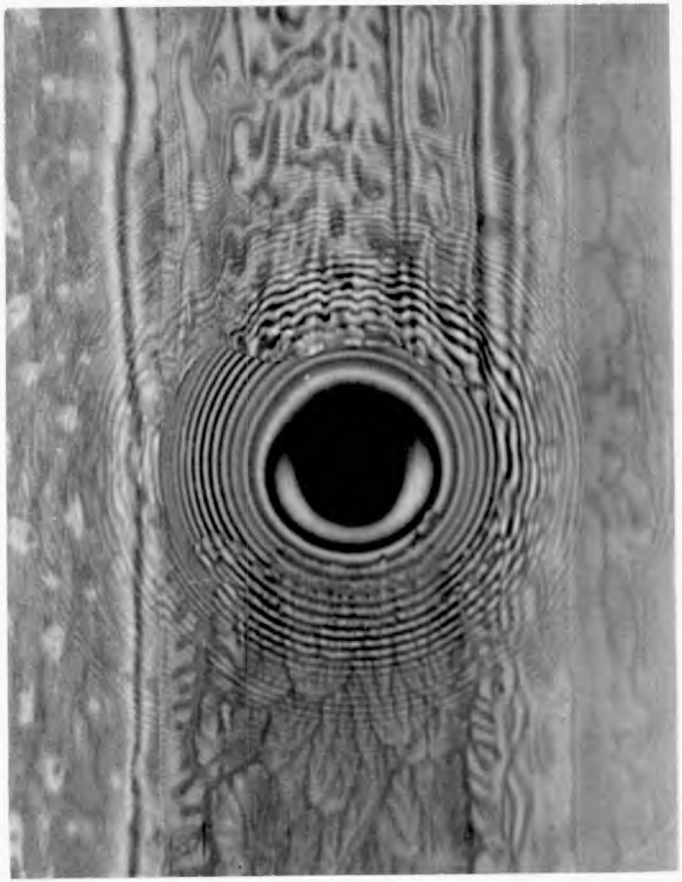


(a) (b) (c)

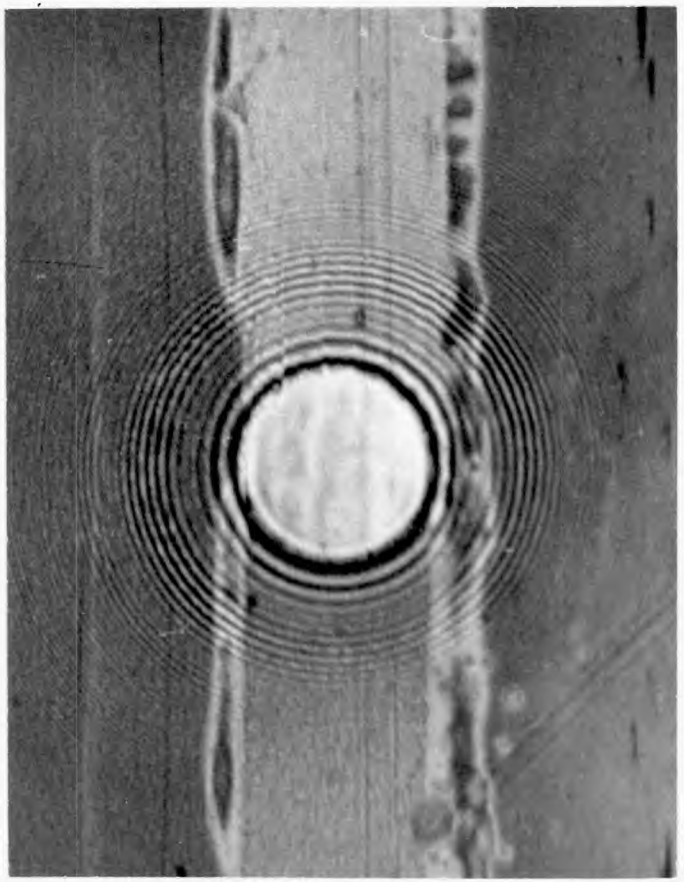
Fig.40 Sample of photomicrographs showing progressive starvation
(continued on next page)



(d)



(e)



(f)

Fig.40

inches). In fig.40a the boundary distance is sufficiently removed from the centre of the Hertzian region to obtain a flooded condition. Figs.40b-f show a diminishing film thickness as the inlet boundary progressively approaches the Hertzian region. The diminishing film thickness is apparent if one observes that the fringe in the centre of the Hertzian region for a given figure is always in the constriction regions in the previous figure. It should also be noted that as starvation progresses, the ehd shape approaches the Hertzian shape. Therefore, the ehd pressure must approach the Hertzian pressure. In the limit, i.e.when h_o equals zero, the Hertzian shape and pressure are attained.

Fig.41 is a plot of the central film thickness against the ratio (x_b/a) where (a) is the calculated Hertzian radius. The results show that film thickness becomes increasingly more sensitive to the inlet boundary as the boundary approaches the Hertzian radius. Indeed, it can be seen that the film thickness approaches zero as the inlet boundary approaches the Hertzian radius. This certainly substantiates the importance of the inlet region in regards to film thickness.

The boundary distance (x_b) required to obtain a flooded inlet region, i.e.when h_o is no longer sensitive to x_b, depends on the initial film thickness. For example, when the initial h_o is 8.13 micro-inches, the required x_b is 1.9a, and when the initial h_o is 29.8 micro-inches, the required x_b is about 4.5a. These results reflect the pressure curves of Dowson and Higginson which were given in fig.26a of Chapter 4. They show that the inlet pressure (which determines film

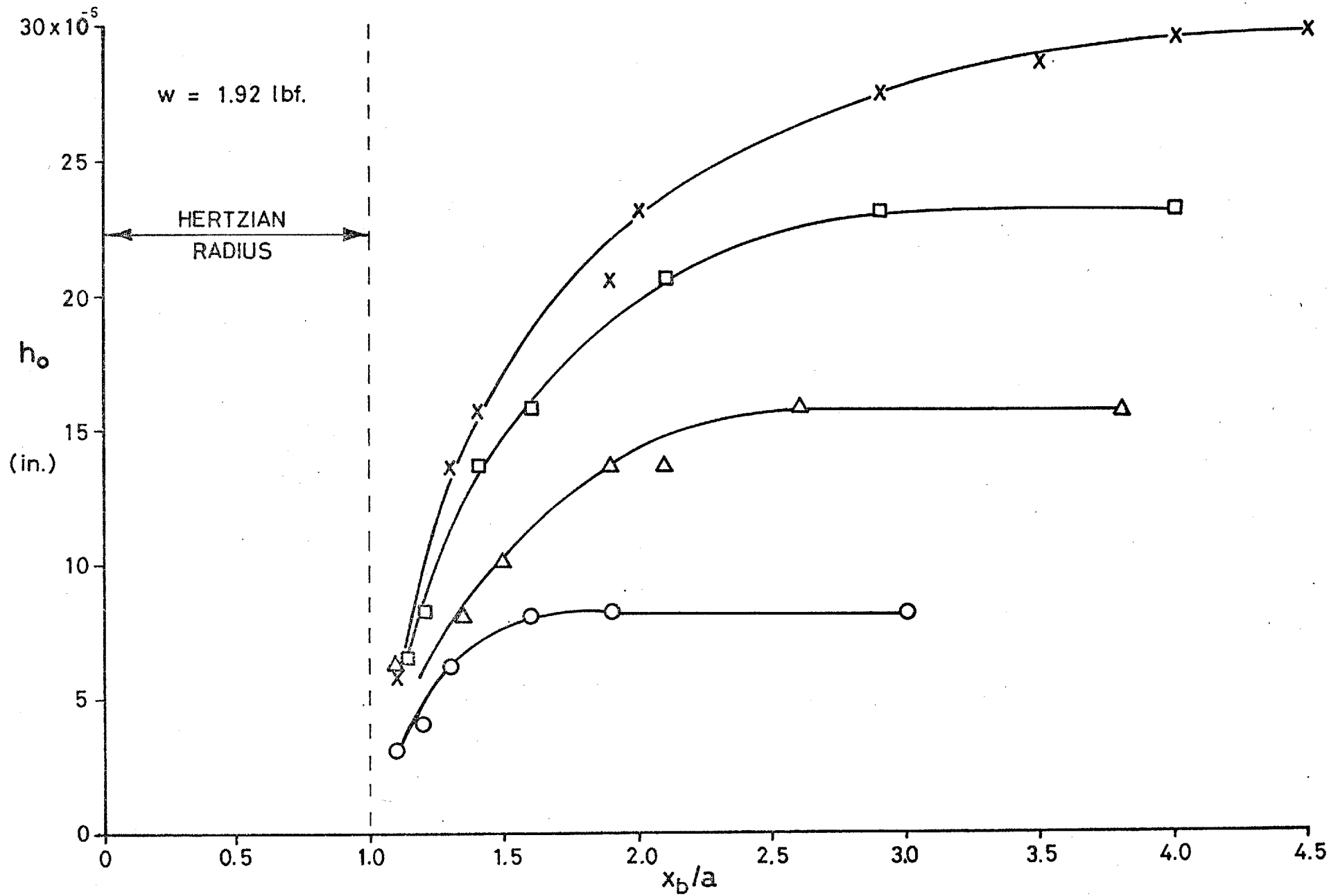
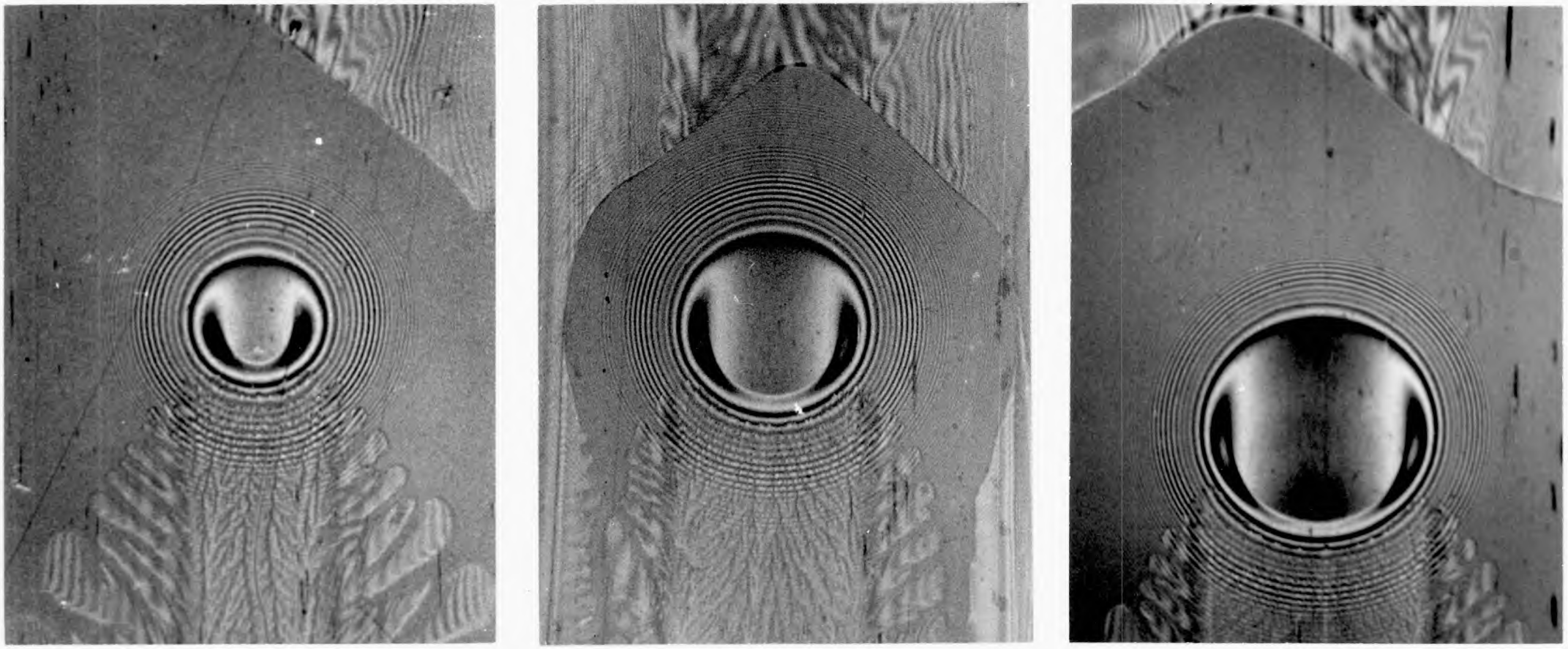


Fig.41 Effect of boundary distance on film thickness for different initial film thicknesses

thickness) commences further ahead of the Hertzian region as the speed parameter, and hence the film thickness, increases.

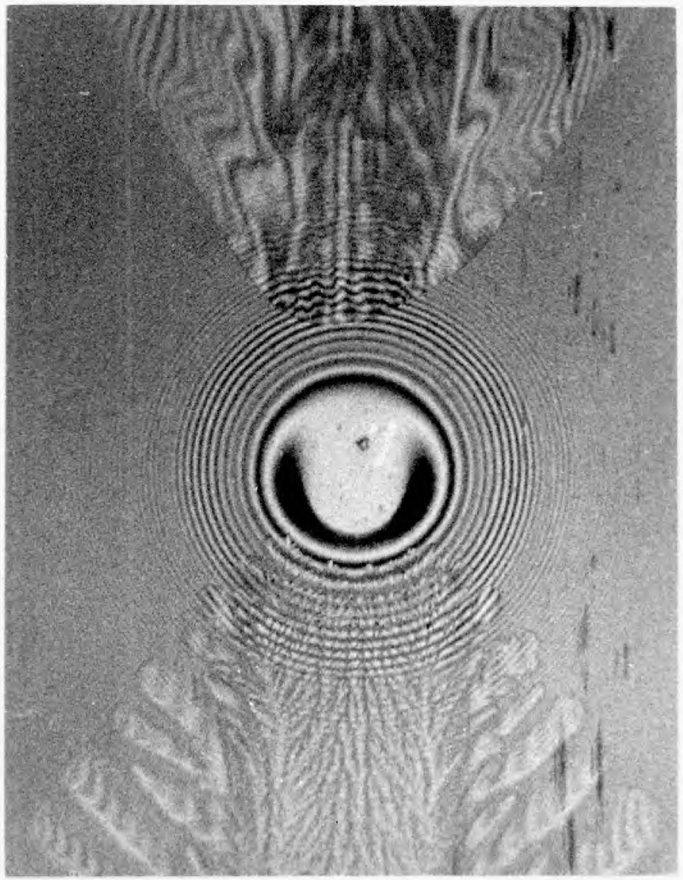
Fig.42 shows a sample of photomicrographs taken from three test runs corresponding to each of the test loads. The initial film thickness corresponds to the second red fringe. Figs.42a, b, and c show the effect of load on film shape for a flooded condition. Load increases the deformation and causes the film shape and hence the pressure distribution to become closer to the Hertzian shape and pressure distribution. It is also seen that as h_o is kept constant the minimum film thickness located at the lateral constrictions decreases slightly with load. The same general trend appears to hold under starvation conditions. This is shown in figs.42d, e, and f where the degree of starvation is slight for each test load, and in figs.42g, h, and i where the degree of starvation is quite severe for each test load. The 'footprint' in the Hertzian region of fig.42i is probably the result of a foreign particle trapped between the surfaces. An important feature of these results is that for each set of starvation conditions the distance between the inlet boundary and the Hertzian region is nearly the same for each test load.

The starvation results of these three load tests are plotted in fig.43. The initial film thickness which corresponds to the second red fringe is slightly different for each load because of the change of refractive index with pressure. To obtain a flooded condition the results show that the required boundary distance in terms of the Hertzian radius becomes smaller as the load increases. These results are consistant with the expected pressure distributions which must

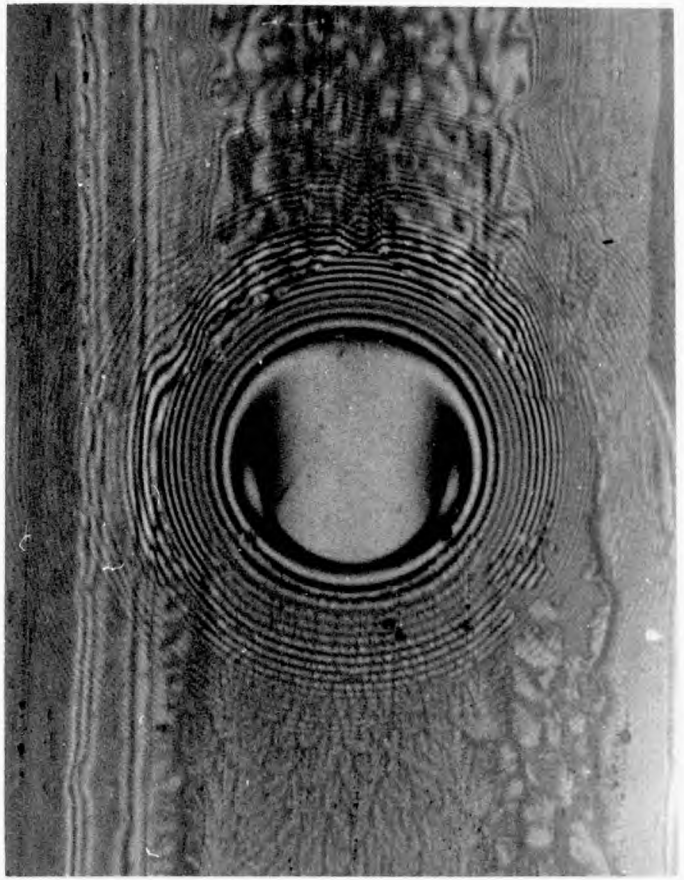


(a) (b) (c)

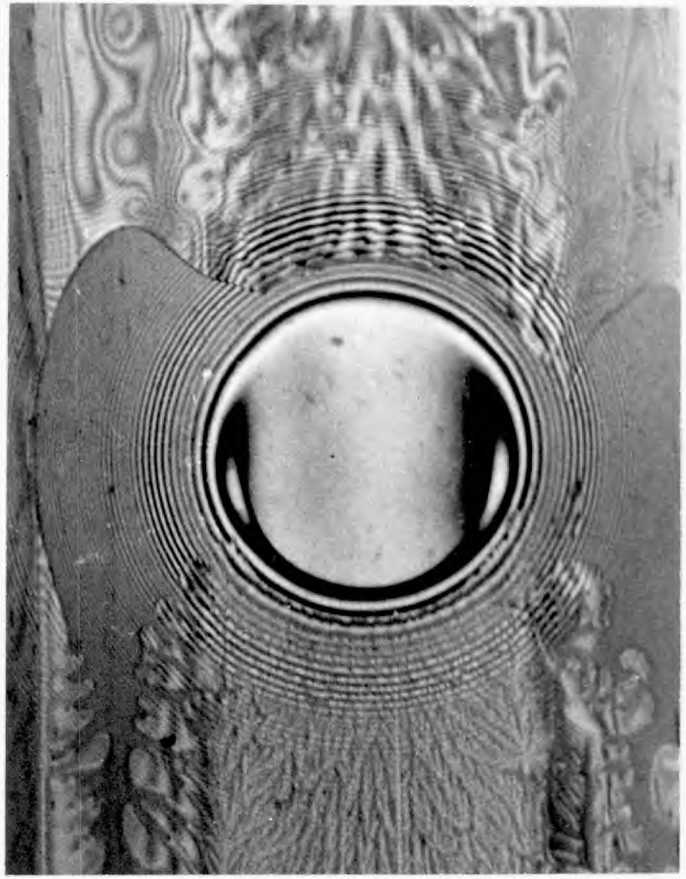
Fig. 42 Sample of photomicrographs showing effect of load
(continued on next page)



(d)

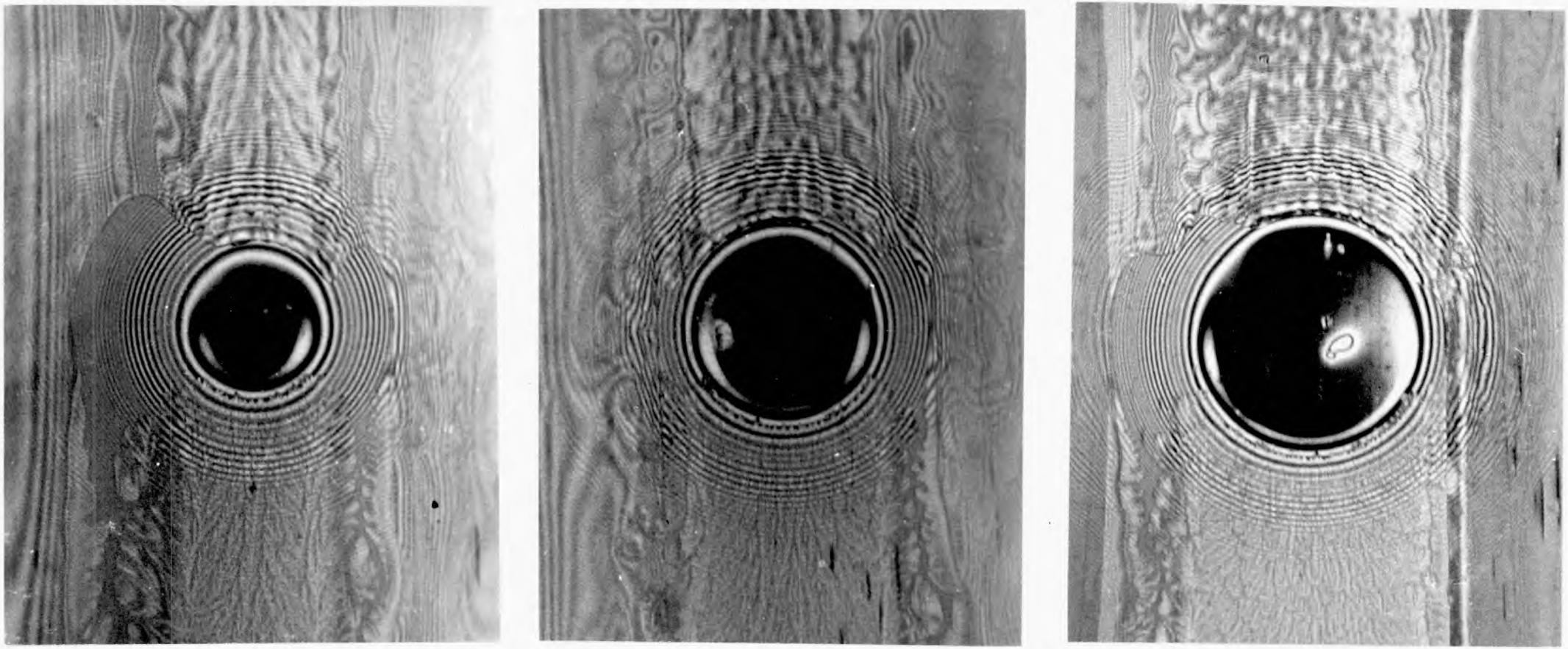


(e)



(f)

Fig. 42 (continued on next page)



(g) (h) (i)

Fig.42

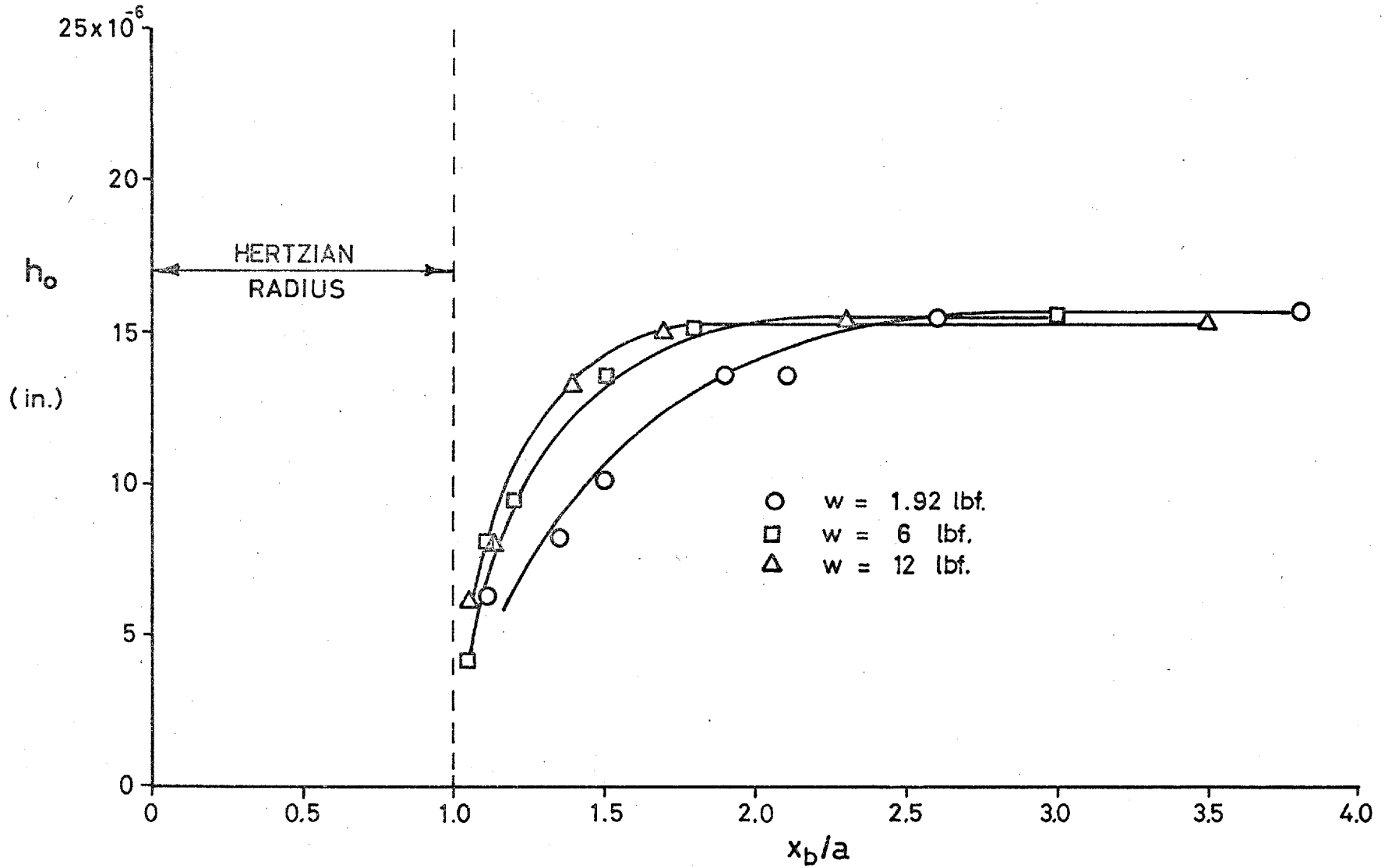


Fig.43 Effect of boundary distance on film thickness for different loads

become closer to Hertzian as the load increases.

6.5 Correlation with Inlet Boundary

The results of the previous section show that starvation is a phenomenon which is associated with the location of the lubricant boundary in the convergent inlet region, and that as the lubricant boundary approaches the Hertzian radius, the film thickness approaches zero. It was also noted that the boundary distance required to obtain a flooded condition is reflected in the expected location of pressure commencement in the inlet region. These results imply that film thickness is determined by the ehd pressure generated in the inlet region. The approximate magnitude of this pressure for the example discussed in section 5.3 is shown in fig.44, which is a plot of the pressure difference between the estimated ehd pressure and the Hertzian pressure.

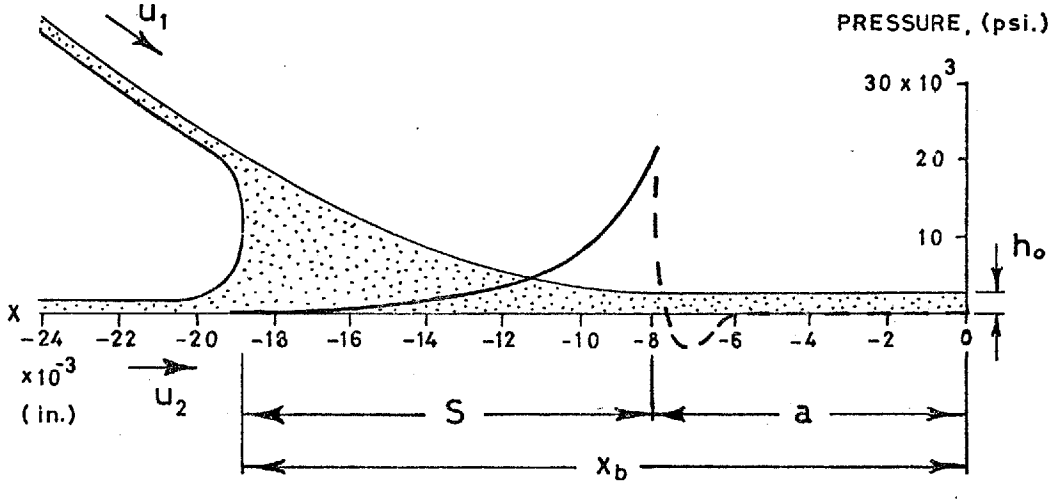


Fig.44 Approximate ehd pressure generated in inlet region

A starved inlet region delays and inhibits the generation of this pressure, and therefore causes a reduction in film thickness.

In the previous section, the boundary distance (x_b), measured from the centre of the Hertzian region to the lubricant boundary, was used to indicate the degree of starvation. This was done to show some of the general features of the results. Since starvation is a phenomenon which occurs in the inlet region, a more relevant measure of the degree of starvation would appear to be the distance S shown diagrammatically in fig.44. S is defined as

$$S \equiv (x_b - a)$$

and is called the 'inlet distance'. It reflects the distance over which the ehd pressure is generated in the inlet region.

Fig. 45 is a plot of the starvation results using the inlet distance as a measure of the degree of starvation. This reduces the experimental data to only four curves, each showing the film thickness approaching zero as the inlet distance approaches zero. Each curve contains the data for all three test loads thus indicating that film thickness is not very sensitive to load under starved as well as flooded conditions. This is consistant with the view that film thickness is primarily determined by the ehd pressure generated in the inlet region. The primary effect of load is merely to alter the size of the Hertzian region and the magnitude of the pressure therein. Load has only a secondary influence on the ehd pressure generated in the inlet region. Therefore, this pressure, and hence the film thickness determined by this

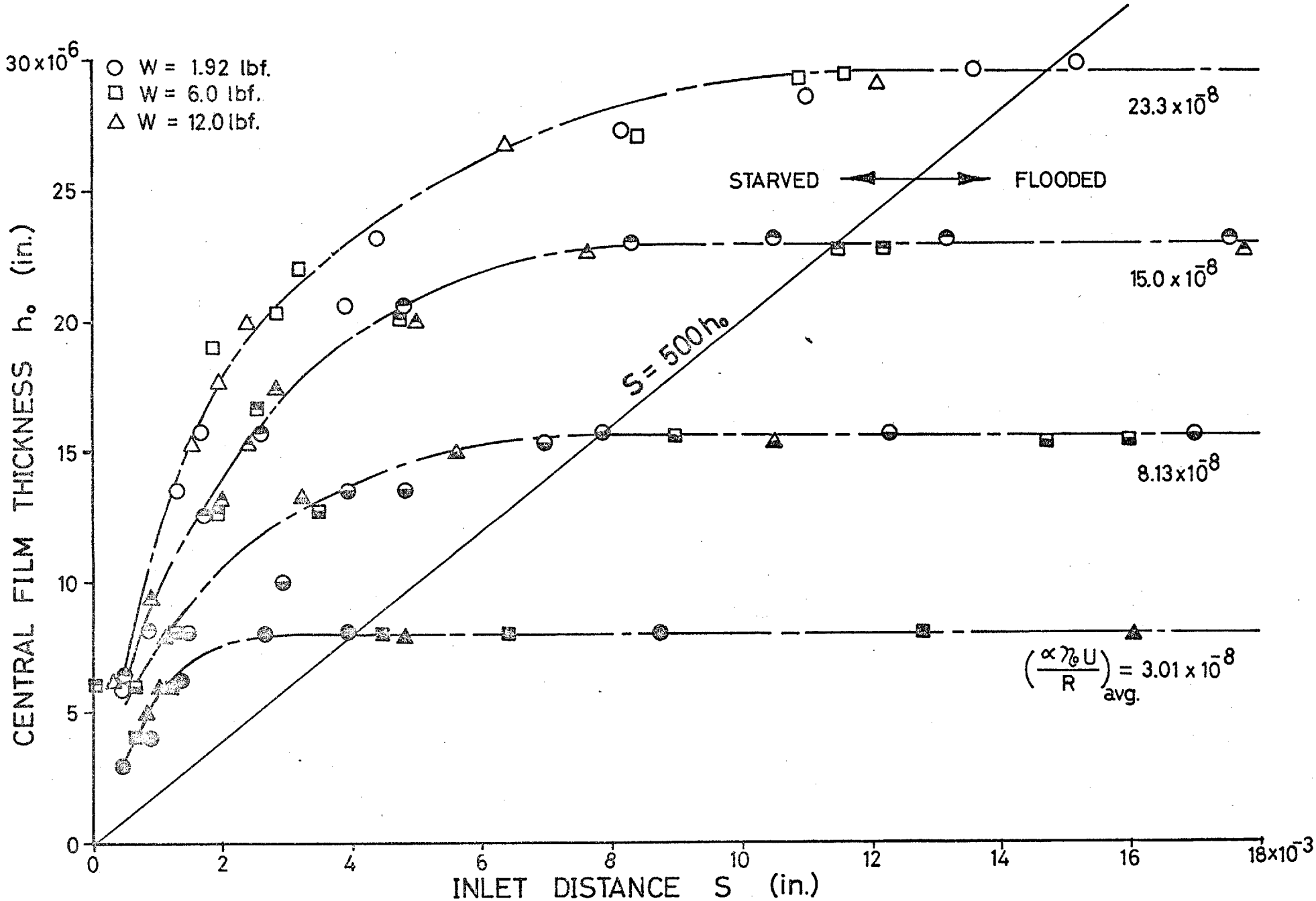


Fig.45 Experimental results plotted as film thickness against inlet distance

pressure, will remain essentially the same regardless of load.

To a first approximation the experimental results can be divided into two regimes by the straight line shown in fig.45. The bearing is starved when operating to the left of this line and flooded when operating to the right of the line. The equation of the dividing line is

$$S = 500\ h_o$$

This relation, which is only applicable to the particular experimental geometry and film thickness range, illustrates the relative magnitude of the dimensions involved.

It was shown in Chapter 4 that film thickness (under flooded conditions) is primarily a function of α, η_o, u and R. These variables combine into a dimensionless group of the form ($\alpha\ \eta_o\ u/R$). The numerical value of this group for each starvation test is given in table 4. An average taken over the three test loads is given in fig.45 for each initial film thickness. Fig.46 is a cross-plot of the experimental curves of fig.45. It shows how film thickness changes with speed (and/or α, η_o) for constant inlet distances. If the inlet distance is greater than 500 times the film thickness, then the film thickness increases along the line marked 'flooded'. If it is less than 500 times the film thickness, then the film thickness increases less rapidly with speed. For severely starved conditions, i.e. for small values of S, the film thickness increases only slightly with speed.

6.6 Inlet Boundary Conditions

Evaluation of pressure from the Reynolds equation (equation 4.1) requires knowledge of two boundary conditions

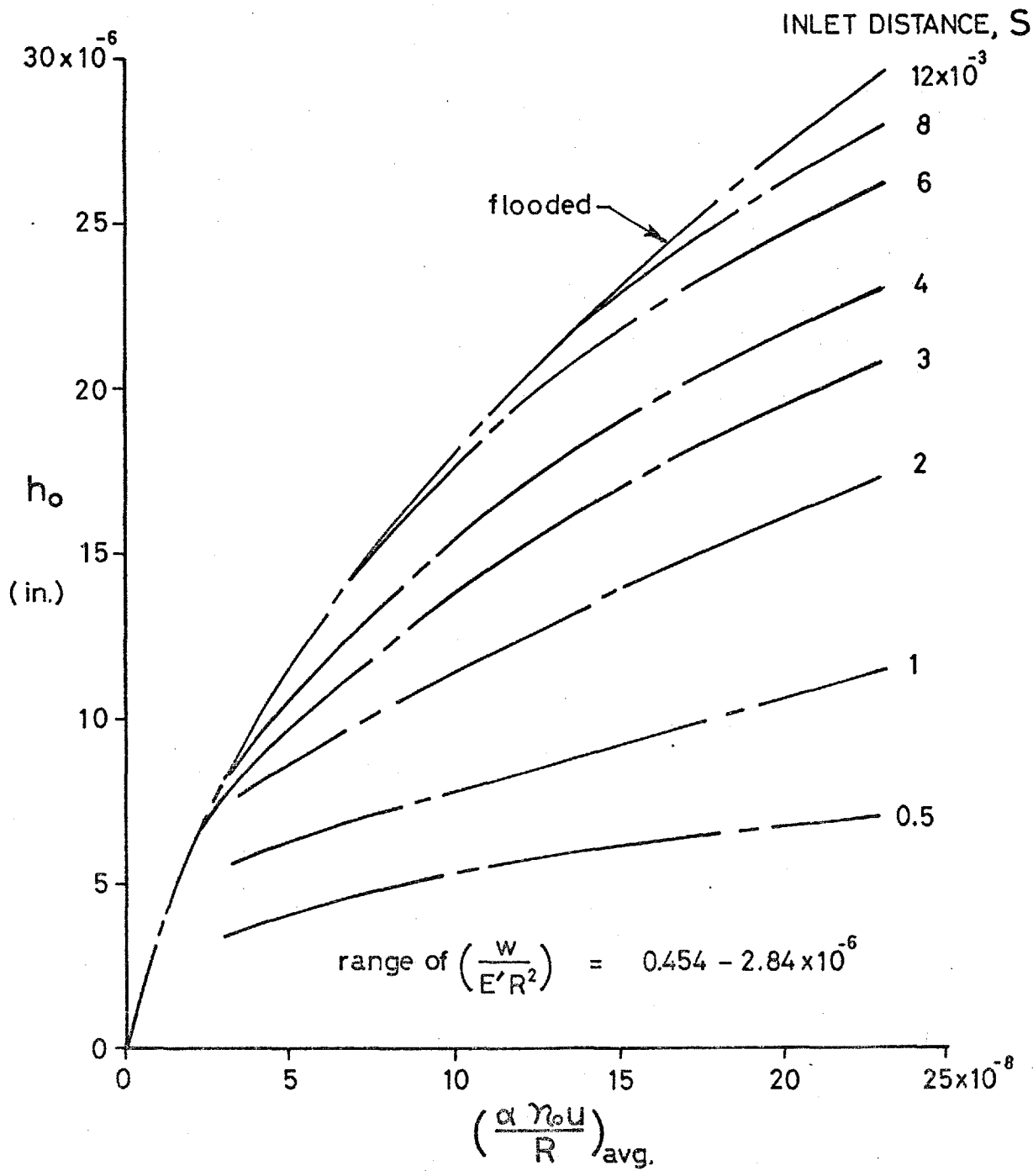


Fig.46 Experimental results showing film thickness as a function of the speed-viscosity parameter for constant inlet distances

when the side-leakage term is neglected. These are usually taken at the beginning and end of the pressure curve. For convenience the inlet boundary condition is usually taken as $p = 0$ at $x = -\infty$. Lauder (130) has questioned this boundary condition on the grounds that it does not conform to experimental facts. Instead, he put forward the boundary condition that $p = 0$ where the fluid velocity and velocity gradient are zero, i.e. $u = \frac{\partial u}{\partial z} = 0$. Furthermore, Whitaker (131) suggested that this inlet condition could usefully be called the 'oil starvation' condition, where any greater quantity of oil would result in an accumulation of oil at inlet; and any less would result in a decrease in film thickness.

The Lauder boundary condition is now applied to the example given in section 5.3. The velocity and velocity gradient can be found from (see e.g. ref.30, p.53)

$$u = \frac{1}{2\eta}\frac{\partial p}{\partial x} \cdot z(z-h) + (u_2-u_1)\frac{z}{h} + u_2 \qquad \ldots(6.1)$$

$$\frac{\partial u}{\partial z} = \frac{1}{\eta} \cdot \frac{\partial p}{\partial x} \cdot (z-\frac{h}{2}) + \left(\frac{u_1 - u_2}{h}\right)$$

The assumptions inherent in these equations are in common with those in Reynolds equation. These were stated in section 4.2. Side-leakage is neglected. This should be a reasonable assumption if one considers only the centre-line in the direction of rolling and very thin ehd films. For pure rolling $u_1 = u_2$, and $\frac{\partial u}{\partial z} = 0$ at mid-film where $z = \frac{h}{2}$. The velocity distribution at various locations within the inlet region was computed from equation 6.1. The pressure gradient and film thickness were obtained from fig.32 and the variation of viscosity with pressure was accounted for by the exponential relation (equation 4.3).

The velocity distribution at x = -10, -11.12 and -12×10^{-3} inches are shown in fig.47; and fig.48 is a plot of the variation of mid-film velocity in the inlet region. From the Hertzian radius, where the fluid velocity is nearly equal to the surface velocity, the mid-film velocity decreases and reaches zero at $x = -11.12 \times 10^{-3}$ inches. Upstream of this point some of the fluid flows backwards relative to the Hertzian region. The pressure at the Lauder boundary condition is not zero but about 4,700 psi, and the location of observable starvation is much further upstream ($x \simeq -17 \times 10^{-3}$ inches).

An interesting relation between h_o and the film thickness at the Lauder boundary condition can be obtained from the fluid velocity relation (equation 6.1) and the integrated form of Reynolds equation (equation 4.2). To satisfy $u = \frac{\partial u}{\partial z} = 0$ for pure rolling, equation 6.1 becomes

$$\frac{dp}{dx} = \frac{8u_2\eta}{h^2}$$

Substituting in Reynolds equation, and remembering that the definition of bearing speed for pure rolling in the Reynolds equation is equivalent to u_2, gives

$$\frac{8u_2\eta}{h^2} = 12u_2\eta\left(\frac{h-h_o}{h^3}\right)$$

which simplifies to

$$\frac{h}{h_o} = 3$$

This agrees very well with the above example where h = 54.5 micro-inches at $u = \frac{\partial u}{\partial z} = 0$, and h_o = 18.3 micro-inches; thus giving $h/h_o = 2.98$. At the location of observable starvation, however, $h/h_o \simeq 10$. This point, which is considerably upstream

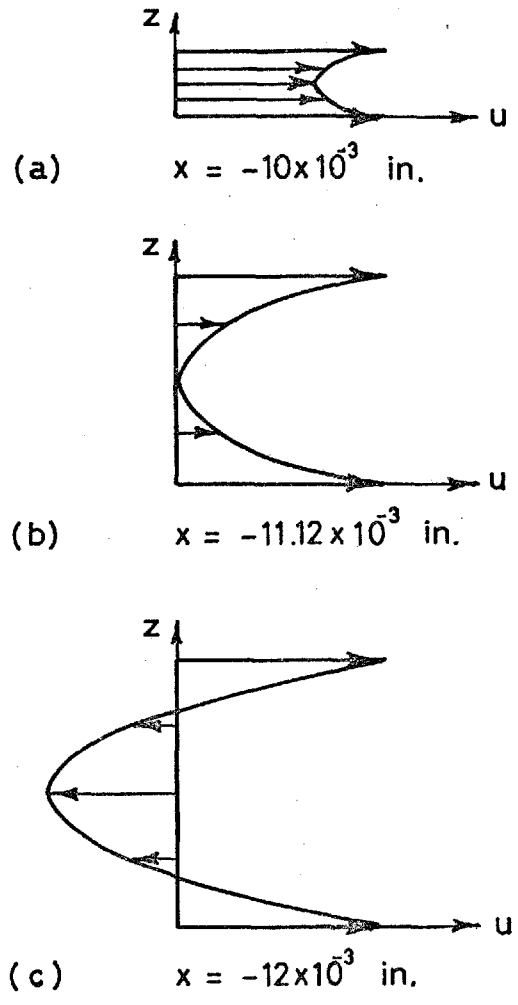


Fig.47 Fluid velocity distribution at different locations within the inlet region

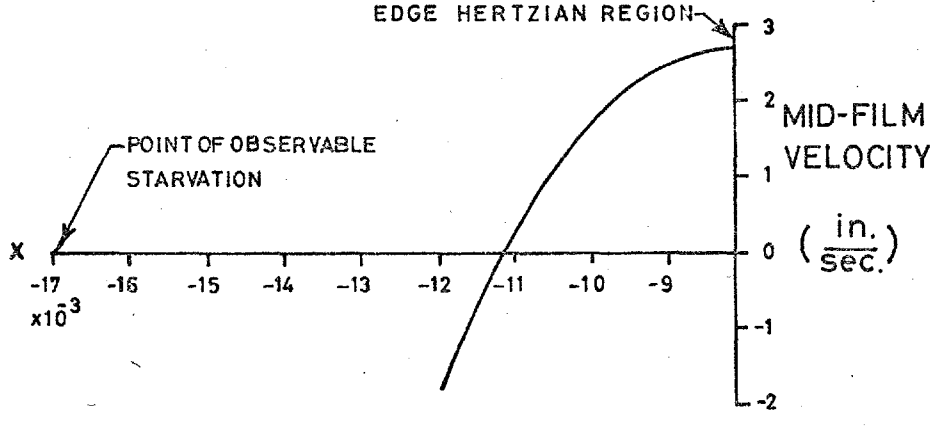


Fig.48 Variation of mid-film velocity within the inlet region

of the location where $u = \frac{\partial u}{\partial z} = 0$, indicates that backflow occurs over a portion of the inlet pressure buildup which is essential for the establishment of a flooded condition. The Lauder boundary condition does not appear, therefore, to provide a very good criteria for starvation. There is no physical reason why the pressure should be zero where the velocity and velocity gradient are zero. The adverse pressure gradient can cause the mid-film velocity to become negative as well as zero.

The state of affairs in the inlet region becomes apparent if one considers the volume of lubricant required in the inlet region to obtain a flooded condition, along with the volume rate of flow. For the above example, the required volume in the inlet region along an elemental width dy on the centre line in the direction of rolling is $8.2 \times 10^{-5} dy$ in.^{3}) and the volume rate of flow ($q = uh_o \cdot dy$) is $0.5 \times 10^{-5} dy$ in^3/sec. The time required for this volume to pass through the Hertzian region, neglecting side-leakage, is 16.4 seconds. The transit time for the fluid adjacent to the surfaces to travel from the inlet boundary to the exit region is only 0.0091 seconds. Thus, while the fluid adjacent to the surfaces is moving rapidly toward the Hertzian region, the central core of the fluid in the inlet region must remain relatively fixed (see also Hartung, ref.132). According to Tipei(133) this core forms a swirl which appears just upstream of the point where negative velocities occur. For pure rolling this swirl develops into two bounded vortices which may look like that found in fig.49. The figure, which is not drawn to scale, is actually very flat as the inlet distance S must be greater than $500h_o$. Tipei suggests that the line AB passing through the centre of the

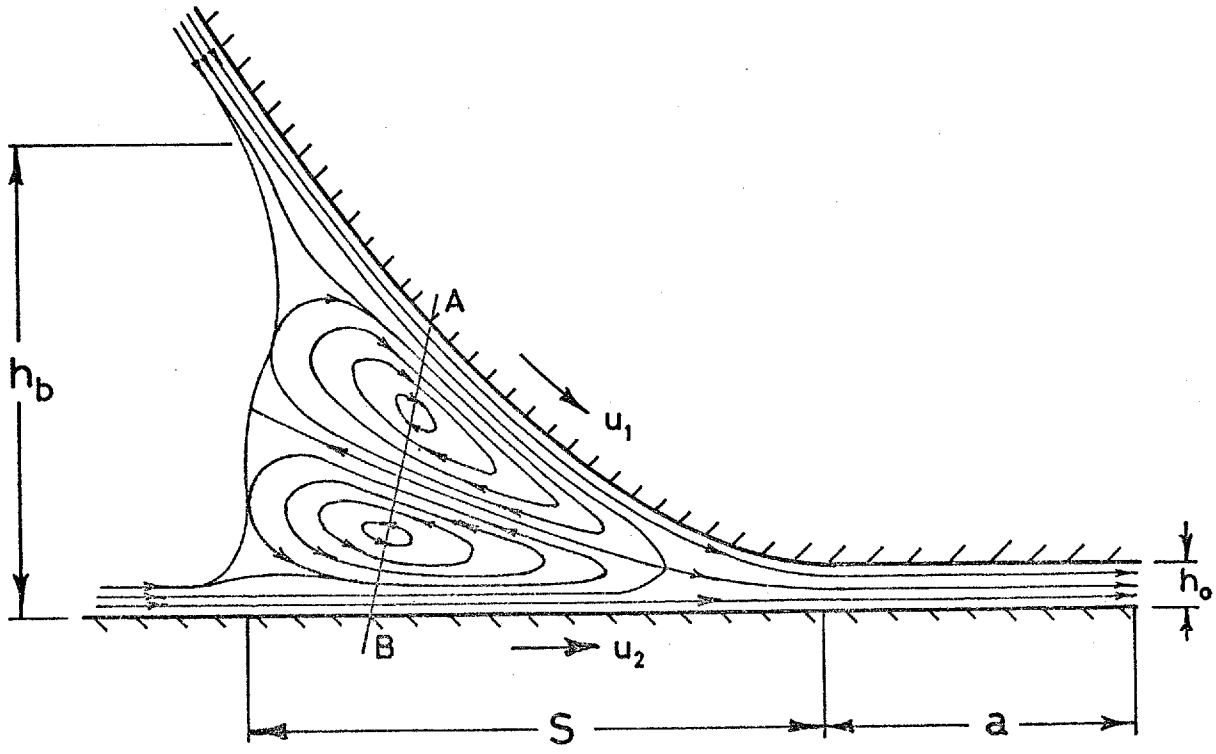


Fig.49 Flow spectrum in inlet region, after Tipei(133)

two vortices is the upstream limit of the fluid film where Reynolds equation applies. He furthermore assumes that this location defines the condition where p = 0. Using potential flow theory Tipei derives the following expression for the pressure gradient at the location of the vortex centres and under pure rolling conditions.

$$\frac{dp}{dx} = \frac{32}{3} \frac{\eta u_2}{h^2}$$

Substituting in Reynolds equation

$$\frac{32}{3} \frac{\eta u_2}{h^2} = 12\ u_2\ \eta \left(\frac{h-h_o}{h^3}\right)$$

gives

$$\frac{h}{h_o} = 9$$

This is in better agreement with the location of the onset of starvation in the above example than the Lauder boundary condition where $h/h_o = 3$.

It will be shown later, in section 6.9, that the particular fluid velocity conditions described by Lauder and Tipei do not appear to be entirely applicable to the starvation problem. The ratio h/h_o, however, is considered further.

If the onset of starvation in the above example agrees with $h/h_o = 9$, then it is possible that the same condition has general application. The central film thickness for all the starvation tests are plotted in fig.50 as a function of h_b/h_o where h_b is the gap thickness at the lubricant boundary(see fig. 49). It was not possible to measure h_b with interferometry for large inlet distances; therefore, it was computed from

$$h_b = h_o + h_H$$

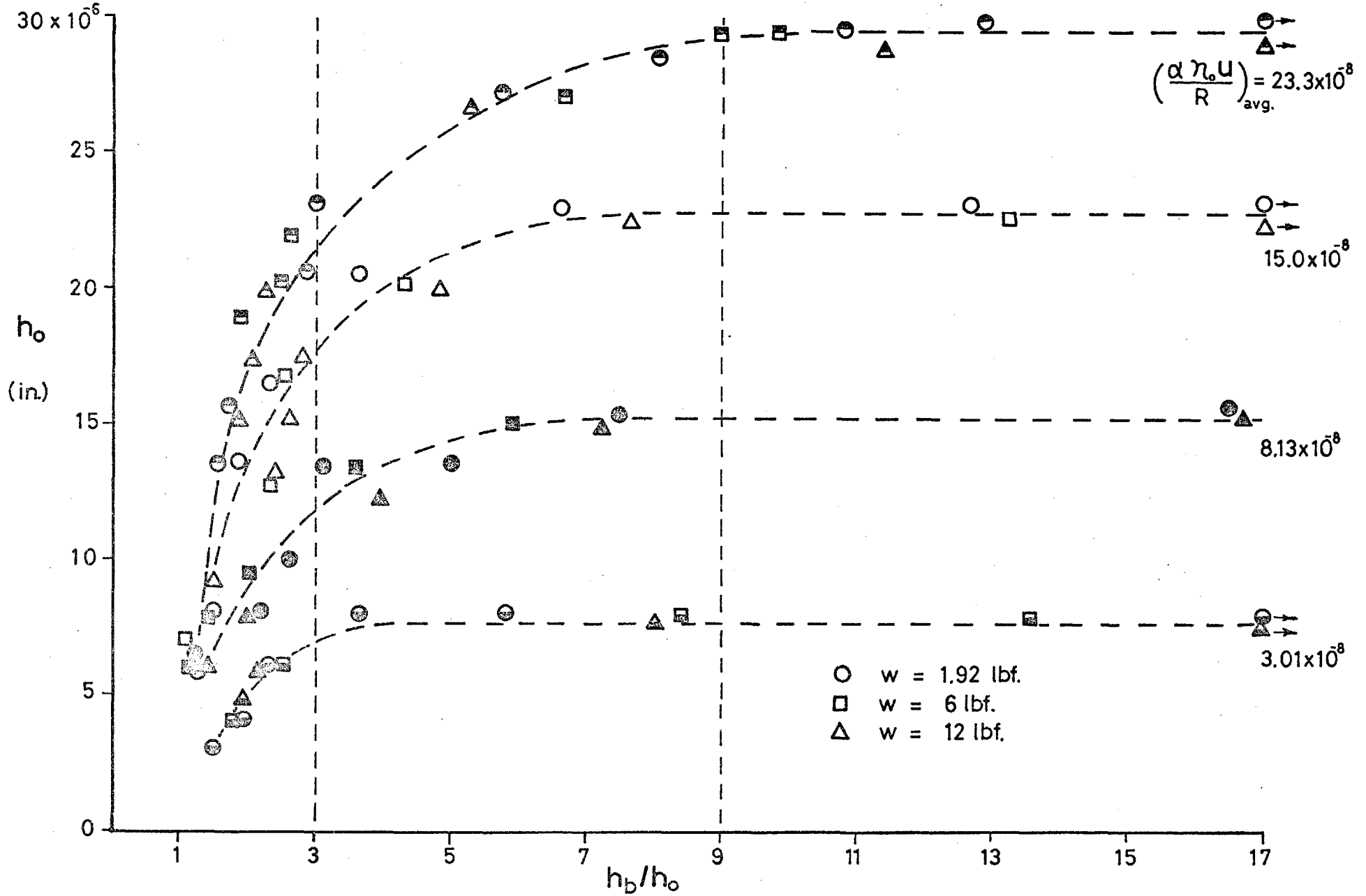


Fig.50 Plot of film thickness against the inlet geometry ratio h_b/h_o

where h_H is the gap thickness for dry contact given by equation 3.6. All the test runs show that starvation commences approximately where $h_b/h_o = 9$, thus suggesting its adoption as a criteria for the onset of starvation. This will be considered further in the next section.

At the location of $h_b/h_o=3$, the film thickness has reduced significantly from its flooded condition; and it should be noted that for each test run, the film thickness has reduced by a constant amount equal to about 22%.

The experimental results also show that the location of the flooded condition is not very well defined. That is, the film thickness approaches the flooded condition asymptotically in the upstream direction. Since film thickness is related to the amount of pressure generated in the inlet region, it should than follow that under flooded conditions the pressure also approaches zero asymptotically in the upstream direction. This can also be concluded from the integrated form of Reynolds equation which shows that dp/dx approaches zero as the gap thickness, h, becomes very large. Therefore, the location of p = 0 will not be very well defined, and according to the experimental results of fig.50 the solution of Reynolds equation should not be effected by the location of p = 0, provided it is at a gap thickness of at least $9h_o$.

Under starvation conditions the results show that h_b/h_o decreases and approaches one as the central film thickness approaches zero. Since the ratios $h_b/h_o = 3$ and $h_b/h_o = 9$ define particular fluid velocity distributions, the results imply that when the inlet region is starved the velocity distribution at or near the lubricant boundary is a function of

the boundary location. For example, both positive and negative velocities occur when the inlet distance gives an h_b/h_o value greater than 3; and only positive velocities occur when the inlet distance gives an h_b/h_o value smaller than 3. Under these conditions it is suggested that p = 0 at the location of the upstream limit of the fluid film where it completely fills the gap between the bearing surfaces and it is assumed that Reynolds equation applies to the continuous film bordered by this upstream limit.

6.7 Starvation Criteria

To predict film thickness in a practical situation it must be established whether or not the bearing is operating in a starved or flooded condition. It was suggested in the previous section that a starvation criteria can be defined by the ratio h_b/h_o. For practical applications a starvation criteria may be more useful in terms of the inlet distance S. This is developed below.

The ratio h_b/h_o can be written as

$$\frac{h_b}{h_o} = \frac{h_o + h_H}{h_o} \qquad \ldots(6.2)$$

For point contact, h_H can be approximated by (see ref.30, p.195)

$$h_H = \frac{ap_{max}}{E'} \left[3.81 \left(\frac{x_b}{a} - 1 \right)^{1.5} \right] \qquad \ldots(6.3)$$

This relation is accurate for small ratios of x_b/a. Fig.41 shows that starvation occurs within $x_b/a = 4.5$ for all the test runs. The calculation of h_H from equation 6.3 for $x_b/a=4.5$ is about 14% less than that calculated from the Hertzian relation given by equation 3.6. For most of the test runs,

starvation occurs within an x_b/a ratio of 3 or 2 where the error is 7.6% and 2% respectively.

Now we can write

$$\left(\frac{x_b}{a} - 1\right) = \frac{S}{a} \qquad \ldots(6.4)$$

and

$$\frac{p_{max}}{E'} = \frac{a}{\pi R} \qquad \ldots(6.5)$$

Substituting equation 6.4, 6.5 and 6.3 in equation 6.2 gives

$$\frac{h_b}{h_o} = 1 + 1.21 \frac{a^{\frac{1}{2}} S^{\frac{3}{2}}}{Rh_o}$$

Solving for the inlet distance S, gives

$$S = \left(\frac{h_b/h_o - 1}{1.21}\right)^{\frac{2}{3}} \frac{(Rh_o)^{\frac{2}{3}}}{a^{\frac{1}{3}}}$$

We will define the onset of starvation by $h_b/h_o = 9$. It must be remembered that this ratio is not well defined. The experimental results of fig.50 show that the ratio h_b/h_o could have been chosen anywhere between 7 and 12. Using $h_b/h_o=9$, the minimum inlet distance S_f to obtain a flooded condition can be approximated by

$$S_f = \frac{3.52\,(Rh_o)^{\frac{2}{3}}}{a^{\frac{1}{3}}} \qquad \ldots(6.6)$$

This is very similar to the statement of Fein and Kreutz (see section 4.12b) that the vast majority of pressure buildup for an ehd contact occurs within about

$$\frac{3\,(Rh_o)^{\frac{2}{3}}}{a^{\frac{1}{3}}}$$

ahead of the Hertzian flat. They give no details, however, about the derivation of this.

The starvation criteria given by equation 6.6 is plotted in fig.51 for each test load. It compares very well with the experimental results. A feature of the starvation criteria is that for a constant central film thickness, starvation will commence at slightly smaller inlet distances when the load increases. The experimental data also reveals this tendency. This variation with load occurs because the gap thickness increases with load for a constant S and h_o, thus making the ratio h_b/h_o larger.

The evaluation of S_f from equation 6.6 presupposes a knowledge of h_o under flooded conditions. It was shown in section 5.5 that the central film thickness for point contact agrees very well with the theoretical film thickness formula of Evans

$$\frac{h_o}{R} = 1.73 \left(\frac{\alpha \eta_o u}{R}\right)^{.714} \left(\frac{w}{E'R^2}\right)^{-.048} \qquad \ldots(6.7)$$

Now,

$$a = \left(\frac{3}{2} \cdot \frac{wR}{E'}\right)^{\frac{1}{3}} \qquad \ldots(6.8)$$

When equation 6.7 and 6.8 are substituted in equation 6.6, the starvation criteria can be written as

$$\frac{S_f}{R} = 4.84\left(\frac{\alpha \eta_o u}{R}\right)^{.476} \left(\frac{w}{E'R^2}\right)^{-.143} \qquad \ldots(6.9)$$

This relation or that of equation 6.6 defines a starvation criteria in terms of the inlet distance for point contacts operating under ehd conditions. This will be used in the next section for the prediction of film thickness under starvation conditions.

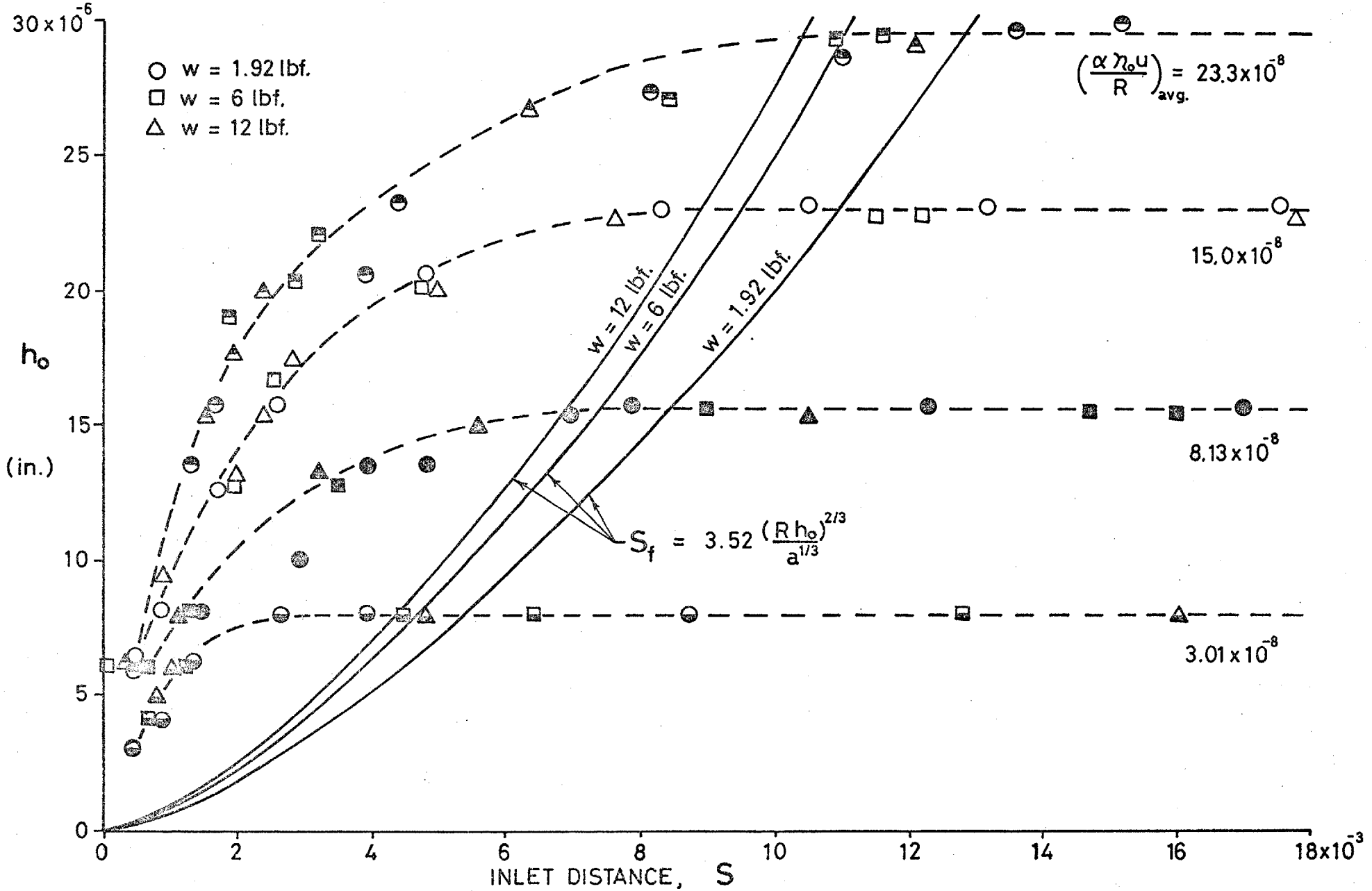


Fig.51 Comparison of starvation criteria with experimental results

6.8 Prediction of Film Thickness Under Starvation Conditions

It was noted in fig.50 that, while the ratio $h_b/h_o=9$ defines the approximate boundary location for the onset of starvation, the ratio $h_b/h_o=3$ describes the boundary location where the film thickness for each test run has diminished from its flooded condition by a constant amount equal to about 22%. This suggests that h_b/h_o defines the degree of starvation for all operating conditions.

Fig.52 shows all of the experimental results plotted as the ratio $h_o/(h_o)_f$ against h_b/h_o where $(h_o)_f$ is the film thickness under flooded conditions. Within the range of experimental scatter the results can be described by a single curve. This curve illustrates some of the general features mentioned before. Namely, that the flooded condition is approached asymptotically, and can not, therefore, be very well defined; and that film thickness approaches zero as the lubricant boundary approaches the Hertzian radius, i.e. when $h_b/h_o = 1$.

If the results of fig.52 have general applications, then it can be used to predict film thickness under starvation conditions. That is, for a given degree of starvation defined by h_b/h_o (or more practically, in equivalent terms of S as shown below), film thickness must be reduced by a constant amount according to fig.52. This constant is defined by

$$K_s = \frac{h_o}{(h_o)_f}$$

and can be employed in the film thickness formula derived by Evans. Thus,

$$\frac{h_o}{R} = 1.73\ K_s \left(\frac{\alpha \eta_o u}{R}\right)^{.714} \left(\frac{w}{E'R^2}\right)^{-.048}$$

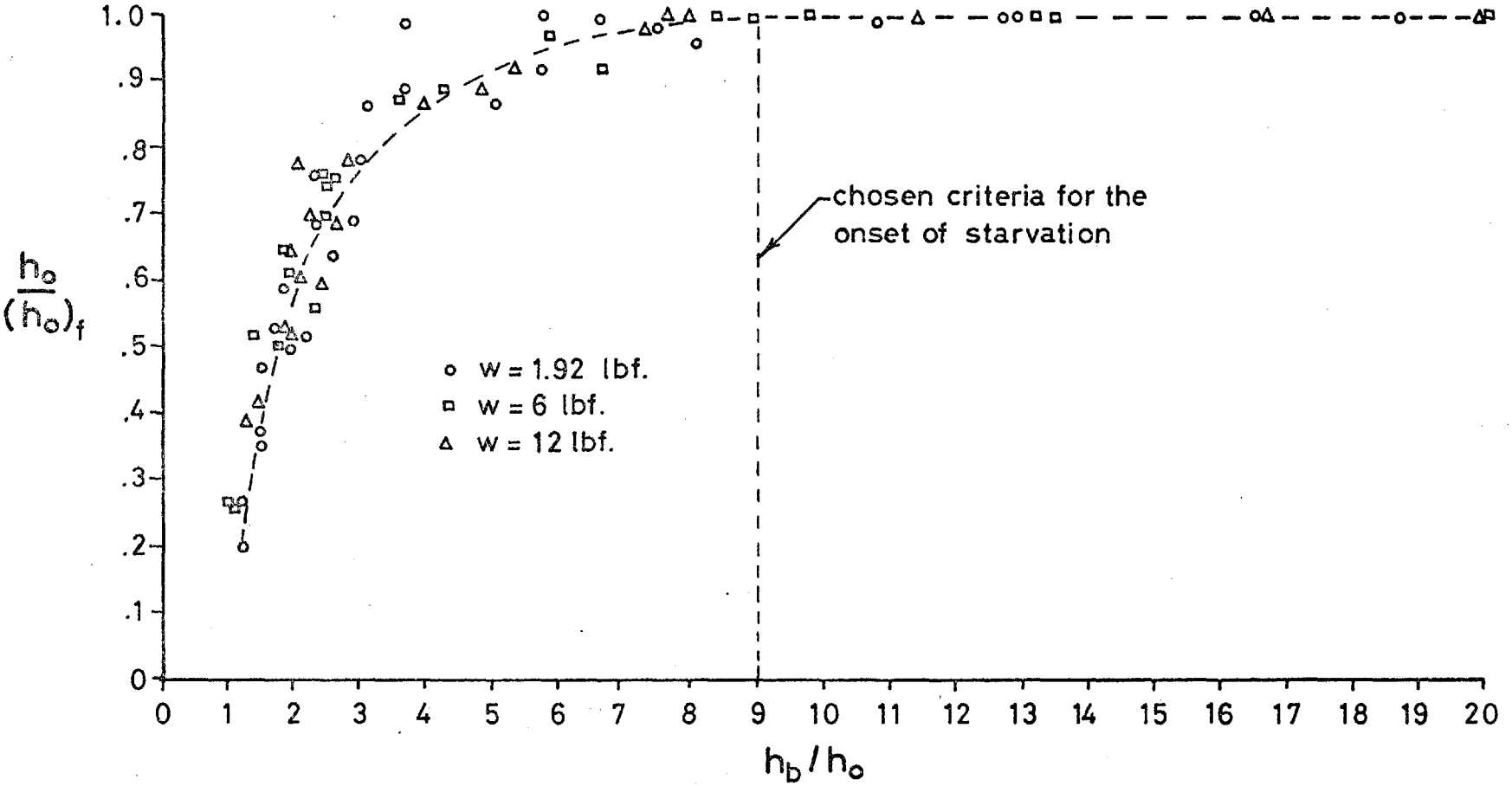


Fig.52 All the starvation data plotted in dimensionless film thickness ratios

which, for a given degree of starvation, implies that film thickness is dependent upon the operating conditions in the same manner as found under flooded conditions.

The dependence of film thickness on the speed-viscosity parameter for constant values of h_b/h_o is shown in fig.53. The graph was obtained by cross-plotting the curves of fig.50 where the speed-viscosity parameter represents an average taken over the three test loads. It can be seen that for various degrees of starvation ($h_b/h_o < 9$) film thickness increases with the speed-viscosity parameter in the same manner as found under the flooded condition ($h_b/h_o = 9$). This is in contrast with fig.46 where the inlet distance S was used to describe the degree of starvation. It shows that film thickness becomes less sensitive to the speed-viscosity parameter as the inlet distance is reduced.

For practical applications h_b/h_o cannot be used as a measure of the degree of starvation since it presupposes a knowledge of h_o which, under starvation conditions, is the unknown which is pursued. It can be shown that h_b/h_o is a function of S/S_f where S_f is defined by equation 6.6 or 6.9. All the starvation results are plotted in fig.54 where S/S_f, as a starvation parameter, now represents the degree of starvation. The experimental data can be approximated by

$$K_s = 1.24 \left(\frac{S}{S_f}\right)^{.45}$$

for values of $S/S_f < 0.62$. The film thickness formula then becomes

$$\frac{h_o}{R} = 2.14 \left(\frac{S}{S_f}\right)^{.45} \left(\frac{\alpha \eta_o u}{R}\right)^{.714} \left(\frac{w}{E'R^2}\right)^{-.048} \quad \ldots(6.10)$$

This semi-empirical formula can be used to predict the central film thickness for point contacts under starvation conditions.

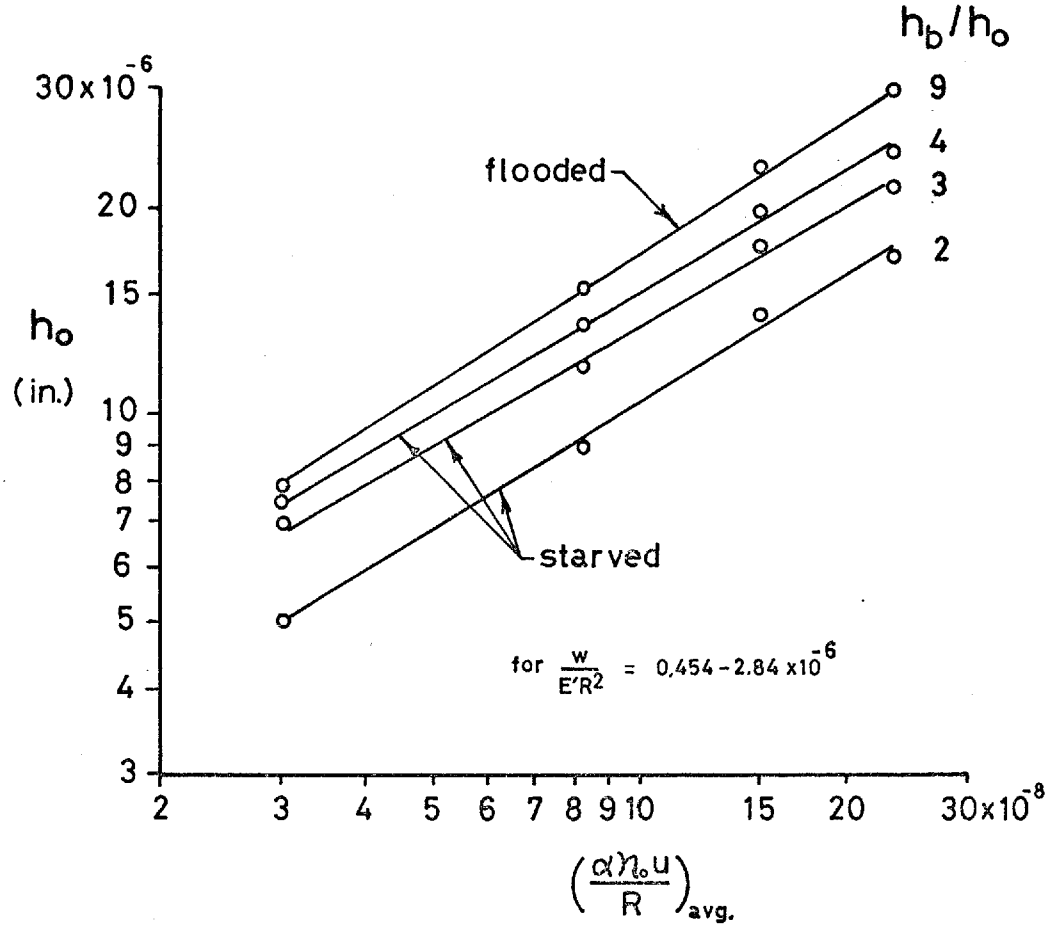


Fig.53 Variation of film thickness with the speed-viscosity parameter for constant values of h_b/h_o

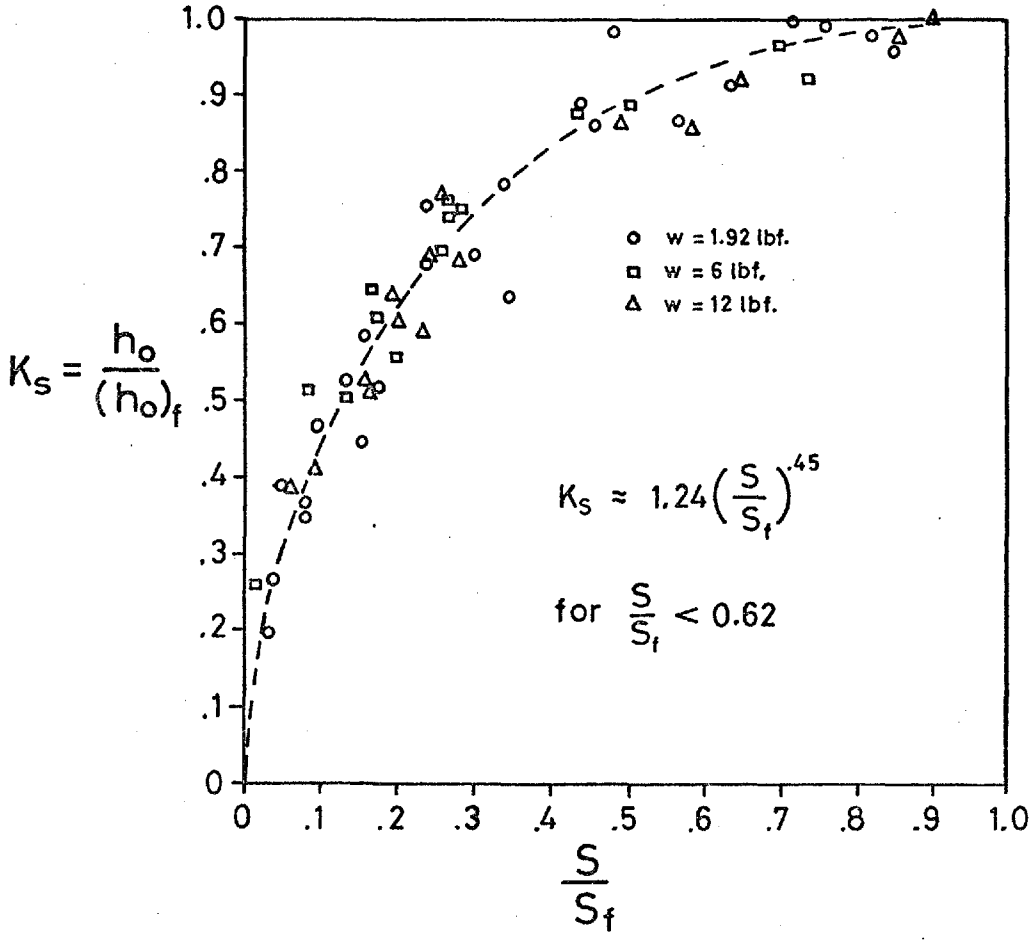


Fig.54 Variation of K_s with the starvation parameter

In section 5.5 an empirical relation was developed for estimating the minimum film thickness (h_m) from the central film thickness. This relation does not apply under starvation conditions. For a given central film thickness the pressure distribution and film shape will always be more Hertzian if the bearing is starved. This causes the lateral and rear constrictions to become less pronounced. The empirical relation will, therefore, underestimate the minimum film thickness.

The greatest practical problem in applying equation 6.10 is the determination of the inlet distance for the particular application. The inlet distance is difficult to measure directly and difficult to predict from external conditions. Besides the method and amount of lubricant supplied, these external conditions must involve those aspects which deal with the transport and distribution of lubricant within the bearing; e.g. viscosity, surface tension, bearing dynamics, and geometry. There is certainly a great need for practical research along these lines.

While the direct application of equation 6.10 may await further theoretical study, the application of the general features of the results can certainly be useful.* These may help to avoid starvation or to control it. For example, in dealing with the 'oil jag' problem discussed in section 4.12a, it is important to remember that film thickness at first rises sharply as the inlet distance increases from zero and then rises slowly, approaching the flooded condition asymptotically.

* These general features have recently been used to redesign a gyro-bearing used in the guidance systems of NASA's Apollo programme.

Therefore, if the bearing can be operated nearer the flooded condition only small changes in film thickness will accompany movement of excess oil onto the bearing raceway.

6.9 Discussion of Starvation Criteria

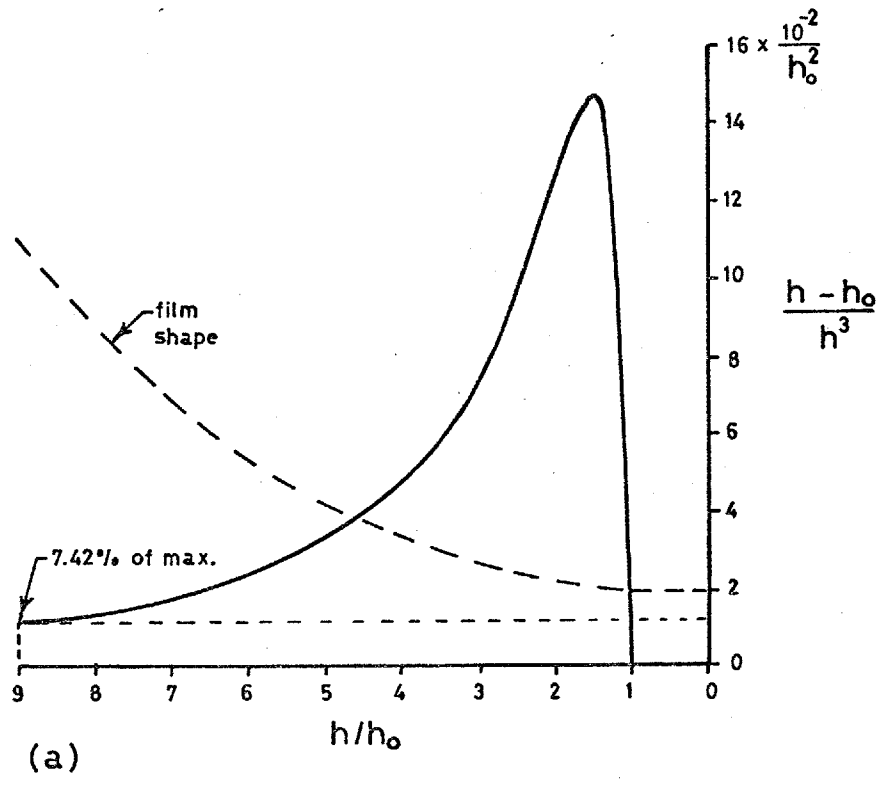
It was noted in the previous two sections that the starvation results correlate very well with the ratio h_b/h_o. It was also noted that the ratio h/h_o defines particular fluid velocity conditions; e.g. for pure rolling $h/h_o = 3$ is the location of zero mid-film velocity; and it can be shown that $h/h_o = 9$ is the location of the vortex centres predicted by Tipei.

Some insight into the significance of h/h_o can be seen from the integrated form of Reynolds equation written in terms of the reduced pressure q.

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

For constant operating conditions the reduced pressure gradient is a function of the geometry term $(h-h_o)/h^3$. This is plotted as a function of h/h_o in fig.55a where it can be seen that the reduced pressure is primarily generated in the region around $h/h_o = 1.5$. Pressure generation diminishes in the upstream direction as h/h_o increases.

The reduced pressure curve is shown schematically in fig.55b. It has an inflection point at $h/h_o = 1.5$ and reaches a maximum of $1/\alpha$ where $h/h_o = 1$. The actual pressure curve will be similar to that of fig.55b for only large values of h/h_o where the pressures are low. For smaller values of h/h_o the actual pressure becomes significantly higher than the reduced pressure due to the enhancement of viscosity



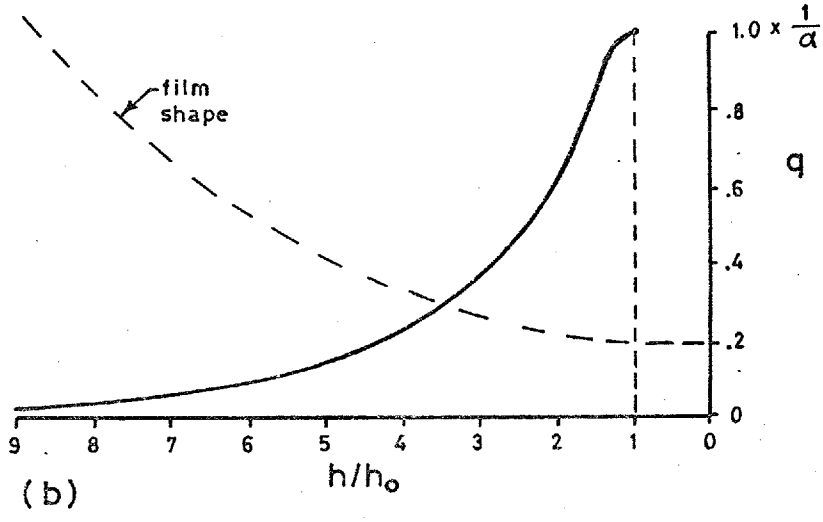


Fig.55 (a) Geometry function, and (b) reduced pressure as a function of h/h_o

with pressure. It is this hydrodynamic pressure which is primarily responsible for film thickness. It will blend in with the Hertzian pressure when h/h_o is approximately equal to one, and will slowly diminish in the upstream direction as h/h_o increases. When $h/h_o = 9$, fig.55a shows that the geometry function, and hence the reduced pressure gradient, is small being only 7.42% of its maximum. The actual pressure gradient will also be small, and it can be shown from the integrated form of Reynolds equation that at $h/h_o = 9$ the pressure gradient is

$$\frac{dp}{dx} = \frac{32}{243} \cdot \frac{u\eta}{h_o^2}$$

The above analysis as well as the experimentation has been restricted to pure rolling conditions. When sliding is introduced, the application of the Tipei boundary condition as a criteria for the onset of starvation appears to meet a major difficulty. For pure sliding only one vortex develops in the inlet region, which can be shown to occur at $h/h_o = 3$. This is in contrast to $h/h_o = 9$ for pure rolling. Therefore, if the location of the Tipei boundary condition is used to define the onset of starvation, then it implies that starvation will commence when the lubricant boundary is further downstream if sliding is introduced.

Reynolds equation implies that equivalent pressure distributions will be generated for sliding or rolling provided the sum of the surface speeds are the same; and, it was shown experimentally by Crook(45) and Dyson et al(60) that sliding has little effect on film thickness. The starvation results presented in section 6.4 suggest that film thickness, whether starved or flooded, is associated with the pressure generated in the inlet region. Therefore, if sliding does

not alter the pressure distribution or film thickness, then the location of the onset of starvation should be the same and the results presented here should apply to both sliding and rolling conditions. It would appear, therefore, that the gap geometry expressed as h_b/h_o is significant in connection with starvation because of the pressure or pressure gradient associated with it, rather than the particular fluid velocity distribution associated with it.

6.10 Extension of Results to Other Operating Conditions

The similarities between ehd line and point contact were pointed out in section 4.8c. It is very likely, therefore, that the features of the starvation results presented here will be similar to the line contact problem. A starvation criteria for ehd line contact can be derived along the lines of the derivation of equation 6.6. For line contact h_H can be approximated for small values of x_b/b by (see ref.30, p.206)

$$h_H = \frac{2\ p_{max}\ b}{E'} \left[2.114 (x_b/b-1)^{1.55} \right] \qquad \ldots(6.11)$$

where b is one-half the Hertzian width.

Now we can write

$$\frac{p_{max}}{E'} = \frac{b}{4R} \qquad \ldots(6.12)$$

and

$$(x_b/b - 1) = S/b \qquad \ldots(6.13)$$

After substituting equation 6.12 and 6.13 in equation 6.11 the ratio h_b/h_o can be written as

$$\frac{h_b}{h_o} = 1 + 1.057 \frac{b^{.45}\ S^{1.55}}{h_o R}$$

solving for S, gives

$$S = \left[\frac{h_b/h_o - 1}{1.057}\right]^{.645} \frac{(Rh_o)^{.645}}{b^{.29}}$$

If the ratio $h_b/h_o = 9$ can also be used to describe the onset of starvation for line contact, then the starvation criteria for the ehd line contact problem can be approximated by

$$S_f = \frac{3.69\ (Rh_o)^{.645}}{b^{.29}} \qquad \ldots(6.14)$$

This is very much like the starvation criteria for point contact given by equation 6.6.

The relation between film thickness and the starvation parameter (S/S_f) for line contact is likely to be similar to that given in fig.54. It may be possible, therefore, to use fig.54 in conjunction with equation 6.14 to estimate film thickness under starvation conditions for the ehd line contact case.

A starvation criteria can also be derived for rigid cylinders lubricated with an isoviscous lubricant. For this case, which is shown schematically in fig.56, it can no longer be assumed that $dp/dx = 0$ at h_o. The Reynolds equation must then be written as

$$\frac{dp}{dx} = 12\ u\eta_o \left(\frac{h-\bar{h}}{h^3}\right)$$

where $\bar{h}$ corresponds to the gap thickness at $dp/dx = 0$. The location of the onset of starvation is now defined as $h_b/\bar{h} = 9$. If the exit boundary condition is taken as $p = dp/dx = 0$, then it can be shown (see ref.30, p.164) that $\bar{h} = 1.226h_o$, giving $h_b/h_o = 11.03$. If the shape of the rigid cylinder is represented by a parabolic profile, h_b can be written as

$$h_b = h_o + \frac{S^2}{2R}$$

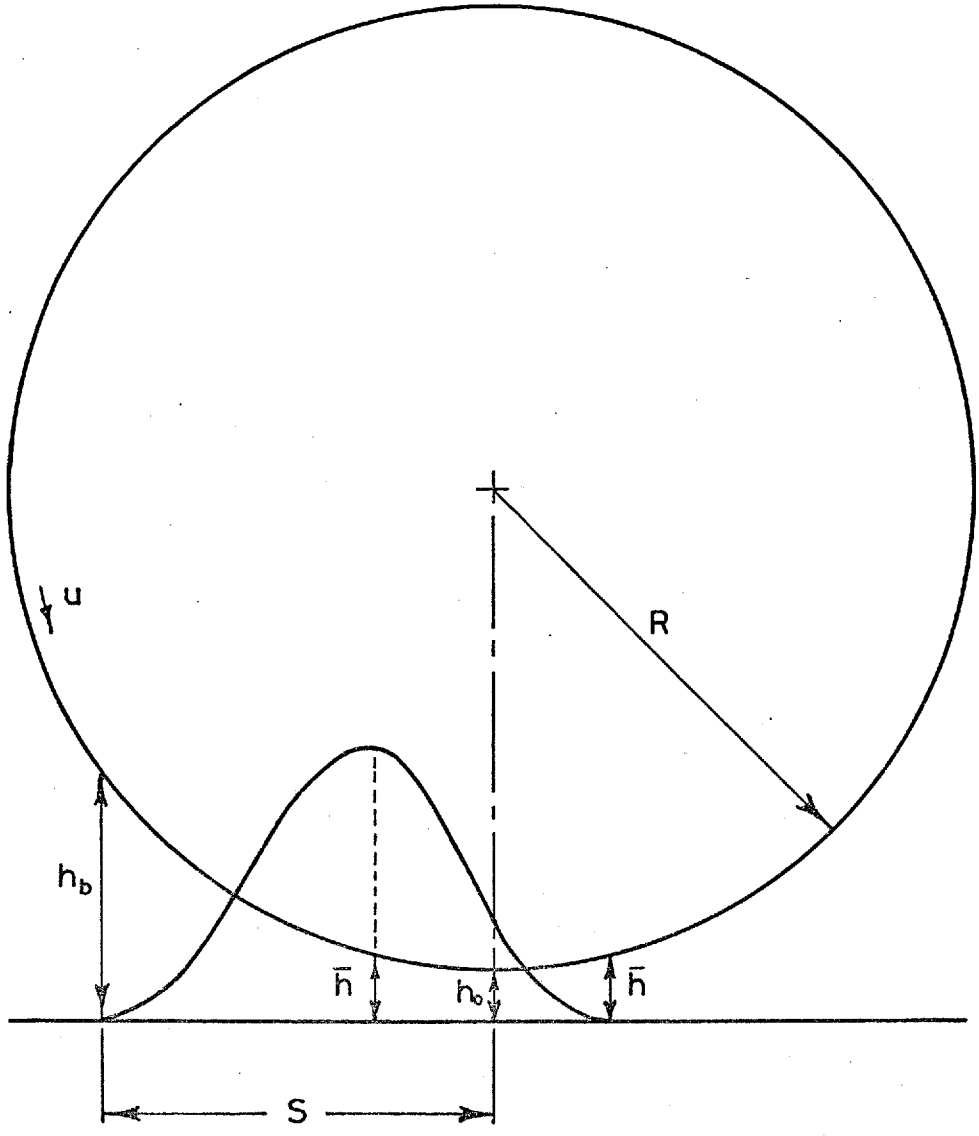


Fig.56 Relevant conditions of pressure and geometry for rigid cylinder

where S now equals x_b. The ratio h_b/h_o then becomes

$$\frac{h_b}{h_o} = 1 + \frac{S^2}{2Rh_o}$$

Solving for S, gives

$$S = \left[\left(\frac{h_b}{h_o} - 1\right) 2Rh_o\right]^{\frac{1}{2}}$$

Using $h_b/h_o = 11.03$ as the criteria for starvation gives

$$S_f = 4.47(Rh_o)^{\frac{1}{2}}$$

When this is compared with the theoretical results of Dowson and Whitaker(42), it shows that over the film thickness range of $h_o/R = 10^{-2}$-10^{-4} the load carried is about 80% of that for a theoretically flooded condition. The agreement may not be as good as the ehd case because under ehd conditions most of the pressure generation occurs close to the conjunction region where pressures are enhanced by increasing viscosities. Therefore, the pressure generated further upstream may be of more relative importance under isoviscous conditions. To establish agreement with theory, the onset of starvation must be defined upstream of $h_b/h_o = 9$. If the prediction of circulation flow (at $h/h_o = 9$) by Tipei is correct, then one must reconsider the assumptions and application of Reynolds equation at remote distances in the inlet region. More work is definitely needed before any conclusions can be made on this subject.

6.11 Simple Starvation Theory

It is possible to approximate the ehd starvation problem with a very simple theory based on Grubin assumptions and the reduced Reynolds equation given by

$$\frac{dq}{dx} = 12u\eta_o \left(\frac{h-h_o}{h^3}\right)$$

where the reduced pressure q is defined as

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

If the location of p = 0 is set at the lubricant boundary, then according to the above equation q also equals zero at the lubricant boundary. When p becomes reasonably large, q approaches $\frac{1}{\alpha}$. It must be very close to $\frac{1}{\alpha}$ at the Hertzian radius. The reduced pressure curve is drawn schematically in fig.57.

For a constant speed u, dq/dx is a function of the inlet geometry term $(h-h_o)/h^3$ as was shown in fig.55a. A maximum occurs at $h = 3/2h_o$. This is the location in the inlet region where dq/dx is maximum. It therefore defines the point of inflection on the reduced pressure curve. Substituting $h = 3/2h_o$ in the reduced Reynolds equation gives

$$h_o = \frac{4}{3}\left[\frac{\eta_o u}{(dq/dx)_i}\right]^{\frac{1}{2}} \qquad \ldots(6.15)$$

where $(dq/dx)_i$ is the slope of the reduced pressure curve at the point of inflection.

As the inlet distance decreases, the slope $(dq/dx)_i$ should increase, and the film thickness predicted by equation 6.15 will decrease. For convenience, the reduced pressure gradient at the point of inflection is assumed to be the slope defined by the straight line drawn from the lubricant boundary to the point on the Hertzian radius where $q = \frac{1}{\alpha}$. This is shown in fig.57. Therefore,

$$\left(\frac{dq}{dx}\right)_i = \frac{1}{\alpha S}$$

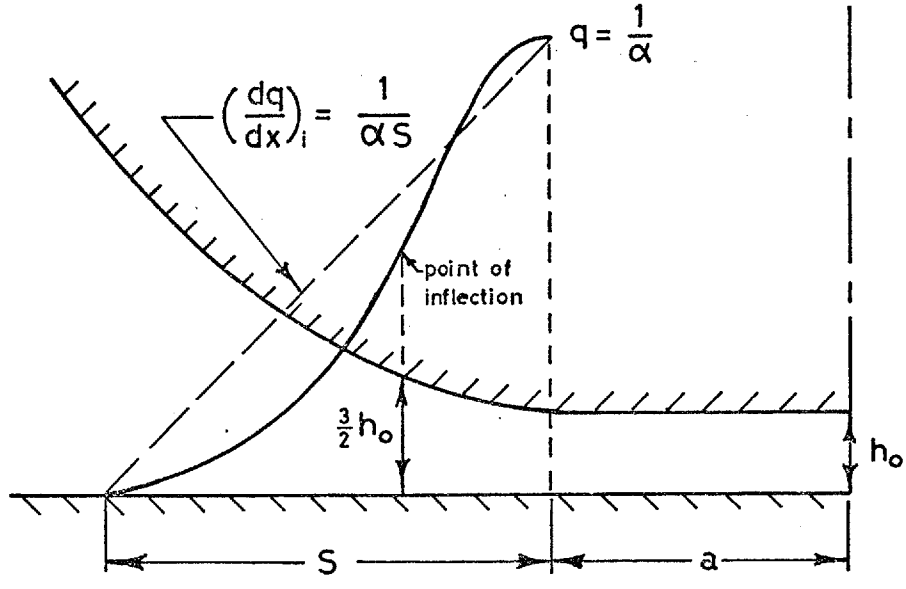


Fig.57 Schematic drawing of the reduced pressure curve and the assumed slope for the simple starvation theory

Substituting, we get

$$h_o = \frac{4}{3} (u\eta_o \alpha S)^{\frac{1}{2}} \qquad \ldots(6.16)$$

The advantage of defining the slope in the above manner is that when equation 6.16 is divided by R, we then obtain a film thickness formula for starvation in terms of three dimensionless groups.

$$\frac{h_o}{R} = \frac{4}{3} \left(\frac{\alpha\eta_o u}{R}\right)^{0.5} \left(\frac{S}{R}\right)^{0.5} \qquad \ldots(6.17)$$

The accuracy of this formula is limited by the assumptions inherent in the Grubin theory and the assumption of no side leakage. The latter will cause the theory to overestimate the film thickness. More important, however, is the assumed definition of $(dq/dx)_i$. This is likely to be smaller than the actual reduced pressure gradient, and it will also tend to overestimate the film thickness, particularly when the inlet region is only moderately starved. The formula does not apply when the inlet region is no longer starved. Indeed, h_o approaches infinity when S approaches infinity.

Under fully flooded conditions, film thickness can be obtained from the Evans formula (equation 6.7). We now have two equations for film thickness, one for starvation, and the other for fully flooded conditions. By equating the two and solving for the inlet distance, we can then define a starvation criteria which will determine when the starvation formula (equation 6.17) applies and when the flooded formula(equation 6.7) applies. The conditions are starved when $S/R < S_f/R$ and flooded when $S/R > S_f/R$ where

$$\frac{S_f}{R} = 2.25\left(\frac{\alpha\eta_o u}{R}\right)^{.488} \left(\frac{w}{E'R^2}\right)^{-.156}$$

The similarity between this equation and equation 6.9, which was derived in section 6.7, should be noted.

Fig.58 is a plot of the central film thickness against the inlet distance for the four starvation tests corresponding to the 6 lbf. load. The solid lines are theoretical, and the dashed lines are experimental. As expected, the theory for starvation overestimates the film thickness, particularly when the inlet region is only moderately starved.

6.12 Comparison of Starvation Results with an Approximate Solution for Point Contact

An approximate solution for point contact employing Grubin assumptions has been computed by Evans(79) using a similar approach to that given by Cameron and Gohar(15). The film thickness formula derived from this solution was presented in section 4.8b and was found in section 5.5 to agree with experiment.

Evans has recently extended the computation to include different locations of the inlet boundary condition where $p = 0$ so that a comparison could be made with the experimental results presented here. From the results of his solution, film thickness was computed for the same operating conditions which were used in the twelve experimental test runs. Some of the relevant details of the solution are given in the Appendix. Only the results are presented here.

The theoretical results for each of the three test loads are shown in figs.59, 60, and 61. The data points which are connected by the solid lines are theoretical, and the dashed lines represent the experimental measurements. The theoretical results show that film thickness diminishes with

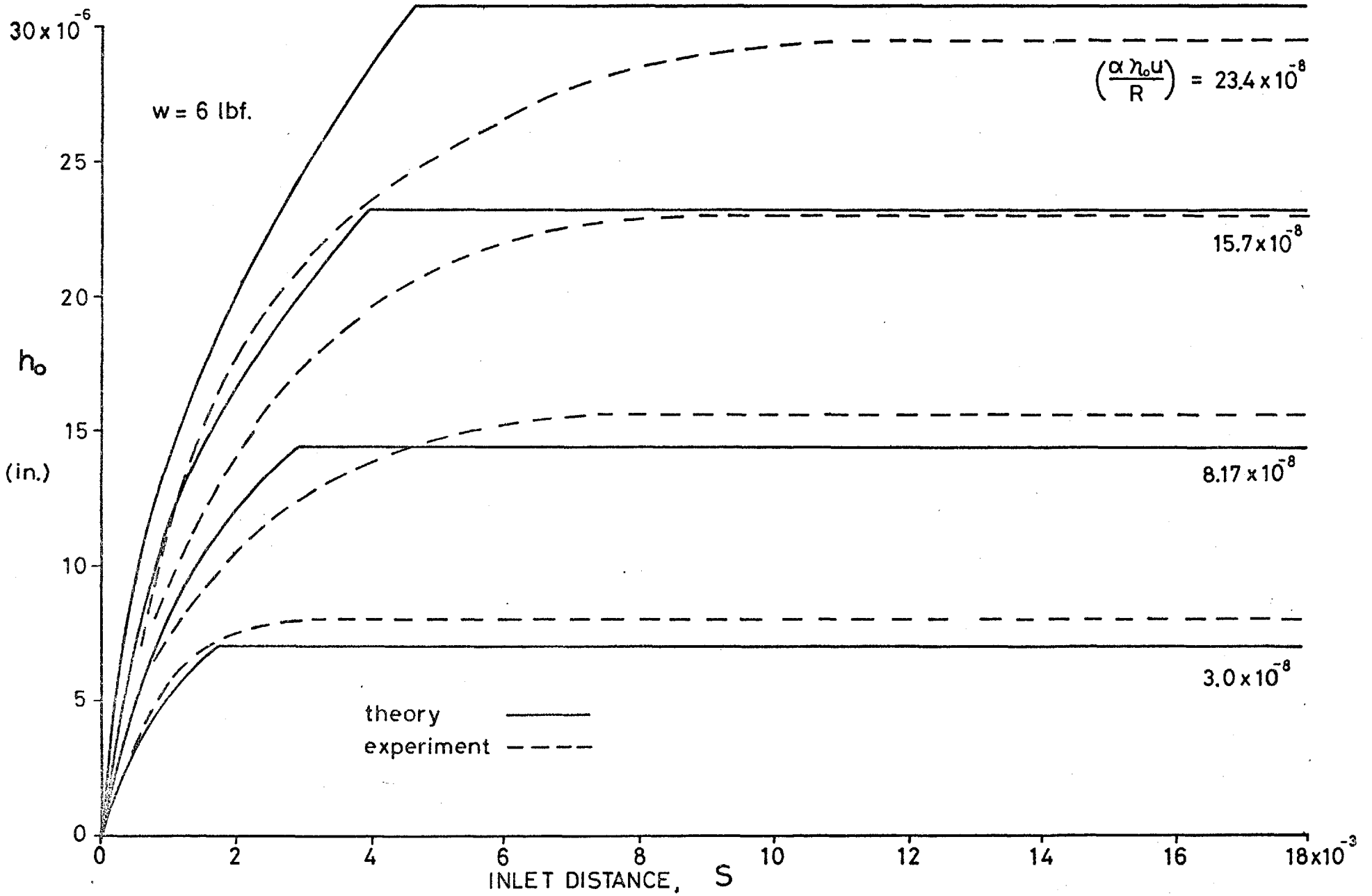


Fig.58 Comparison of simple starvation theory with experimental results

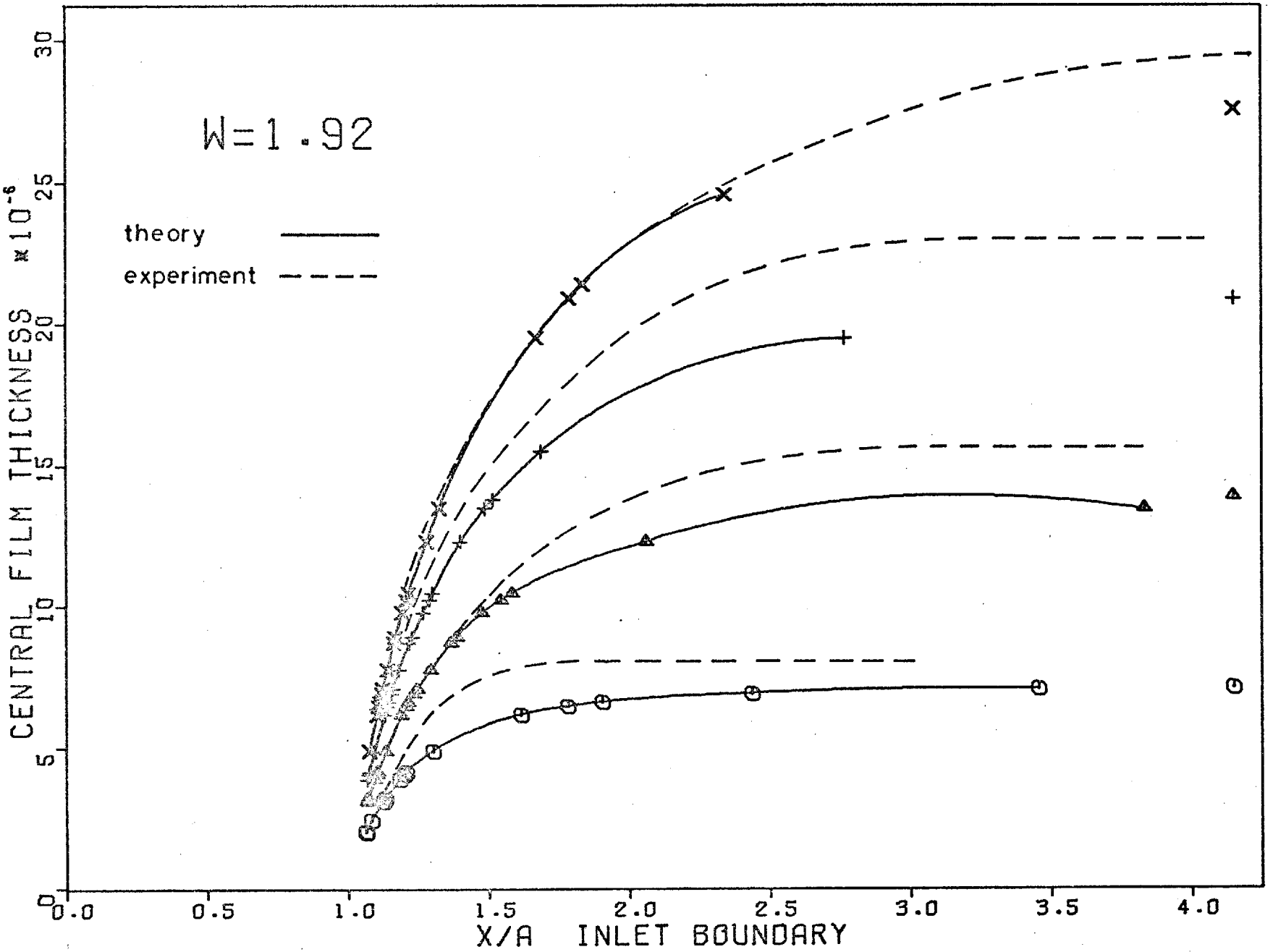


Fig.59 Comparison of theory with experiment for w = 1.92 lbf.

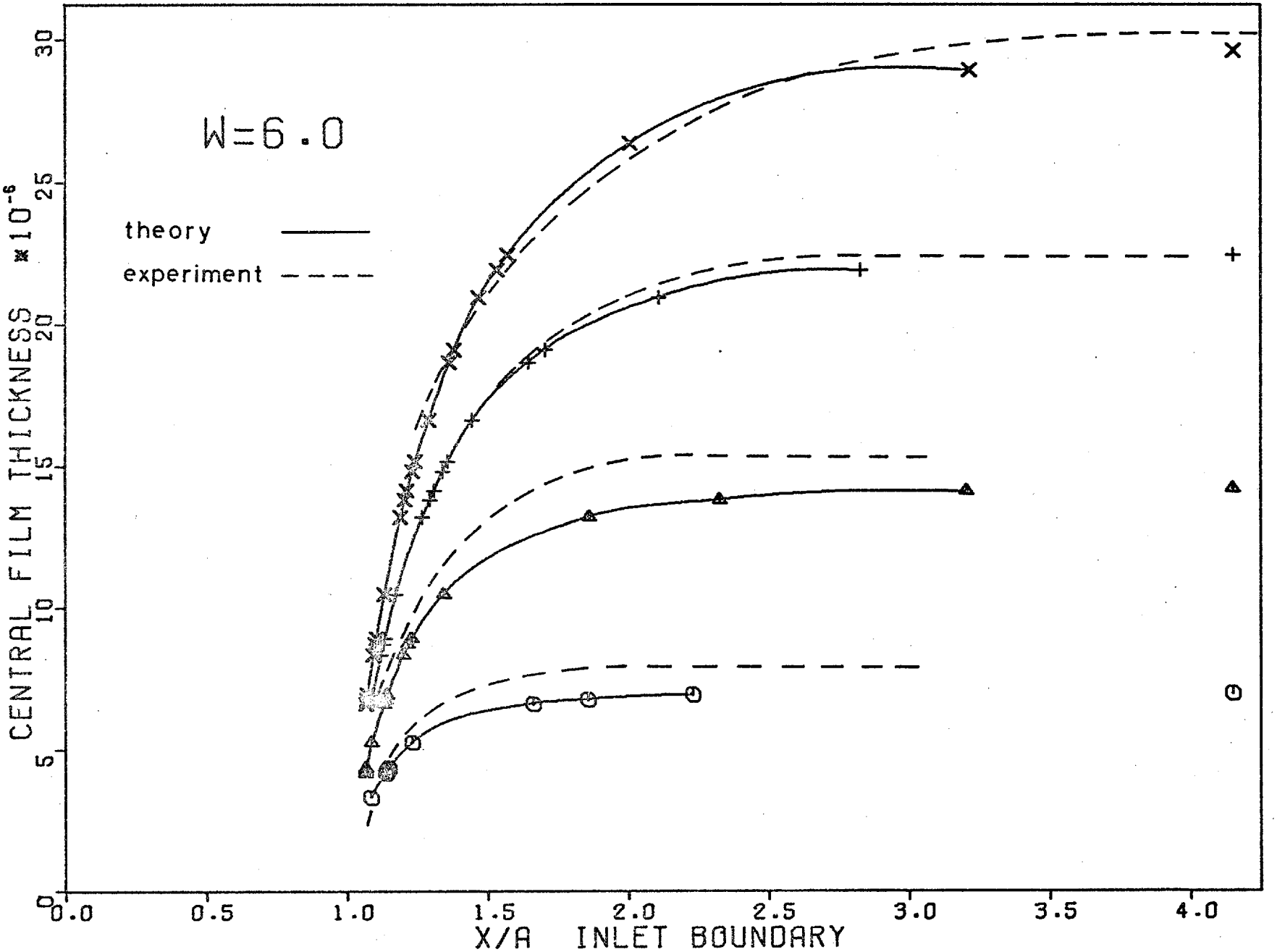


Fig.60 Comparison of theory with experiment for w = 6 lbf.

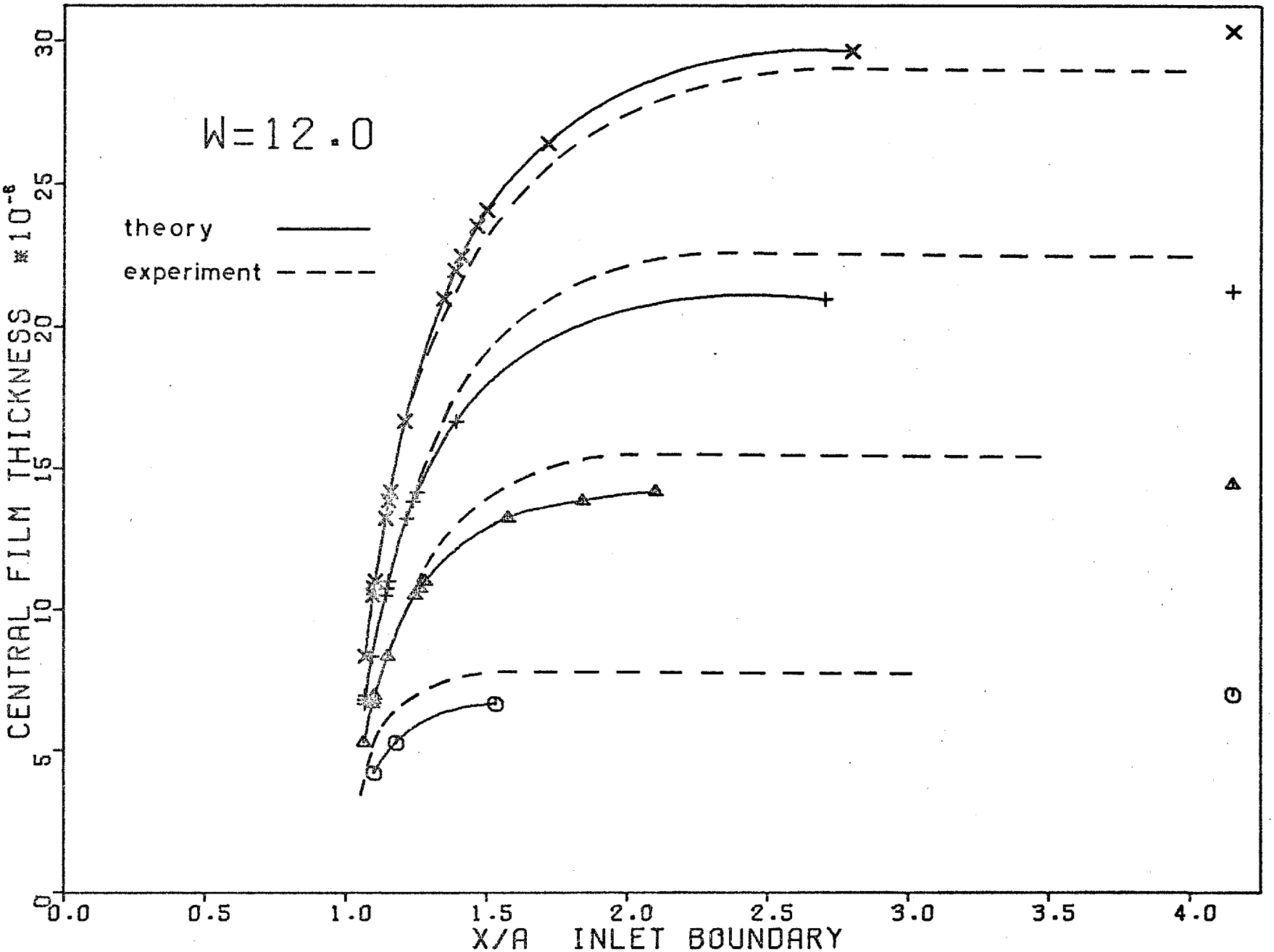


Fig.61 Comparison of theory with experiment for w = 12 lbf.

the inlet boundary in the same manner as was observed experimentally, thus emphasizing the same general features which were pointed out before. The data points on the right-hand side of each graph indicate the theoretically flooded condition for each curve, i.e. for p = 0 located at large values of x/a.

The agreement between theory and experiment is quite good when one considers the scatter in the experimental data together with the Grubin assumptions inherent in the theoretical solution. It must also be remembered that the film thickness predicted by theory is subject to the accuracy of the measured experimental operating conditions, particularly the lubricant properties of α and η_o. It should also be noted that the agreement between theory and experiment is no worse for starvation conditions than it is for flooded conditions. In fact, it is likely to be better under starvation conditions since the film shape becomes more like the assumed Hertzian shape when the inlet region is starved.

A theoretical plot of $h_o/(h_o)_f$ against h_b/h_o is shown in fig.62 along with the experimental data. There is very good agreement between theory and experiment except near the flooded condition. The reason for this is not known. The significance of the theoretical result, however, is that starvation, over a range of loads and initial film thicknesses, can be described by a single curve. This emphasizes the significance and generality of the inlet geometry ratio h_b/h_o.

6.13 Effect of Starvation on Other Aspects of Elastohydrodynamics

While a starved inlet region influences film thickness, it also influences other aspects of ehd which may be important.

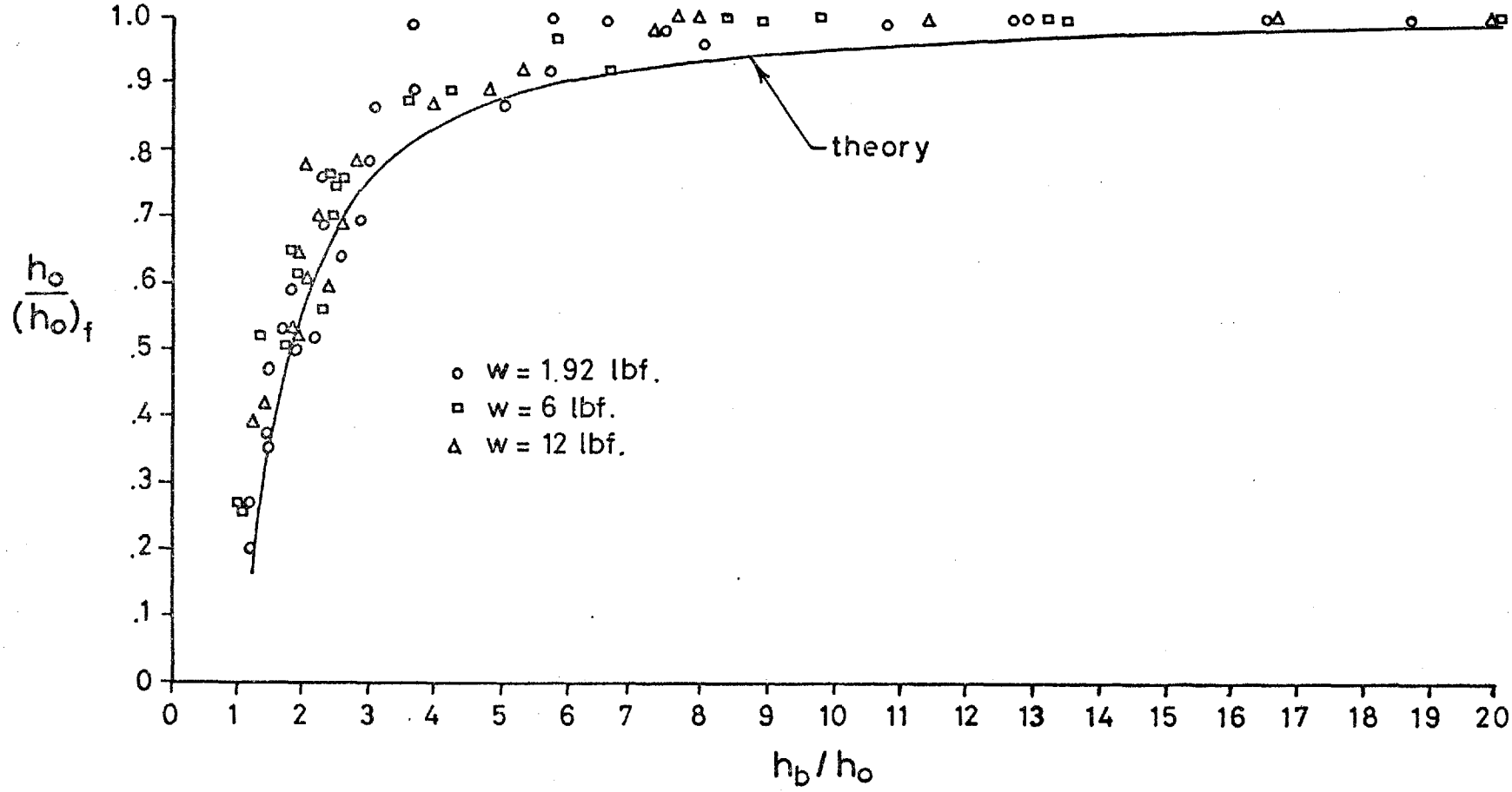


Fig.62 Comparison of theory and experiment for all the test conditions

Two of these aspects, friction and stress distribution, will be considered briefly below.

a. Friction

Friction arises from the viscous shear forces acting on the bearing surfaces. The total viscous frictional force for an elemental width dy is given by (see ref.30)

$$F = \pm \frac{dy}{2} \int_{x_1}^{x_2} h \cdot \frac{\partial p}{\partial x} dx + (u_2 - u_1)\, dy \int_{x_1}^{x_2} \frac{\eta}{h} dx$$

where x_1 and x_2 define the beginning and end of the pressure zone. The first integral identifies the rolling component of friction, while the second integral identifies the sliding component of friction.

For ehd contacts Crook(57) has shown that rolling friction is associated with the convergent inlet region where $\partial h/\partial x \neq 0$, while sliding friction is associated with the Hertzian region where the viscosities are very high and where $h \simeq h_o$. Using the inlet region to define rolling friction and the Hertzian region to define sliding friction, the friction components can be written as

$$F_R = \pm \frac{dy}{2} \int_{-\infty}^{-a} h \cdot \frac{\partial p}{\partial x} dx \qquad \ldots(6.18)$$

$$F_S = \frac{(u_2 - u_1)}{h_o} dy \int_{-a}^{+a} \eta dx \qquad \ldots(6.19)$$

Equation 6.18 shows that rolling friction is associated with the pressure generation in the inlet region. This is very much like the view brought out in section 6.5 for film thickness. The ehd pressure generated is primarily a function of α, η_o, u and R. It is expected, therefore, that rolling friction will

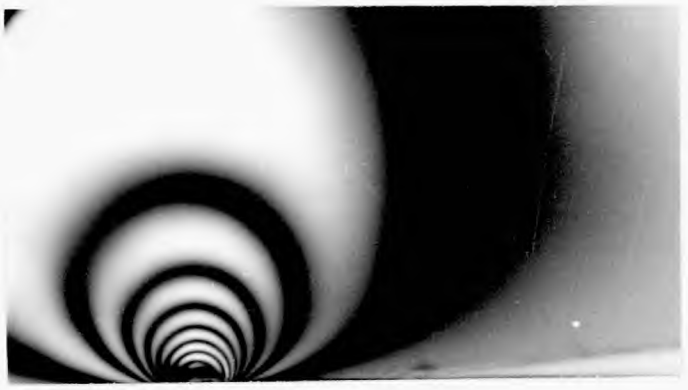
vary with these parameters in a manner similar to film thickness. Indeed, Crook(57) has experimentally shown this to be true for $u\eta_o$. A starved inlet region inhibits the ehd pressure generation and should therefore reduce the rolling friction as well as film thickness. In the limit, i.e. when the inlet distance is zero, only the Hertzian pressure prevails. Thus, rolling friction must be zero since $\partial p/\partial x = 0$ over the region of integration in equation 6.18.

Equation 6.19 shows that sliding friction is inversely proportional to h_o. Therefore, sliding friction will increase as h_o is reduced by starvation. Both the reduction in rolling friction and the increase in sliding friction with starvation have practical implications. In rolling element bearings, for example, it is desireable to have low rolling friction and high sliding friction or traction. The latter is important for preventing slip and hence bearing failure between the rolling elements and the raceways.

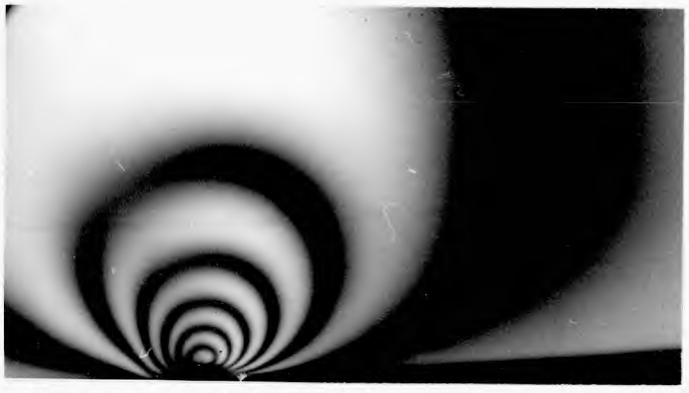
b. Stress Distribution

It was shown in section 6.4 that as starvation progresses, the ehd film shape approaches the Hertzian shape. It should follow,therefore, that the pressure distribution and the stress caused by that pressure distribution will also approach the Hertzian condition.

The effect of starvation on the photoelastic stress pattern between a glass disc and a steel disc has recently been recorded by Klemz(126). Fig.63a shows the stress pattern for static conditions. The maximum Hertzian pressure is 21,300 psi. and it is noted that there are eight symmetrical fringes radiating from the Hertzian stress field. Fig.63b shows the stress pattern under rolling conditions with a copious

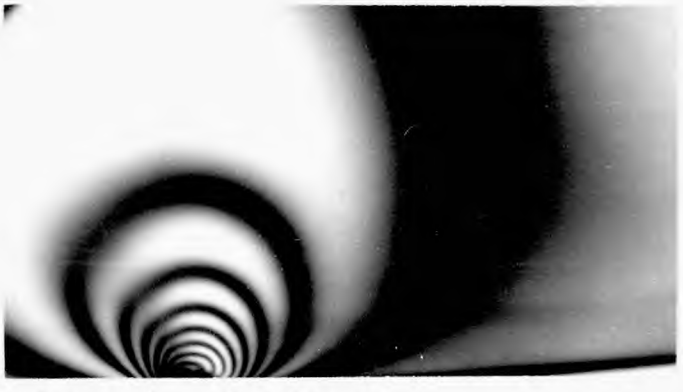


(a) static, 8 fringes



(b) dynamic, 6 fringes

(c) starved, 7 fringes



(d) starved, 8 fringes

Fig.63 Effect of starvation on the photoelastic stress pattern; taken from Klemz(126)

supply of lubricant. The ehd film has redistributed the pressure and hence the stress distribution, which now shows only six fringes. When the oil supply was cut off, the stress pattern gradually returned to the Hertzian condition of eight fringes as shown in figs.63c and d. Fig.63d was taken approximately 25 minutes after the oil supply was cut off.

It is possible to explain the gradual return of the stress distribution to Hertzian in terms of starvation. After the oil supply is cut off, the build-up of lubricant in the inlet region is gradually depleted due to flow in the axial direction. This starves the inlet region, thus inhibiting the ehd pressure generation which causes the overall pressure distribution to become more Hertzian.

There is an important difference between line and point contact in connection with starvation. In the point contact experiment, much of the lubricant which would apparently be lost from the inlet region - due to axial flow - is recaptured by surface tension. The excess lubricant on the sides of the track act as reservoirs from which lubricant is drawn into the inlet region by surface tension. For moderate speeds this mechanism is often sufficient to maintain a flooded condition. In a line contact situation, however, this mechanism will be far less effective because of the relatively long contact length in the axial direction. Cutting off the oil supply is, therefore, more likely to cause starvation since it is the only source which can successfully resupply the inlet region. It would be interesting to study the effectiveness of artificial reservoirs - such as a series of shallow grooves cut along the periphery of a roller or disc surface.

6.14 Conclusions on Starvation

Starvation is associated with the location of the lubricant boundary in the inlet region. Under starvation conditions the local film thickness in the Hertzian region is primarily a function of the boundary location immediately upstream. The results show that the onset of starvation is reflected in the expected location of pressure commencement, and that film thickness diminishes to zero as the lubricant boundary approaches the Hertzian region. This implies that film thickness is established by the ehd pressure generated in the inlet region, and that starvation inhibits the generation of this pressure. The inlet distance S, which is the distance between the lubricant boundary and the Hertzian region, reflects the physical length over which the ehd pressure is allowed to generate. It is a physical measurement which can be used to determine the degree of starvation.

Under constant operating conditions the hydrodynamic pressure generation is a function of the inlet gap geometry. The ratio h/h_o describes the relative thickness of this gap and reflects the relative magnitude of the hydrodynamic pressure generation along the inlet region. The experimental results correlate with the inlet geometry ratio h_b/h_o where h_b is the gap thickness at the lubricant boundary. The onset of starvation is not very well defined, but can be approximated by $h_b/h_o = 9$. In terms of the inlet distance it can be approximated by

$$S_f = \frac{3.52\,(Rh_o)^{\frac{2}{3}}}{a^{\frac{1}{3}}}$$

A similar expression is derived for line contact.

The results show that film thickness diminishes from its flooded condition according to h_b/h_o or S/S_f. The latter defines a starvation parameter which can be incorporated in a film thickness formula for flooded conditions to predict film thickness under starvation conditions.

The primary effect of starvation is that it inhibits the generation of hydrodynamic pressure. The effect of this was measured in terms of film thickness and was found to be in harmony with ehd theory based on Grubin assumptions. Film thickness, however, is only one aspect of starvation. The suppression of hydrodynamic pressure by starvation causes the overall pressure and stress distribution to conform to the Hertzian condition. It also reduces rolling friction and increases traction. Because these aspects have practical implications, starvation should not be considered as just a situation to avoid, but should be dealt with as an additional ehd parameter which can be usefully employed.

Chapter 7 ADDITIONAL EXPERIMENTAL WORK

7.1 Introduction

The material in this chapter represents experimental work which was initiated but not thoroughly investigated or analyzed. The results have interesting implications; but they do not always lend themselves to very definite conclusions. Some of the material presented here may, therefore, indicate fruitful areas for future work.

7.2 Results from Stroboscope Arrangement

In the stroboscope arrangement which was described in section 2.2 the balls have curvilinear as well as rotational motion. These motions are shown in fig.64a. They induce two dynamic loadings on the balls in addition to the static normal load. One dynamic loading is the centrifugal force F_c which can be calculated from the relation

$$-F_c = -mR_c\omega_c^2$$

where m = mass of the ball
R_c = radius of pitch circle
ω_c = angular velocity of cage

This force is proportional to the square of the cage speed, and its direction is shown in fig.64a.

The other dynamic loading is due to a gyroscopic moment M_g resulting from the rate of change of angular momentum of the balls. The angular momentum vector ($I\omega_b$) is shown in fig.64b. If the balls are not allowed to slip, the rate of precession $d\psi/dt$ is equal to the cage speed ω_c in

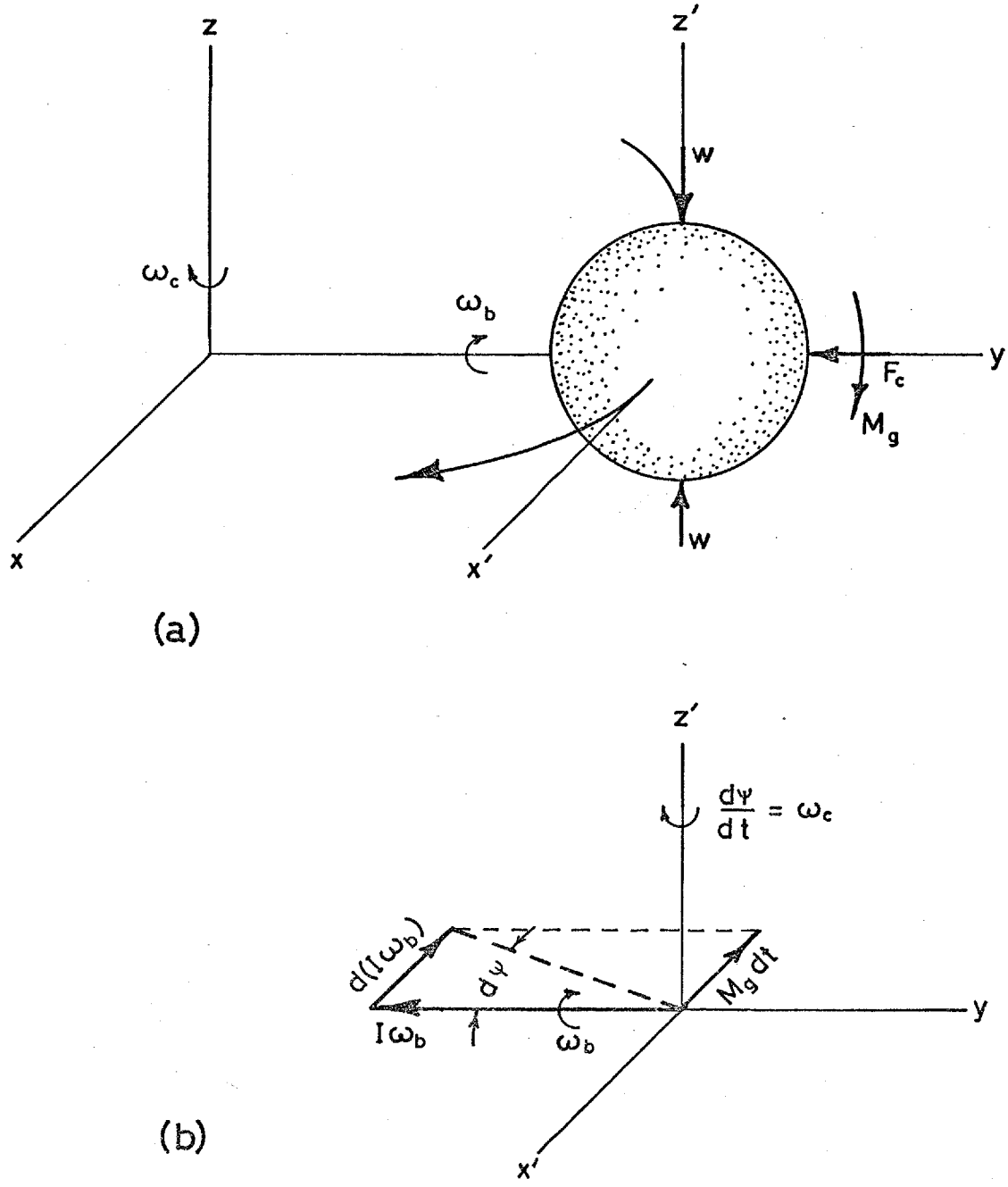


Fig.64 Diagramatic representation of stroboscope arrangement showing: (a) ball motion and loading, and (b) vector diagram for gyroscopic moment

magnitude and direction. Dynamic requirements state that the change in angular momentum $d(I\omega_b)$ during the time dt must be equal to the angular impulse $M_g dt$ These are shown vectorially in fig.64b. Thus,

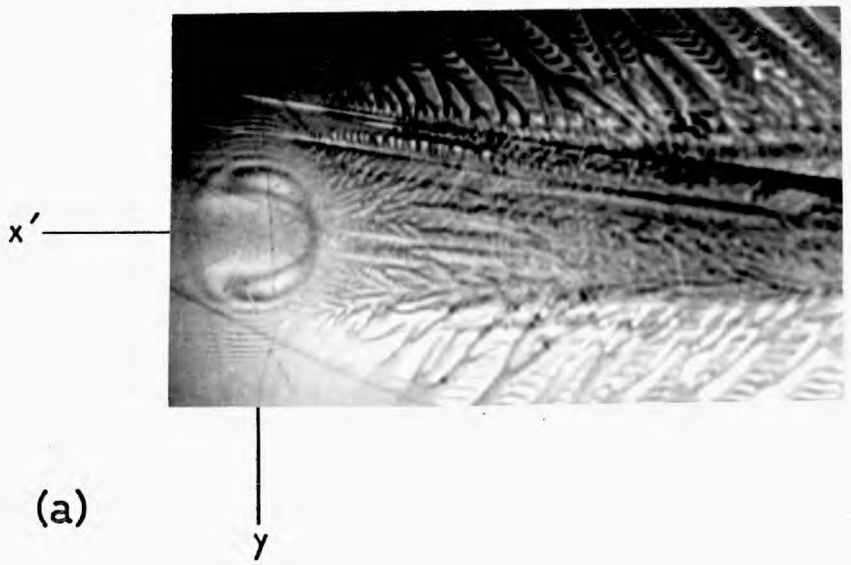
$$M_g dt = d(I\omega_b) = I\omega_b (d\psi)$$

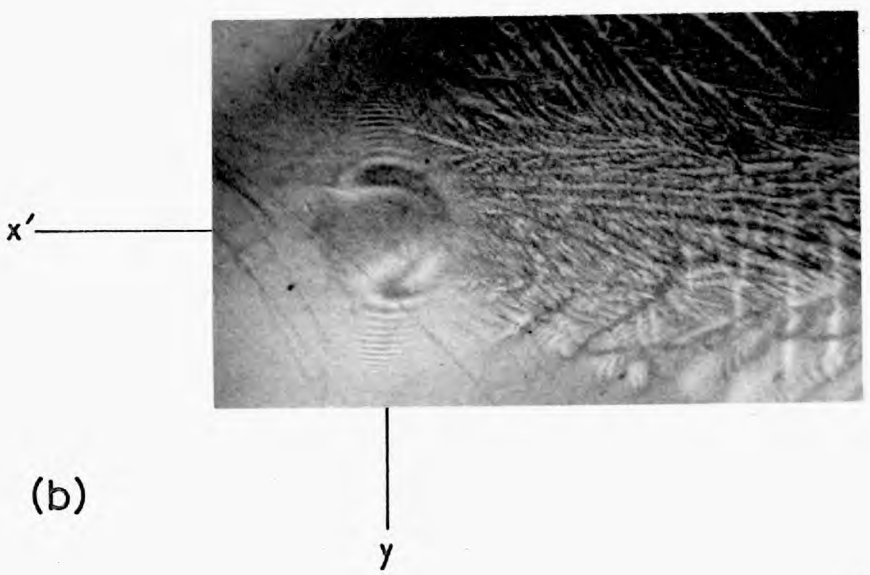
and

$$M_g = I\omega_b \omega_c$$

where I is the mass moment of inertia of the ball, and ω_b and ω_c are the ball and cage rotational speeds respectively. The direction of the gyroscopic moment is shown in fig.64a. To prevent sliding in the transverse direction this moment must be resisted by the frictional forces at the cage and bearing raceways.

Fig.65a shows a high-speed photomicrograph of the contact region under a normal load of 5 lbf. and a cage speed of 263 rpm (ω_c = 27.6 rad./sec., ω_b = 104 rad./sec.). The ehd film shape and wake pattern are similar to that seen when the bearing is operating in a counterrotation manner. The quality of the photomicrograph is lacking because at high object speeds the flash duration is not short enough to prevent image blur. This causes the closely spaced fringes outside the contact region to be visible only on the sides where they are parallel to the orbital motion of the ball. Under the given cage speed the centrifugal force is 0.53 lbf. Since the bearing races are flat, this force is resisted by the cage causing a cage loading w_c equal to 0.53 lbf. If no slip occurs, the cage loading is located on the rotational axis of the ball which produces a theoretical film of zero thickness. This





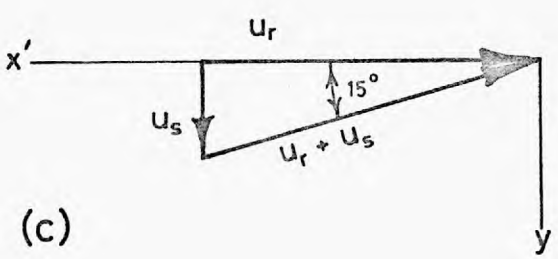


Fig.65 Illustration of ball spin: (a) no spin at cage speed of 263 rpm, (b) ball spins at cage speed of 300 rpm and (c) vector diagram of spin and roll velocities

condition has caused the generation of scratches on the ball surfaces which can be seen in figs. 65a and b.

If no gyroscopic slip is to occur, then the gyroscopic moment must be successfully resisted by the frictional forces at the cage and bearing races. Thus,

$$M_g = \mu R(2w+w_c)$$

where μ is the effective coefficient of friction required to prevent gyroscopic slip. For the given operating conditions $\mu = 0.02$ from the above equation.

The conditions shown in fig.65b were obtained under a slightly higher cage speed of 300 rpm. The orbital direction of the ball can be obtained from the visible fringes on the sides of the contact region. Both the ehd shape and the wake pattern are skewed on the order of 15° away from the orbital direction of the ball. This suggests that the ball is spinning as well as rolling, and that the spin velocity (u_s) can be determined from the vector diagram in fig.65c. The roll velocity (u_r) is 59in./sec.; therefore, the spin velocity must be 16 in./sec., giving a spin/roll ratio of 0.27.

In addition to the skew ehd shape, fig.65b shows that the film thickness is thinner at the lateral constriction, which is farthest removed from the bearing centre. This may be due to differential sliding velocities around the contact region.

It is unlikely that the ehd shape shown in fig.65b is the result of a starved inlet region. According to the results of Chapter 6, the inlet lubricant boundary, which does not appear to be in the field of view, must be approximately 1.7a ahead of the Hertzian region to establish a fully flooded

condition for the given central film thickness, which is on the order of 20 micro-inches. This requirement appears to be met. Furthermore, if the results are to be explained in terms of starvation, the inlet region must be more severely starved along the side corresponding to the lateral constriction of least film thickness. This condition of starvation can give the impression of a skewed ehd shape; but the apparent skew is in the opposite direction (see e.g. fig.6 of ref.134 and fig.15.3 of ref.135).

7.3 Edge Effects in Line Contact

Preliminary work was carried out on a tapered roller thrust bearing shown schematically in fig.66. The bottom race which was supported by the air bearing was driven while the top race remained stationary. The interference fringes were observed stroboscopically.

The tapered rollers are 0.57 inches long and have an average diameter of 0.35 inches. Fig.67a shows a roller loaded statically to a maximum Hertzian pressure of approximately 27,000 psi. The deformation of the glass race beyond the roller end causes an additional compressive stress at the end of the roller. This edge loading produces the 'dog bone' shape at the roller end.

Fig.67b shows the film shape under dynamic conditions. A very prominant lateral constriction can be seen at the contact edge and a less severe constriction at the rear. This has previously been shown by Gohar and Cameron(18). It should also be noted that the edge loading shown in fig.67a is also apparent in fig.67b. This edge loading should not significantly

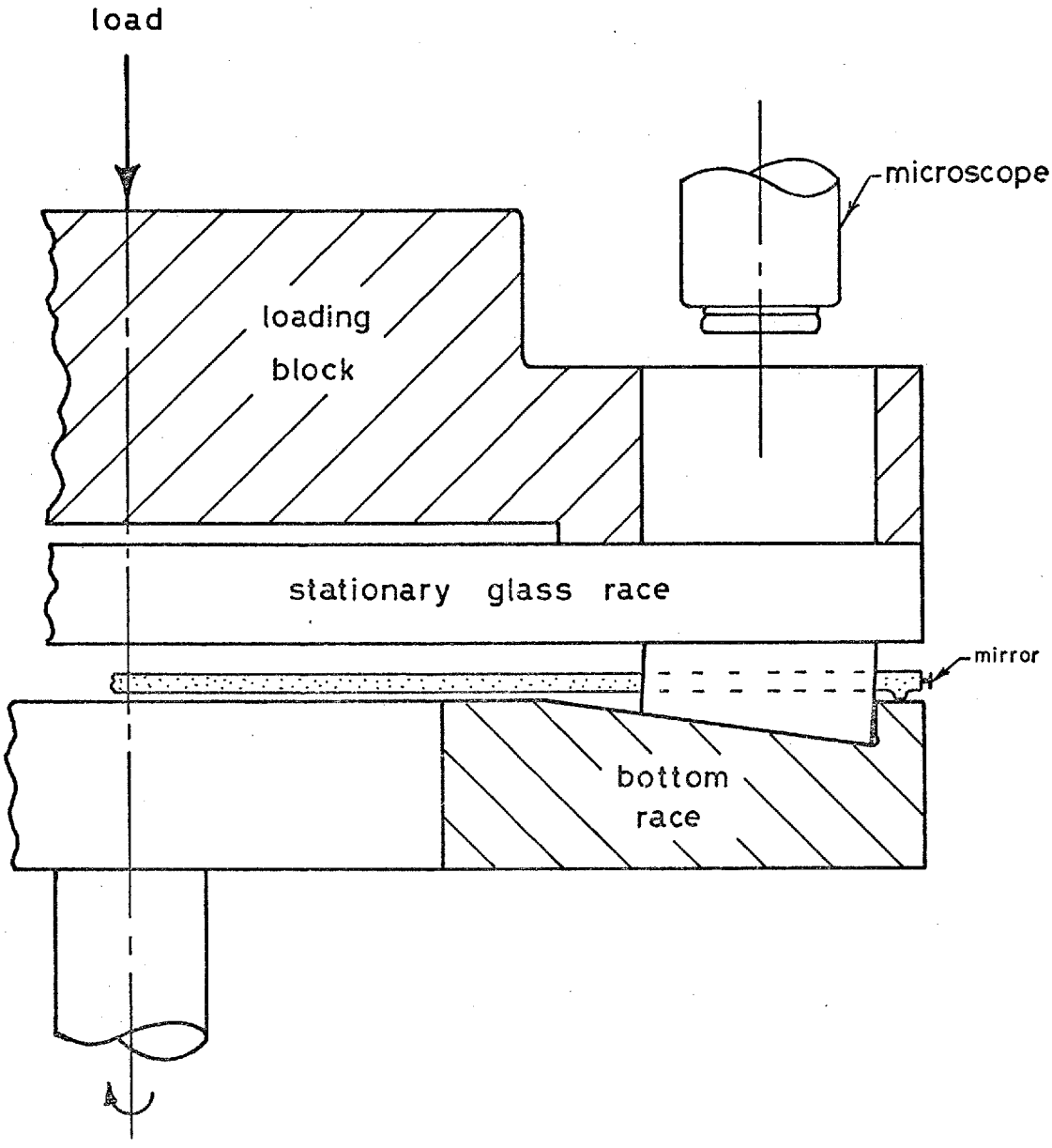
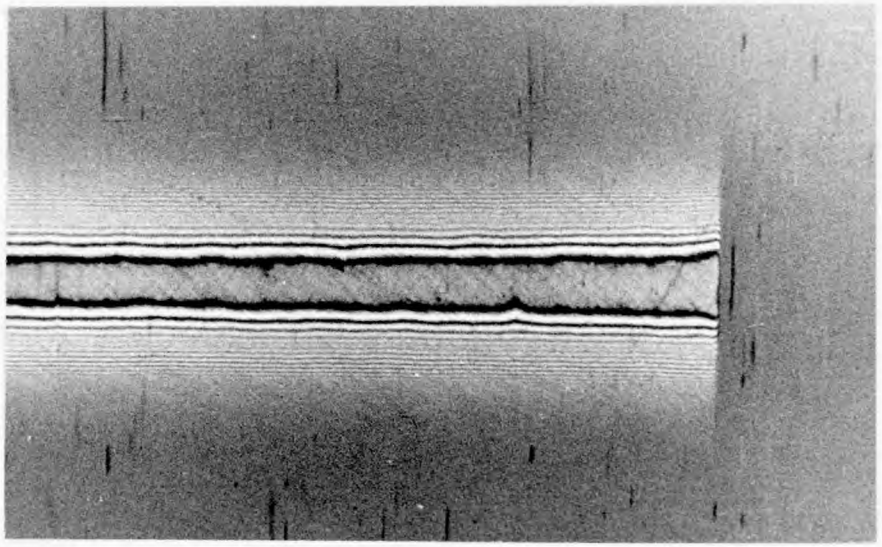
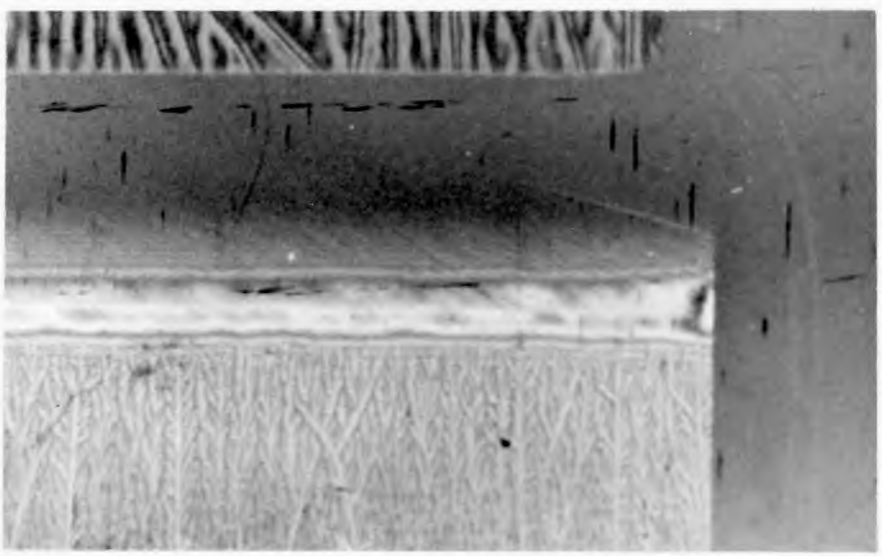


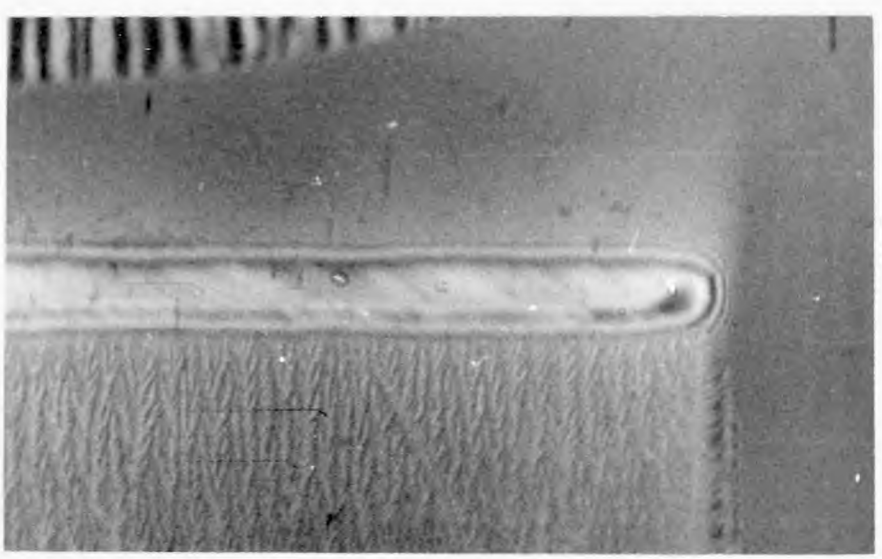
Fig.66 Schematic drawing of tapered roller thrust bearing



(a)



(b)



(c)

Fig.67 Edge effects in line contact: (a) static; (b) dynamic, unblended edge; and (c) dynamic, blended edge

effect the film thickness since film thickness has been shown to be relatively insensitive to load. The severe reduction in film thickness at the edge is apparently due to side leakage effects.

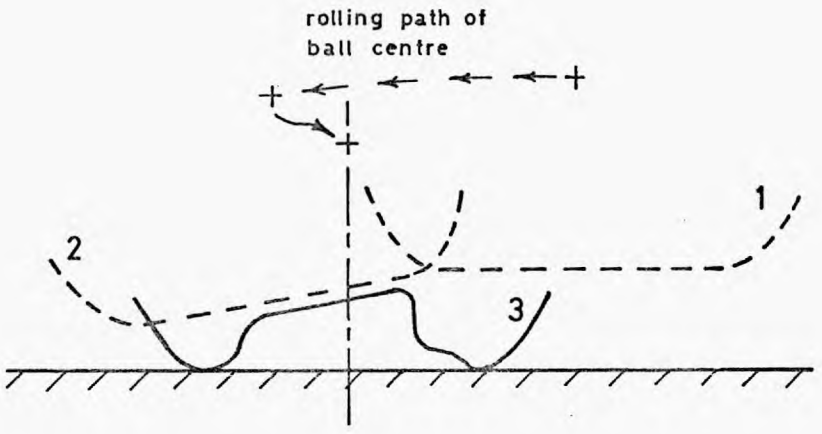
It is well known that failure invariably initiates at the edges, see e.g.Kelley and Lemanski(104). This failure which is apparently associated with the thinner films and higher stresses at the roller ends, can be reduced or eliminated by crowning the roller profile or blending the roller ends. Fig.67c shows a blended roller end. The central film thickness is the same as that in fig.67b. The blended end, which inhibits side leakage and also reduces the edge loading, has eliminated most of the lateral constriction.

Some interesting implications arise from the film shapes shown in figs.67b and c. The film shape at the edge of the blended roller is very similar to that found in point contacts. Even the oil film cavitation pattern along this edge is similar to the point contact case. In this respect point contacts can be considered as 'naturally' blended or crowned. While point contacts may not have the optimum crown to eliminate the lateral constriction, it is possible that an elliptical contact found, for example, between a ball and grooved raceway may provide the required geometry for its elimination. These results further imply that the lateral constriction for an unblended roller as found in fig.67b may be more severe than those associated with point contacts.

7.4 Entrapment from 'Rolling-Normal Approach' Conditions

An oil entrapment between the ball and glass race is shown in fig.68a. The surfaces are in contact (or very

(a)



(b)

Fig.68 Entrapment under 'rolling-normal approach' conditions: (a) photomicrograph of entrapment, and (b) schematic representation of its formation

nearly so) only along the edge of the Hertzian region. The fringe in this area is grey due to the phase change on reflection. The dark and light areas within the Hertzian region correspond to the first two fringes of the monochromatic green system. The photomicrograph was taken approximately five minutes after the entrapment was formed. During that time some of the lubricant leaked out along the scratches in the bearing surfaces.

The particular shape of the entrapment and its formation is somewhat different from the normal approach entrapments described by other workers mentioned in section 4.9a. The bearing was supplied with lubricant (LUB D) and rotated very slowly under a load of 12 lbf. (p_{max} = 87,000 psi.). The generated ehd film caused the first light fringe to appear across the Hertzian region. The corresponding cross-sectional film thickness profile is shown schematically as curve 1 in fig.68b. The bearing was then stopped by applying a breaking torque. Curve 2 of fig.68b shows schematically the film thickness at the instant when all forward motion had ceased. The shape of the film is apparently the result of the decreasing surface speed which generates a thinner film in the inlet region as the motion stops. The breaking torque then produced a slight amount of reverse motion, thus causing the surfaces to come to rest at the position represented by curve 3. The thickness of the film in the centre of the Hertzian region, where the pressure and viscosity are very high, remained essentially unchanged during the reverse motion. However, the lubricant rapidly leaked away near the edges of the Hertzian region where the viscosity is lower and the pressure gradient is higher.

The formation of an entrapment under the 'rolling-normal approach' conditions described above explains why some voltage discharge measurements taken in the laboratory several years ago on a four-ball machine did not reach zero immediately after the machine had stopped(136). The above results also imply that in a bearing application where the surfaces change direction it is possible to maintain a full ehd film even though the instantaneous surface velocity may be zero.

7.5 Greases

Because of its complicated rheological behaviour, grease has received very little fundamental attention in connection with ehd lubrication. Since it is extensively used as a lubricant in rolling element bearings it is of interest to consider briefly the following experimental results.

Fig.69 is a plot of film thickness vs speed for three lithium hydroxystearate soap greases and their base oils. The relevant details of the greases and their base oils are given in table 5. Each grease gives a higher film

TABLE 5

Properties of lithium hydroxystearate soap greases*

GREASE DESIGNATION	SHELL REF.	SOAP CONTENT	BASE OIL VISCOSITY, cs 100°F	210°F	REF. INDEX
G1	687	10.5%	33.23	4.74	1.4863
G2	688	10.5%	133.5	10.07	1.4981
G3	692	6%	133.5	10.07	1.4981

*Lubricant samples and data (except ref. index) supplied by Thornton Research Centre.

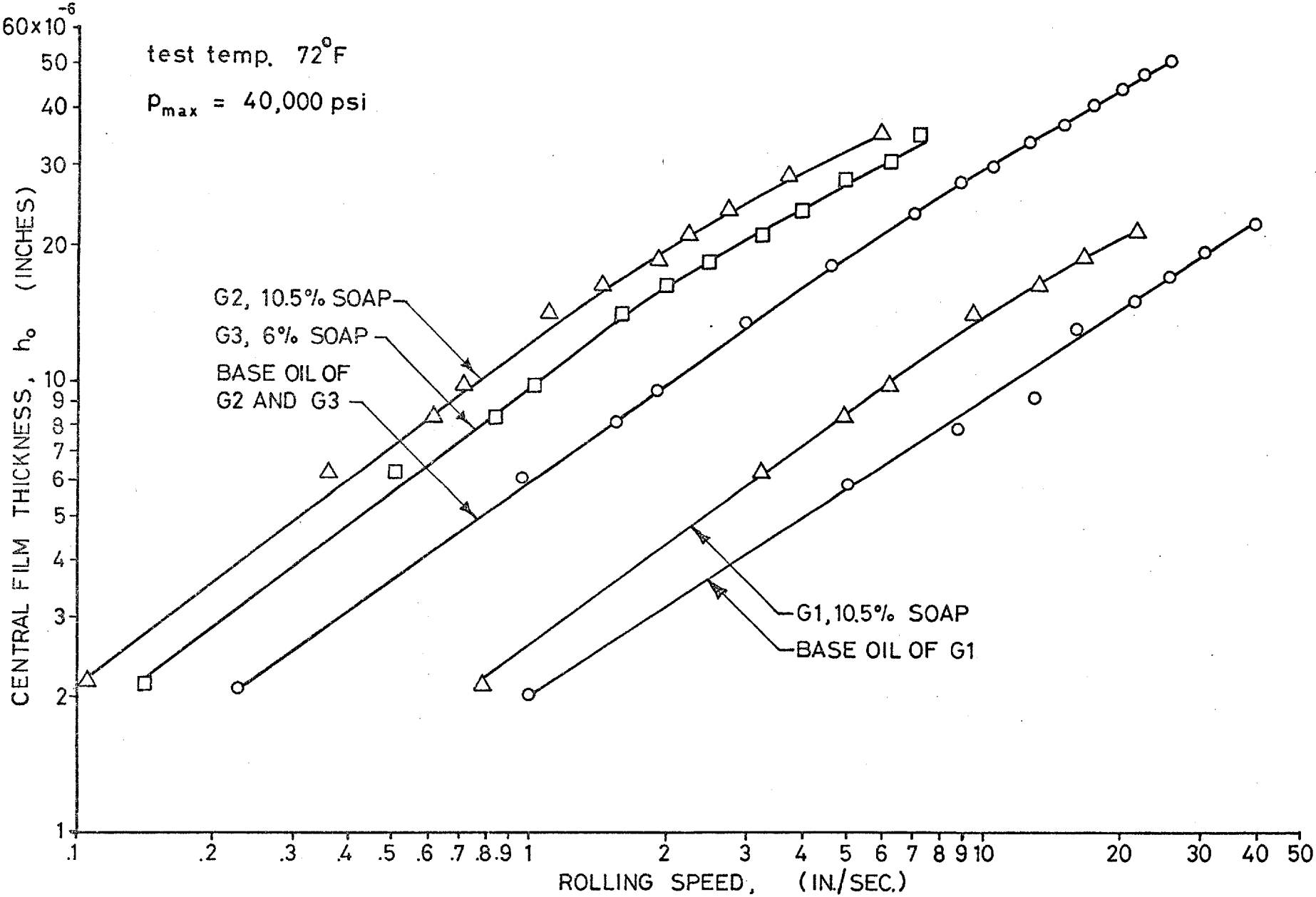
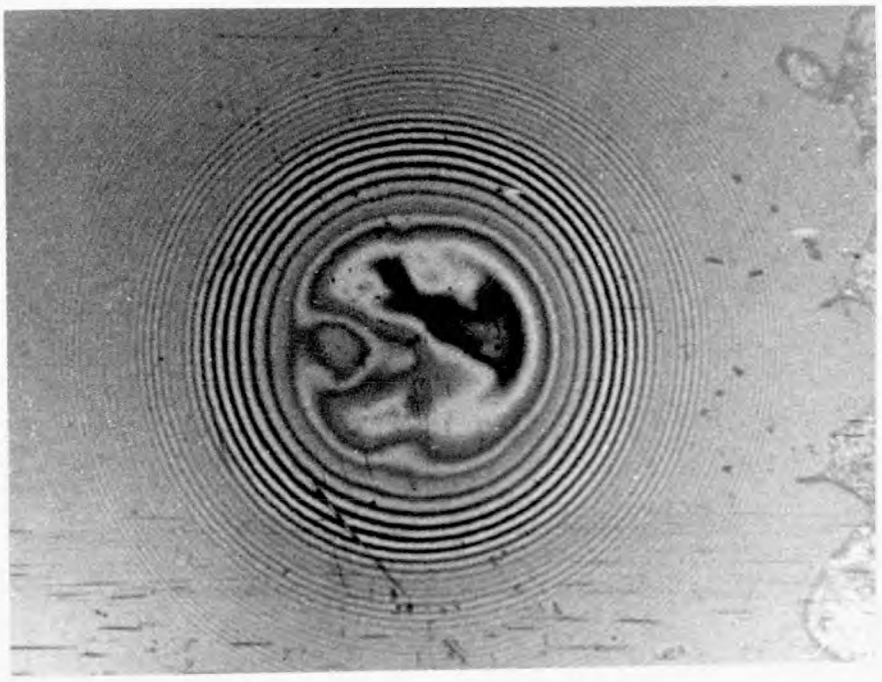


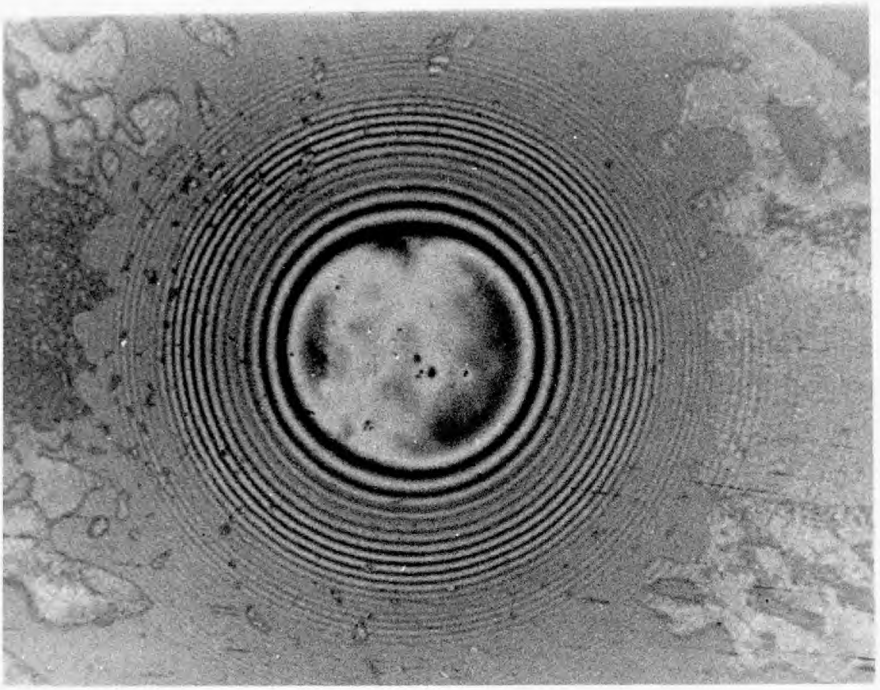
Fig.69 Film thickness vs speed for three greases and their base oils

thickness than its base oil, thus indicating that the 'apparent' viscosity of the grease is greater than the base oil viscosity, even at very high rates of shear (approximately 10^6 reciprocal seconds). It is also noted that the grease G2 with a 10.5% soap content gives a greater film thickness than the grease G3 with a 6% soap content. It has been shown by Milne et al and Scarlett(137) that high rates of shear cause the soap structure of the grease on the balls and tracks of rolling element bearings to break down. Recent work at Thornton Research Centre(138) on the same greases used in fig.69 has indicated that the soap breaks down into hard spheroidal particles. These particles have a thickening effect on the base oil which prevents the apparent viscosity of greases G1 and G2 from reaching less than 22% of its base oil viscosity and the apparent viscosity of G3 from reaching less than 16% of its base oil viscosity. This is in line with the above results.

Fig.70 shows two photomicrographs of the Hertzian region under static conditions where the maximum Hertzian pressure is about 40,000 psi. Fig.70a was obtained after the ball had been rolled very slowly in fresh grease (G1). The fringes within the Hertzian region are very irregular in shape, thus indicating a non-uniform film shape or possibly large changes in refractive index due to the heterogeneous structure of the grease. Whatever the cause, it can still be concluded that the surfaces are completely separated by a film of grease. An interesting feature of fig.70a is that the fringes do not indicate an entrapment of the form discussed in the previous section where the point of closest approach is along the edges of the Hertzian region. The minimum



(a)



(b)

Fig.70 Grease films under static conditions obtained:
(a) after rolling in fresh grease (G1), and
(b) after rolling several times over the same track

film thickness, however, is located near the centre of the Hertzian region in the area bordered by the black fringe. This area corresponds to the first red fringe of the duo-chromatic system, thus indicating a minimum film thickness on the order of 8 micro-inches. These grease films do not last forever but may take several hours before any significant change can be observed.

Fig.70b shows a grease film under static conditions after the ball had been rolled several times over the same track. The resulting film is thinner (note the position of the black fringe in each photomicrograph) and more uniform, thus suggesting a break down in soap structure. It was only possible to obtain a film of the thickness shown in fig.70a by resupplying the bearing with unworked grease. A feature of fig.70b is the appearance of small circular spots. It is possible that these spots are broken down soap particles.

Further observations on grease G1 showed that, while relatively thick and irregular films would form under very slow speeds, an increase in speed often resulted in thinner, but more uniform, films. This may be the result of a break down or an orientation of the soap structure due to slightly higher shear rates. At more moderate speeds photomicrographs show that the ehd grease and oil film shapes are exactly the same, provided thay are compared at the same nominal film thickness. A test, run at constant speed, showed no change in film thickness over a period of 30 minutes. It is likely, therefore, that most of the soap structure had broken down soon after the bearing was started. The test was terminated after 45 minutes when the film thickness became clearly effected

by starvation. This was always found difficult to avoid.

Fig.71 shows the results of two starvation tests performed with grease G2 and its base oil (TN631). There is no apparent difference in the way the film thickness diminishes with the inlet boundary from its initially flooded value. However, important differences between grease and oil lubrication in connection with starvation are likely to be found in the mechanisms by which the lubricant is transported and redistributed in the vicinity of the contact region.

7.6 Cavitation

a. Introduction

In the inlet region the normal approach of the surfaces and the viscous shear of the lubricant generate a pressure subjecting the lubricant to high compressive stresses. In the outlet region the surfaces diverge, causing a decreasing pressure which eventually subjects the lubricant to tensile stresses. Since ordinary lubricants display little or no tensile strength, the lubricant film ruptures.

Ordinary lubricants contain between 7 and 11% (by volume) of dissolved gas or air at atmospheric pressure. The equilibrium amount of dissolved gas in a lubricant is directly proportional to the absolute pressure. In a bearing, where the lubricant may be agitated by moving parts, a certain amount of entrained air may also be present. This entrained air is dispersed in the lubricant in the form of bubbles. Also, gas in the form of microscopic or sub-microscopic bubbles may be present at solid surfaces or attached to small particles within the lubricant. When the lubricant pressure is reduced below the saturation pressure, the dissolved gas diffuses into

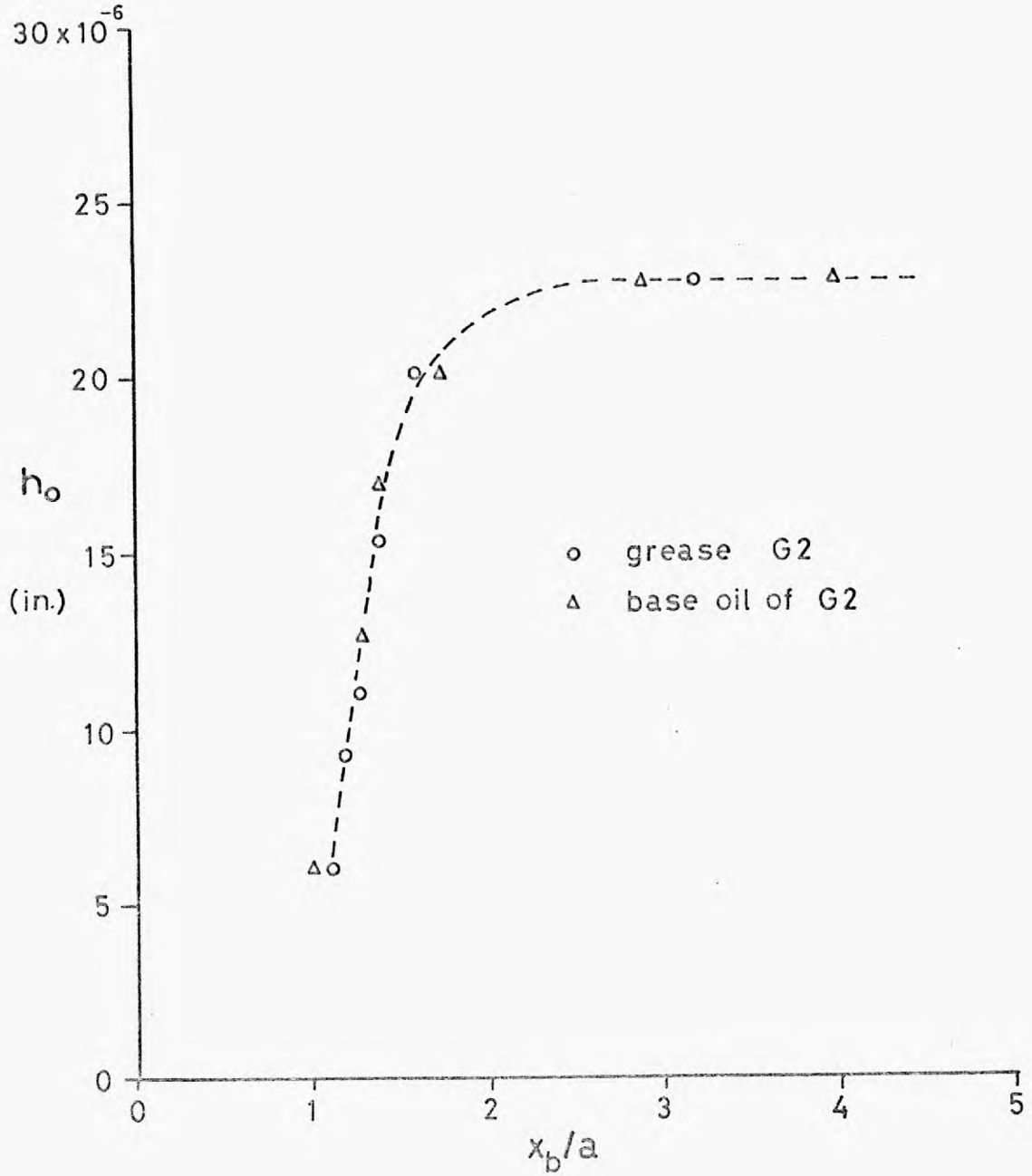


Fig.71 Starvation test for grease G2 and its base oil

the small gas bubbles causing the bubbles to grow in size. This form of bubble growth, described by Schweitzer and Szebehely (139) and Floberg(140), is the usual form of cavitation which occurs in lubricating films. Most observations of the cavitated region have been confined to line contact conditions. Notable exceptions, however, are the observations of Dowson(141), Archard and Kirk(13) and Gohar(1) on sliding point contact, and Hingley(142) on a rolling point contact in a very viscous fluid.

b. Observations of the cavitated region under rolling point contact conditions

The cavitated region, under slow speeds, may often be contained within the lubricant meniscus surrounding the contact region. At very slow speeds it may consist of a single gas bubble like that observed by Archard and Kirk(13); or, at higher speeds it may develop into a large number of gas bubbles. If the bearing is stopped the bubbles often coalesce into a single bubble which detaches itself from the contact region and slowly goes back into solution as it moves downstream. This is shown in fig.72. Fig.72a, which was taken about one second after the bearing had stopped, shows a bubble just prior to its separation from the contact region. Fig.72b shows a bubble approximately two seconds after the bearing stopped. The most striking feature is the interference pattern which has developed, apparently between the glass race and the surface of the gas bubble. It indicates that the bubble and glass race are separated by an 'ehd film' which has the features of the classical ehd point contact shape. Both the lateral and rear

(a)

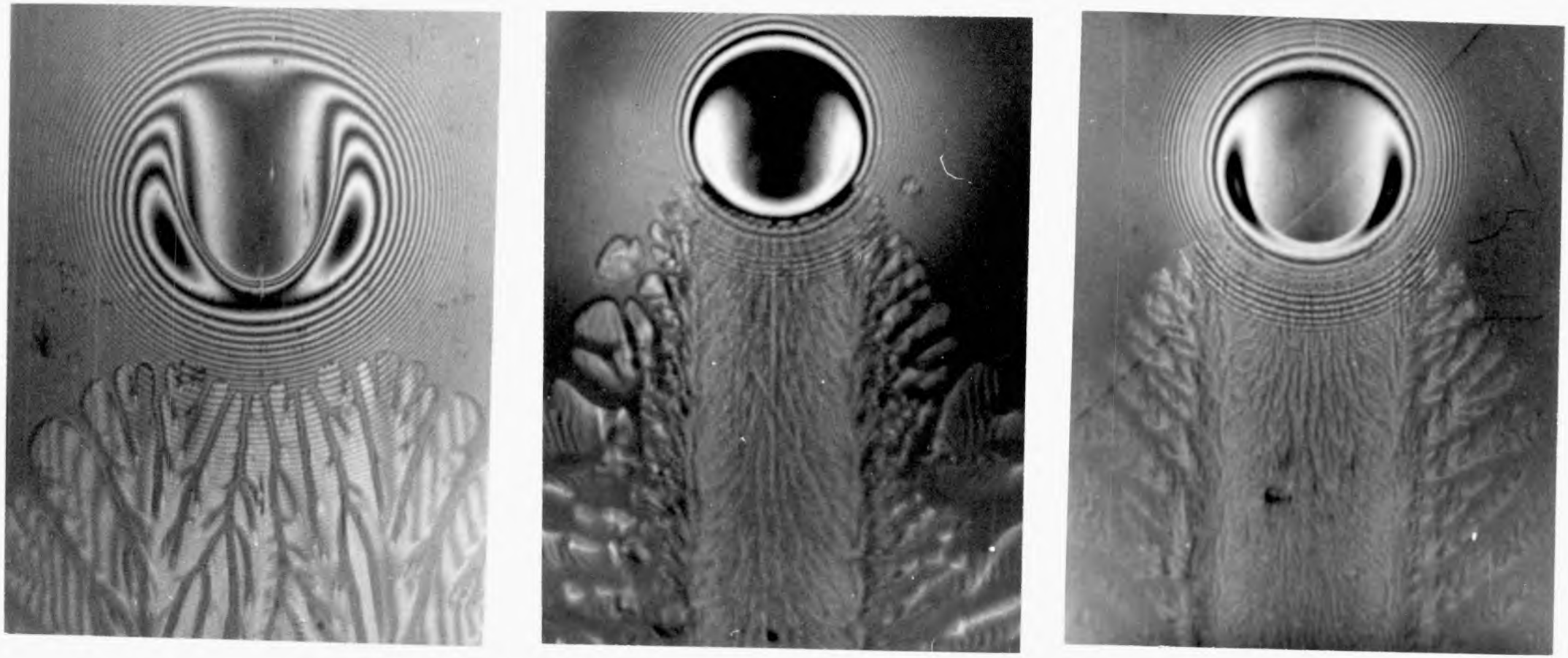
(b)

Fig.72 The coalescence and separation of the cavitation bubbles:
(a) about one second after the bearing was stopped,
(b) about two seconds after the bearing was stopped showing the 'ehd lubrication' of an air bubble

constrictions are visible as well as the slight increase in film thickness just prior to the rear constriction. Apparently, the applied load is due to the buoyant force on the bubble, and the motive force is due to the pressure gradient created by the lubricant leaking out of the entrapment formed between the glass race and steel ball.

Under normal bearing speeds the cavitated region breaks through the lubricant meniscus on the downstream side, thus exposing itself to the atmosphere. This has little effect on the general appearance of the cavitation pattern. Fig. 31 shows the usual cavitation pattern obtained under rolling point contact conditions. The cavitation pattern associated with the lubricant film which has passed through the Hertzian region appears to be distinctively different from the cavitation of the lubricant which has escaped around the Hertzian region. Although some of their features may be different, they should not be classified as two separate phenomenon such as 'separation' and 'cavitation' as has been done in the past(143). The lubricant which has passed through the centre of the Hertzian region ruptures at a film thickness which is usually about 1.5 times the central film thickness. The cavitation pattern consists of thin ribs of lubricant separated by air spaces. This is most clearly seen under large film thicknesses like that shown in fig.73a.

The lubricant which has escaped around the Hertzian region ruptures into cavitation fingers which grow perpendicular to a line which makes an angle of approximately 35^{o} with the direction of motion. The shape of the cavitation



(a) (b) (c)

Fig. 73 Cavitation patterns in rolling point contact: (a) for large film thickness (h_o = 67 micro-inches) showing striations of air and oil, (b) for high surface tension fluid (glycerol), and (c) for a concentrated dispersion of collidal graphite (Oildag)

fingers is partially determined by surface tension. A high surface tension fluid causes the cavitation fingers to become more rounded. This is shown in fig.73b where the lubricant is glycerol. It has a surface tension of 63 dynes/cm which is twice that of most lubricating oils. The small bubble to the right of the Hertzian region has cavitated prematurely. This may be due to a foreign particle within the lubricant which has provided a preferential nucleating site for cavitation to take place. The lubricant in fig.73c is a concentrated dispersion of very fine colloidal graphite (Oildag) in SAE 30 oil. These graphite particles provide additional nucleating sites thus causing the branches to form on the cavitation fingers. Surface defects such as scratches also provide preferential nucleating sites. This is shown in fig.42a and fig.15.3 of (135).

c. Importance of the cavitated region

Since cavitation terminates the hydrodynamic pressure curve, its importance has usually been associated with the relevant boundary conditions required for the solution of Reynolds equation. Recent investigations on this subject are reported by Floberg(140, 144) and Scott(145). In most bearing applications a cavitated outlet region is immediately followed by an inlet region. The influence of the cavitated region on the lubricant supply condition in the inlet region is an important practical aspect which is often overlooked. It is of interest, therefore, to consider the distribution of lubricant within the cavitated region.

The relative shift of the interference fringes due

to the different refractive indices of air and oil indicate the distribution of lubricant within the cavitated region. The relative shift of the fringes in the track of the cavitated region shown in fig.31 indicate that the air spaces, as well as the ribs of lubricant, essentially fill the gap between the bearing surfaces – at least for part of the way into the diverging outlet region. The shape of the cavitation fingers adjacent to the track can be derived from the two sets of fringes which appear within the cavitation fingers located near the Hertzian region. One set of fringes arise from the reflections of the ball and glass race. The shift of these fringes relative to their position in oil indicate that the cavitation fingers extend across most of the gap between the bearing surfaces. It is also noted that these fringes bend slightly toward the Hertzian region as they approach the tips of the cavitation fingers, thus indicating that the fingers are getting thinner. The other set of fringes run parallel to the track. These are first order fringes which increase in fringe order as they extend away from the track. These fringes remain visible even after the ball has passed. Thus, they must form between the glass race and the top surface of the cavitation fingers; and, therefore, they indicate the thickness of the lubricant film separating the cavitation fingers and the glass race.

Fig.74a shows schematically the distribution of lubricant within the cavitated region at a constant radius from the ball centre (i.e. where the gap thickness is constant). After the ball has passed, the glass race is left with the situation shown schematically in fig.74b. Narrow ribs of lubricant remain on the track while thin waves of lubricant

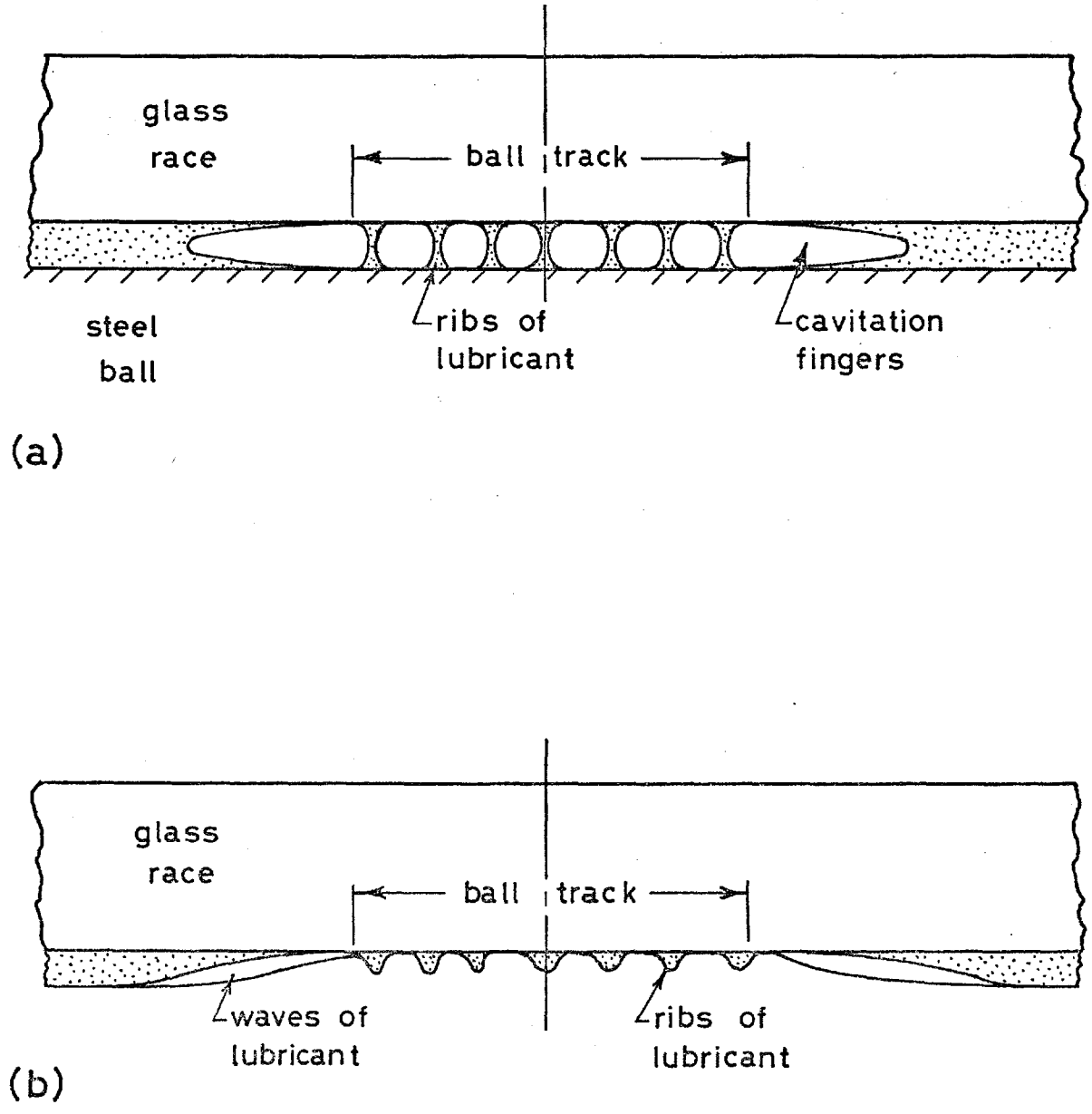


Fig.74 Schematic representation of the cavitated region: (a) at a constant radius from the ball centre and near the Hertzian region, and (b) showing the distribution of lubricant on the glass race after the ball has passed

form adjacent to the track. The same general lubricant distribution, though somewhat flattened out, appears in the inlet region as shown by the photomicrographs in Chapter 6. Unless the track is filled in, it is this depleted supply of lubricant which must provide a sufficient quantity of lubricant in the inlet region to maintain a flooded condition. It would appear, therefore, that the mechanisms associated with the transport and distribution of lubricant within the vicinity of the inlet region are very important. If the design of a bearing and the method of lubricant supply do not fulfill the lubricant requirements in the inlet region, the bearing may be operating under starved conditions even though it is 'well lubricated'.

7.7 Suggestions for Future Work

a. Investigate experimentally and theoretically the starvation problem under ehd line contact conditions. The theoretical work should begin with a solution using Grubin assumptions and end with the full solution for ehd line contact. It would be very interesting to compare the line contact results with the point contact results presented in Chapter 6.

b. Investigate the fluid velocity distribution, backflow, and circulation within the inlet region for line and point contact under rolling and sliding conditions, considering also its influence on some of the assumptions in Reynolds equation.

c. Investigate the influence of cavitation in the outlet region on the lubricant supply in the inlet region, and also the fluid flow mechanisms associated with the transport and distribution of lubricant within the inlet region.

d. Study the effect of starvation on other aspects of ehd such as stress distribution, and particularly rolling and sliding friction.

e. Attempt the complete solution for ehd point and elliptical contacts including starvation conditions.

f. Study the effect of ball spin on film thickness and shape.

g. Investigate edge effects in line and elliptical contact considering both stress distribution and the elimination of the lateral constrictions. Also, investigate the effect of side leakage in connection with the lateral constrictions.

h. Investigate the behaviour of greases under ehd conditions.

APPENDIX

POINT CONTACT SOLUTION OF EVANS INCLUDING STARVATION CONDITIONS

Some of the details of the approximate point contact solution using Grubin assumptions are given below. The formulation and computation of the problem were kindly carried out by Evans(79). He was assisted by the writer, especially in connection with the method for extracting the required film thickness information from the dimensionless parameters of the results.

The dimensionless Reynolds equation, including the side leakage term, in polar coordinates is

$$\frac{1}{r^*} \cdot \frac{\partial}{\partial r^*}\left(r^* h^{*3} \frac{\partial q^*}{\partial r^*}\right) + \frac{h^{*3}}{r^{*2}} \cdot \frac{\partial^2 q^*}{\partial \phi^2} = \cos\phi \cdot \frac{\partial h^*}{\partial r^*}$$

where $r^* = \frac{r}{a}$

$h^* = \frac{h}{h_o}$

$q^* = \frac{h_o^2 \, q}{12 u \eta_o a}$ ...(A1)

The solution of Reynolds equation concerns the region ABCD shown in fig.75

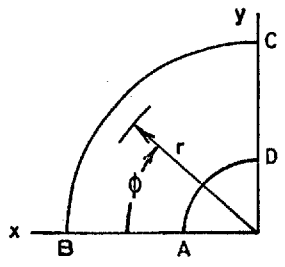


Fig.75 Representation of inlet region in polar coordinates

The film shape in this region is assumed to be Hertzian and is given by

$$h^* = 1+B\left[(r^{*2}-1)^{\frac{1}{2}} - (2-r^{*2})\cos^{-1}\frac{1}{r^*}\right]$$

where

$$B = \frac{3w}{2\pi a E' h_o} \qquad \ldots(A2)$$

To set up in finite difference form the transformation $s = \log_e r^*$ was used. The boundary conditions are:

along AB Due to symmetry along AB, $\partial q^*/\partial\phi = 0$

BC $q^* = 0$ at some specified value of r^*. This is the inlet boundary condition which is located at various values of r^* to obtain starvation conditions.

CD q^* was assumed to fall off according to $1/h^n$, where $n = 10$ was found to be compatible with the boundary condition along DA. The solution was not very sensitive to n, particularly for high loads or thin films.

DA The Grubin solution shows $q = 1/\alpha$ at the Hertzian radius. Therefore, $q^* = \text{const.}$ and $\partial q^*/\partial r^* = 0$ along DA. Only the zero derivative, however, was held along DA. The solution proceeded until the correct value of q^* at point A was found to give the zero derivative constraint along DA.

The problem was computed for the range of B and r^* values used in the twelve experimental test runs. For flooded conditions (i.e. large values of r^*) the solution gave

$$q^* = 0.093\ B^{-.6}$$

The results for starved conditions, i.e. for q* = 0 located at decreasing values of r*, are shown in fig.76 where q* is plotted against r* for constant values of B.

To facilitate the comparison of the theoretical and experimental results we need h_o as a function of r* for the experimental operating conditions used in each starvation test. Both q* and B are functions of h_o as shown by equations A1 and A2. These equations can be written as

$$h_o = (12\alpha\eta_o u a q^*)^{\frac{1}{2}} \qquad \ldots(A3)$$

$$h_o = \frac{3}{2\pi} \cdot \frac{w}{aE'B} \qquad \ldots(A4)$$

Equating the two and solving for q*, gives

$$q^* = \frac{3}{16\pi^2 B^2} \cdot \frac{w^2}{\alpha\eta_o u (E')^2 a^3}$$

Substituting equation 6.8 for a^3 gives q* as a function of B^2 for constant speed-viscosity and load parameters. Thus we get

$$q^* = \frac{1}{8\pi^2 B^2} \left(\frac{\alpha\eta_o u}{R}\right)^{-1} \cdot \frac{w}{E'R^2} \qquad \ldots(A5)$$

For each starvation test the speed-viscosity and load parameters are constant; and, under starvation conditions B must always be larger than under flooded conditions (see equation A4). From equation A5 q* is calculated for each B larger than the flooded condition. The required h_o is given by equation A3 and the required value of r*, corresponding to the given B and calculated q*, is found by interpolation between the data points shown in fig.76. This procedure was iterated for all the B values larger than the flooded condition. The computed results for each starvation test are given in figs. 59, 60 and 61 of Chapter 6.

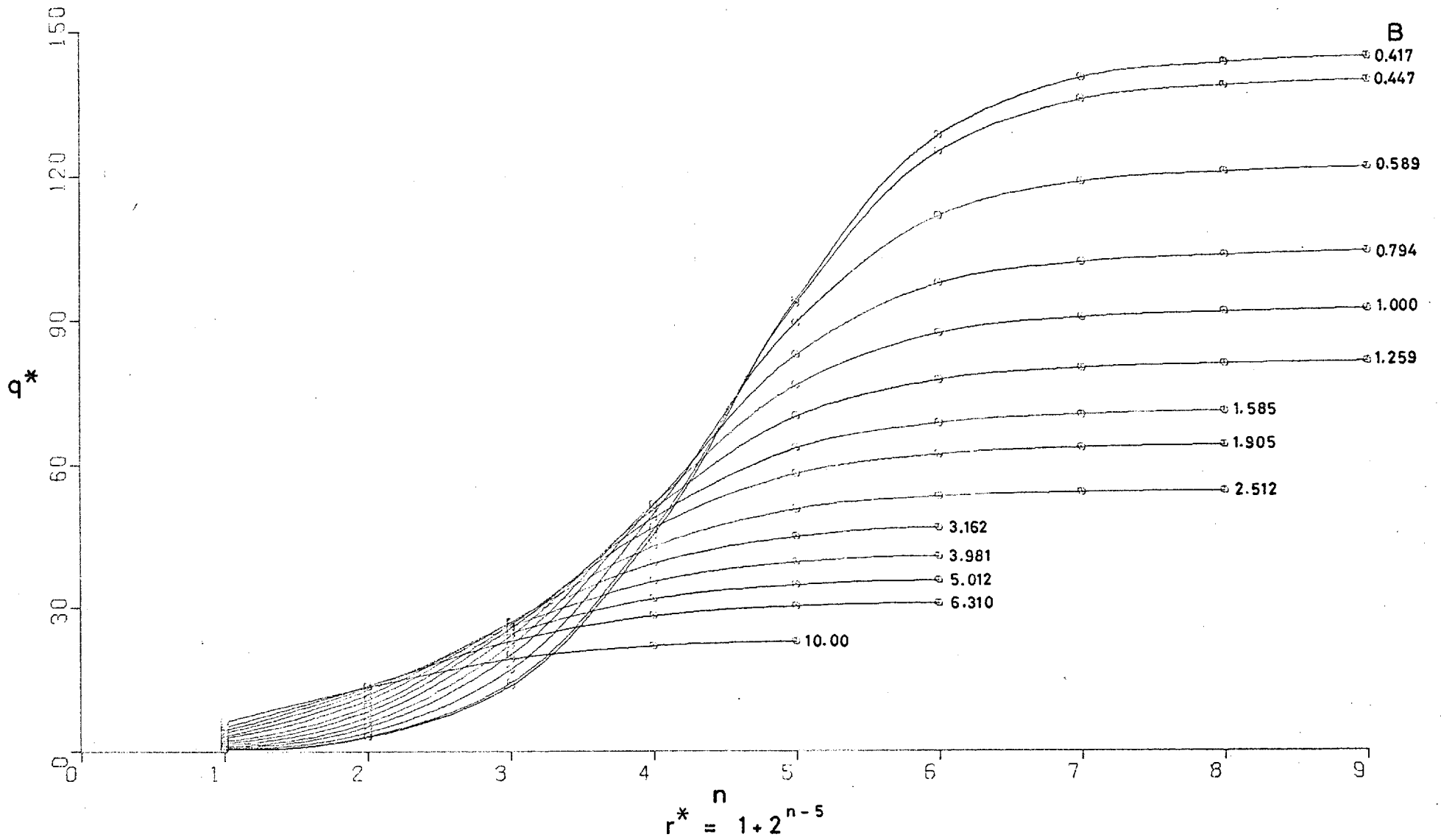


Fig.76 Point contact solution for starvation conditions
(only 40% of the total number of curves are shown)

REFERENCES

(1) Gohar, R., 'Lubrication in Point Contact', Ph.D. Thesis, University of London (Imperial College), 1965.

(2) Foord, C.A., 'Pitting and Film Thickness in Rolling Contact', Ph.D. Thesis, University of London (Imperial College), 1967.

(3) Grassam, N.S. and Powell, J.W., Gas Lubricated Bearings, Butterworths, London, 1964.

(4) Loch, E., 'Aerostatic (Externally-Pressurized Gas Lubricated) Thrust Bearings', Translation report T-A2049-26; Translated from a doctoral thesis submitted to the Technical University of Graz, Austria, The Franklin Institute Research Laboratories, 1965.

(5) Constantinescu, V.N., Inst. of Fluid Mechanics, Bucharest, Rumania, private communication, April 1967.

(6) Tolansky, S., An Introduction to Interferometry, Longmans, London, 1966.

(7) Tolansky, S., Multiple Beam Interferometry of Surfaces and Films, Oxford University Press, London, 1948.

(8) Tolansky, S., Surface Microtopography, Longmans, London, 1960.

(9) Francon, M., Optical Interferometry, Academic Press, New York, 1966.

(10) Steel, W.H., Interferometry, Cambridge University Press, London, 1967.

(11) Hardy, W.B., Collected Scientific Papers of Sir William Bate Hardy, Cambridge University Press, London, 1936, pp. 609-23.

(12) Kirk, M.T., 'Hydrodynamic Lubrication of Perspex', Nature, 1962, 194, p. 965.

(13) Archard, J.F., and Kirk, M.T., 'Influence of Elastic Modulus on the Lubrication of Point Contacts', Inst. Mech. Engr., 1963, Lubrication and Wear Convention, paper 15, pp. 181-89.

(14) Blok, H. and Koens, Ir H.J., 'The "Breathing" Film Between a Flexible Seal and a Reciprocating Rod', Proc. Inst. Mech. Engrs. 1965-66, 180, Pt3B, pp. 221-223.

(15) Cameron, A. and Gohar, R., 'Theoretical and Experimental Studies of the Oil Film in Lubricated Point Contacts', Proc. Roy. Soc., (A), 1966, 291, pp. 520-536.

(16) Foord, C.A., Hammann, W.C. and Cameron, A., 'Evaluation of Lubricants Using Optical Elastohydrodynamics', ASLE Trans. 11, 31-43, 1968.

(17) Michelson, A.A., Light Waves and Their Uses, The Decennial Publications, University of Chicago Press, Chicago, Second Series, Vol. III, 1907.

(18) Gohar, R., and Cameron, A., 'The Mapping of Elastohydrodynamic Contacts', ASLE paper No. 66LC-21, Lubrication Conference, Oct. 1966.

(19) Westlake, F.J. and Cameron, A., 'Fluid Film Interferometry in Lubrication Studies', Nature 1967, 214, 633.

(20) Westlake, F.J. and Cameron, A., 'A Study of Ultra-Thin Lubricant Films Using an Optical Technique', Proc. Inst. Mech. Engrs. 1967-68, 182, Pt 3 G, Report 12, pp. 75-78.

(21) Foord, C., Wedeven, L., Westlake, F.J. and Cameron, A., 'Optical Elastohydrodynamics' , to be published in Proc. Inst. Mech. Engrs., 1969-70, 184, Part 1.

(22) Westlake, F.J. and Cameron, A., 'High Speed Photographic Study of Lubricated Contacts Using Optical Interferometry', Journal of Photographic Science, 1969, 17, No.4, pp.137-140.

(23) Frisk, Optical Works Ltd., Ealing, London, private communication, 28, August 1969.

(24) Hass, G. and Hadly, L., American Institute of Physics Handbook, McGraw-Hill, 1963, pp.6-103.

(25) Westlake, F.J., 'A Study of Ultra-Thin Films', Ph.D. Thesis, Imperial College, London, to be submitted to the University of London.

(26) Barakat, N. and Mokhtor, S., 'Experimental Determination of Phase Change of Reflection from Glass-Metallic Layer Interface Using a Koster Interference Comparator', Journal of the Optical Soc. of Amer., Oct. 1963, 53, No. 10, pp.1153-55.

(27) Tolansky, S. private communication, 28 Jan. 1969.

(28) Holden, J., 'Multiple-Beam Interferometry: Intensity Distribution in the Reflected System', Proc. Physical Soc., July 1949, Section B, 62, Part 7, No.355B, pp. 405-17.

(29) Hopkins, H.H., University of Reading, formerly of Imperial College.

(30) Cameron, A., Principles of Lubrication, Longmans, London, 1966.

(31) Pinkus, O. and Sternlicht, B., Theory of Hydrodynamic Lubrication, McGraw-Hill Book Company, 1961.

(32) (Martin), 'The Lubrication of Gear Teeth', Engineering, Aug.11, 1916, 102, pp.119-21.

(33) Dowson, D. and Higginson, G.R., Elasto-Hydrodynamic Lubrication, The Fundamentals of Roller and Gear Lubrication, Pergamon Press Ltd., Oxford, 1966.

(34) Blok, H., Communications, Proc. Inst. Mech. Engrs., 1965-66, 180, Pt 3B, p.237.

(35) Grubin, A.N. and Vinogradova, I.E., 'Investigation of Scientific and Industrial Research', Central Scientific Research Institute for Technology and Mechanical Engineering, 1949, Book No. 30, pp.115-66 (D.S.I.R. Translation No.337).

(36) Chu, P.S.Y. and Cameron, A., 'Pressure Viscosity Characteristics of Lubricating Oils', J. Inst. Petrol., 1962, 48, 147-55.

(37) 'Viscosity and Density of Over Forty Lubricants', Report of ASME Research Committee on Lubrication, ASME New York, 1953.

(38) Dowson, D. and Higginson, G.R., 'A Numerical Solution to the Elasto-Hydrodynamic Problem', J. Mech. Engr. Sci., 1959, 1, No.1, pp.6-15.

(39) Dowson, D. and Higginson, G.R., 'The Effect of Materials Properties on the Lubrication of Elastic Rollers', J. Mech. Engr. Sci., 1960, 2, No.3, pp.188-194.

(40) Archard, G.D., Gair, F.C., and Hirst, W., 'The Elasto-hydrodynamic Lubrication of Rollers', Proc. Roy. Soc., (A), 1961, 51, pp.51-72.

(41) Dowson, D., Higginson, G.R., and Whitaker, A.V., 'Elasto-hydrodynamic Lubrication: A Survey of Isothermal Solutions', J. Mech. Engr. Sci., 1962, 4, No.2, pp. 121-26.

(42) Dowson, D. and Whitaker, A.V., 'The Isothermal Lubrication of Cylinders', ASLE Trans., 1965, 8, pp.224-34.

(43) Archard, J.F., 'The Temperature of Rubbing Surfaces', Wear, 1959, 2, p.438.

(44) Crook, A.W., 'The Lubrication of Rollers III., A Theoretical Discussion of Friction and the Temperature in the Oil Film', Phil. Trans. of the Roy. Soc., (A), 1961, 254, pp.237-58.

(45) Crook, A.W., 'The Lubrication of Rollers, II., Film Thickness with Relation to Viscosity and Speed', Phil. Trans. of the Roy. Soc. (A), 1961, 254, pp.223-36.

(46) Cheng, H.S. and Sternlicht, 'A Numerical Solution for the Pressure, Temperature, and Film Thickness Between Two Infinitely Long, Lubricated Rolling and Sliding Cylinders Under Heavy Loads', J.Mech. Engr. Sci., 1965, 87, pp.695-707.

(47) Cheng, H.S., 'A Refined Solution to the Thermal-Elastohydrodynamic Lubrication of Rolling and Sliding Cylinders', ASLE Trans. 1965, 8, pp.397-410.

(48) Dowson and Whitaker, 'A Numerical Procedure for the Solution of the Elastohydrodynamic Problem of Rolling and Sliding Contacts Lubricated by a Newtonian Fluid', Proc. Inst. Mech. Engrs., 1965-66, 180, Part 3B, p.57.

(49) Dowson, D. and Higginson, G.R., 'New Roller-Bearing Lubrication Formula', Engineering, 1961, 192, pp.158-9.

(50) Dowson, D. and Higginson, G.R., 'The Lubrication of Elastic Rollers', Xth International Congress of Applied Mechanics, 1960.

(51) Blok, H., Communication;'Elastohydrodynamic Lubrication', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt 3B, p.237; Moes, H., Communication, Ibid., p.244.

(52) Greenwood, J.A., 'Presentation of Elastohydrodynamic Film-Thickness Results', J. Mech. Engr. Sci., 1969, 11, No.2, pp.128-32.

(53) Halling, J., Communication, Proc. Inst. Mech. Engrs., 1967-68, 182, Pt. 3A, p.259.

(54) Archard, J.F., 'Non-dimensional Parameters in Isothermal Theories of Elastohydrodynamic Lubrication', J. Mech. Engr. Sci., 1968, 10, No.2, pp.165-67.

(55) Dowson, D., 'Elastohydrodynamics', Proc. Inst. Mech. Engrs. 1967-68, 182, Part 3A, paper no. 10.

(56) Crook, A.W., 'The Lubrication of Rollers', Phil. Trans. of the Roy. Soc. (A), 1958, 250, pp.387-409.

(57) Crook, A.W., 'The Lubrication of Rollers IV., Measurements of Friction and Effective Viscosity', Phil. Trans. of the Roy. Soc., (A), 1963, 255, pp.281-312.

(58) Crook, A.W., 'Developments in Elasto-Hydrodynamic Lubrication', J. Inst. Petrol, 1963, 49, pp.295-307.

(59) Sibley, L.B. and Orcutt, F.K, 'Elasto-Hydrodynamic Lubrication of Rolling-Contact Surfaces', ASLE Trans., 1961, 4, pp.234-49.

(60) Dyson, A., Naylor, H., and Wilson, A.R., 'The Measurement of Oil-Film Thickness in Elasto-hydrodynamic Contacts', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, p.119.

(61) Archard, J.F., 'Experimental Studies of Elastohydrodynamic Lubrication', Proc. Inst. Mech. Engrs., 1965-66, 180, Part 3B, paper no. R2.

(62) Crook, A.W., 'Elastohydrodynamic Lubrication of Rollers', Nature, 1961, 190, 1182.

(63) Cheng, H.S. and Orcutt, F.K., 'A Correlation Between the Theoretical and Experimental Results on the Elastohydrodynamic Lubrication of Rolling and Sliding Contacts', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt 3B, paper no.13.

(64) Hamilton, G.M. and Moore, S.L., 'A Modified Gauge For Investigating an Elastohydrodynamic Contact', Proc. Inst. Mech. Engrs. 1967-68, 182, Pt. 3A, p.251.

(65) Higginson, G.R., 'A Model Experiment in Elasto-hydrodynamic Lubrication', Int. J. Mech. Sci., 1962, 4, p.205.

(66) Dowson, D. and Longfield, M.D., 'Distribution of Pressure and Temperature in a Highly Loaded Lubricated Contact', Inst. Mech. Engrs., Lubrication and Wear Convention, 1963, pp.27-33.

(67) Dowson, D. and Longfield, M.D., 'The Lubrication of Rolls of Finite Width: An Investigation of Oil-Film Characteristics', Inst. Mech. Engrs., Third Annual Meeting of the Lubrication and Wear Group, October 1964, Cardiff, paper no.7.

(68) Longfield, M.D., 'Pressure Distribution in a Highly Loaded Lubricated Contact', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B., paper no.9.

(69) Niemann, G. and Gartner, F., 'Distribution of Hydrodynamic Pressure on Counterformel Line Contacts', ASLE Trans. 1965, 8, No.3, pp.235-49.

(70) Kannel, J.W., Bell, J.C., and Allen, C.M., 'Methods for Determining Pressure Distributions in Lubricated Rolling Contact', ASLE Trans, 1965, 8, pp.250-70.

(71) Kannel, J.W., 'Measurements of Pressures in Rolling Contact', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.92.

(72) Fein, R.S., 'Possible Role of Compressional Viscoelasticity in Concentrated Contact Lubrication', Trans. ASME, J. Lub. Tech., April 1967, pp.127-133.

(73) Orcutt, F.K., 'Experimental Study of Elastohydrodynamic Lubrication', ASLE Trans. 1965, 8, pp.381-96.

(74) Kannel, J.W., Walowit, J.A., Bell, J.C. and Allen, C.M., 'The Determination of Stress in Rolling-Contact Elements', ASME-ASLE Lubrication Conference, Minneapolis, Minn., Oct. 1966, Paper No. 66-Lub-16.

(75) Archard, J.F., and Kirk, M.T., 'Lubrication at Point Contacts', Proc. Roy. Soc. (A), 1961, 261, pp. 532-50.

(76) Roberts, A.D. and Tabor, D., Short Communication; 'Fluid Film Lubrication of Rubber- An Interferometric Study', Wear, 1968, 11, pp.163-6.

(77) Archard, J.F. and Cowking, E.W., 'Elastohydrodynamic Lubrication at Point Contacts', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.3.

(78) Snidle, R.W. and Archard, J.F., 'Lubrication at Elliptical Contacts', Proc. Inst. Mech. Engrs., 1968-69, 183, Pt. 3P, paper no.17.

(79) Evans, D., Imperial College Computer Dept., private communication.

(80) Cheng, H.S., 'A Numerical Solution of the Elastohydrodynamic Film Thickness in an Elliptical Contact', Lubrication Symposium, San Francisco, June 1969, ASME paper No.69-LubS-17.

(81) Christenson, H., 'The Oil Film in a Closing Gap', Proc. Roy. Soc. 1962, (A), 266, 312.

(82) Christenson, H., Elastohydrodynamic Theory of Spherical Bodies in Normal Approach Motion', Report MTI-67 TR21, sponsored by Department of the Navy Office of Naval Research, Contrast NOOOl4-66-COO37 (1967).

(83) Dowson, D. and Jones, D.A., 'Lubricant Entrapment Between Approaching Elastic Solids', Nature, 1967, 214, No.5091, p.947.

(84) Dowson, D. and Jones, D.A., 'An Optical-Interference Method of Measurement of Time-Dependent Elastohydrodynamic Film Profiles', Proc. Inst. Mech. Engrs., 1967-68, 182, Pt. 3G, pp.49-52.

(85) Smith, F.W., 'The Effect of Temperature in Concentrated Contact Lubrication', ASLE Trans., 1962, 5, pp.142-148.

(86) Johnson, K.L. and Cameron, R., 'Shear Behaviour of Elastohydrodynamic Oil Films at High Rolling Contact Pressures', Proc. Inst. Mech. Engrs. 1967-68, 182, Pt. 1, paper no.2.

(87) Plint, M.A., 'Some Recent Research on the Perbury Variable-Speed Gear', Proc. Inst. Mech. Engrs. 1965-66, 180, Pt. 3B, pp.225-227.

(88) Plint, M.A., 'Traction in Elastohydrodynamic Contacts', Proc. Inst. Mech. Engrs. 1967-68, Pt. 1, paper no.3.

(89) Bell, J.C., Kannel, J.W., and Allen, C.M., 'The Rheological Behaviour of the Lubricant in the Contact Zone of a Rolling Contact System', Trans. ASME, J. Basic Engr., Sept. 1964, 86, Series D, No.3, pp.423-35.

(90) Dyson, A., 'Flow Properties of Mineral Oils in Elasto-hydrodynamic Lubrication', Phil. Trans. of the Roy. Soc. (A) 1965, 258, p.529.

(91) Barlow, A.J. and Lamb, J., 'The Visco-Elastic Behaviour of Lubricating Oils Under Cyclic Shearing Stress', Proc. Roy. Soc. (A), 1959, 253, pp.52-69.

(92) Oldroyd, J.G., 'Non-Newtonian Effects in Steady Motion of Some Idealized Elasto-Viscous Liquids', Proc. Roy. Soc.(A), 1958, 245, pp.278-97.

(93) Archard, J.F. and Kirk, M.T., 'Film Thickness for a Range of Lubricants Under Severe Stress', J. Basic Engr. Sci. 1964, 6, no.1, pp.101-2.

(94) Dyson, A. and Wilson, A.R., 'Film Thickness in Elastohydrodynamic Lubrication of Silicone Fluids', I. Mech. Engrs., 1966, Lubrication and Wear Convention, Paper 3.

(95) Hamilton, G.M. and Robertson, W.G., 'Lubrication of Rollers with Oils Containing Polymer', Proc. Inst. Mech. Engrs. 1966-67, 181, Pt. 30, paper no.3.

(96) Tallian, T.E., 'Rolling Contact Failure Control Through Lubrication', Proc. Inst. Mech. Engrs. 1967-68, 182, Pt. 3A, paper no. 14,

(97) Dawson, P.H., 'Effect of Metallic Contact on the Pitting of Lubricated Rolling Surfaces', J. Mech. Engr. Sci., 1962, 4, no.1, pp.16-21.

(98) Dawson, P.H., 'The Effect of Metallic Contact and Sliding on the Shape of the S-N Curve for Pitting Fatigue', Inst. Mech. Engrs., Symposium on Fatigue in Rolling Contact, March 1963, paper no.4, pp.41-45.

(99) Dawson, P.H., 'Further Experiments on the Effect of Metallic Contact on the Pitting of Lubricated Rolling Surfaces', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.7.

(100) Dudley, D.W., 'Elastohydrodynamic Behaviour Observed in Gear Tooth Action', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.17.

(101) Tallian, T.E., Chiu, Y.P., Huttonlocher, D.F., Kamenshine, J.A., Sibley, L.B., and Sindlinger, N.E., 'Lubricant Films in Rolling Contact of Rough Surfaces', ASLE Trans. 1964, 7, pp.109-126.

(102) Tallian, T.E., McCool, J.I. and Sibley, L.B., 'Partial Elastohydrodynamic Lubrication in Rolling Contact', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.14.

(103) Christensen, H., 'Nature of Metallic Contact in Mixed Lubrication', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.12.

(104) Kelley, B.W., and Lemanski, A.J., 'Lubrication of Involute Gearing', Proc. Inst. Mech. Engrs., 1967-68, 182, Pt. 3A, paper no.11.

(105) Garnell, P. and Higginson, G.R., 'The Mechanics of Roller Bearings', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper no.16.

(106) Dowson, D. and Higginson, G.R., 'Theory of Roller-Bearing Lubrication and Deformation', Inst. Mech. Engrs., Lubrication and Wear Convention, 1963, pp.216-227.

(107) Dowson, D. and Higginson, G.R., 'A Theory of Involute Gear Lubrications', Inst. Petrol., Gear Lubrication Symposium, 1964, pp. II 7-14.

(108) Howell, P.G., 'Experiments in the Aerosol Lubrication of Power Gearing', Proc. Inst. Mech. Engrs. 1966-67, 181, Pt.30, paper no.11.

(109) Jones, F.C. and Wilcock, D.F., 'The Mechanism of Lubrication Failure in High-Speed Ball Bearings', ASME Trans. 1950, 72, No.6, pp.817-23.

(110) Booser, E.R. and Wilcock, D.F., 'Minimum Oil Requirements of Ball Bearings', Lubrication Engr., 1953, 9, No.3, pp. 140-3.

(111) Kingsbury, E.P., 'Experimental Observations on Instrument Ball Bearings', M.I.T. Instrument Laboratory Publication E-2316, Prepared for presentation at the Dartmouth Bearings Conference, Dartmouth College, Hanover, N.H., Sept. 1968.

(112) Horsch, J.D., 'Correlation of Gyro Spin-Axis Ball Bearing Performance with the Dynamic Lubricating Film', ASLE Trans., 1963, 6, pp. 112-24.

(113) Freeman, A.P., 'Gyro Ball Bearings - Technology Today', M.I.T. Instrumentation Laboratory, Prepared for presentation at the Sixth AGARD Guidance and Control Meeting, "Inertial Navigation: Components," at Braunschweig, Germany, May 1968.

(114) Blok, H., discussion, Proc. Inst. Mech. Engrs., 1965-66, pp. 108-109.

(115) Boness, R.J., communication, Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, pp. 238-39.

(116) Boness, R.J., 'Hydrodynamic Lubrication of Lightly Loaded Finite Cylinders', Proc. Inst. Mech. Engrs., 1966-67, 181, Pt.30, paper 14.

(117) Boness, R.J., 'The Lubrication of Cylinders with Particular Reference to Roller Bearings', Ph.D. Thesis, 1969, Royal Military College of Science, Schrivenham.

(118) Dowson, D. and Whomes, T.L., 'Side-Leakage Factors for a Rigid Cylinder Lubricated by an Isoviscous Fluid', Proc. Inst. Mech. Engrs., 1966-67, 181, Pt. 30, paper 8.

(119) Fein, R.S. and Kreutz, K.L., 'Interdisciplinary Approach to Friction and Wear', NASA SP-181, Symp. held at Southwest Research Inst., San Antonio, Texas, Nov. 28-30, 1967, pp. 364-6.

(120) Orcutt, F.K. and Cheng, H.S., 'Lubrication of Rolling-Contact Instrument Bearings', Proc. of the Gyro Spin-Axis Hydrodynamic Bearing Symposium, Vol. II, The Instrumentation Laboratory, M.I.T., Cambridge, Mass., U.S.A., Dec. 1966, Tab. 5., pp. 1-25.

(121) Castle, P., Dept. of Mech. Engr., University of Leeds, private communication, 19, June 1969.

(122) Glasstone, S., Textbook of Physical Chemistry, Macmillan and Co. Ltd., London, 1948.

(123) Poulter, T.C., Ritchey, C. and Benz, C.A., 'The Effect of Pressure on the Index of Refraction of Paraffin Oil and Glycerine', Physical Review, Aug.1, 1932, 41, pp. 366-7.

(124) Chu, P.S.Y. and Cameron, A., 'Compressibility and Thermal Expansion of Oils', J. Inst. Petrol, May 1963, 49, No. 473, pp. 140-5.

(125) Galvin, G.D., Naylor, H., and Wilson, A.R., 'The Effect of Pressure and Temperature on Some Properties of Fluids of Importance in Elastohydrodynamic Lubrication', Proc. Inst. Mech. Engrs., 1963-64, 178, Pt. 3N, pp. 283-90.

(126) Klemz, B., Imperial College, private communication, also 'Stress Measured by Photoelasticity', Ph.D. Thesis, to be submitted to the University of London.

(127) Fresco, G.P., 'Measurement and Prediction of Viscosity-Pressure Characteristics of Liquids', MSc. Thesis, Penn. State University, Report No. PRL 3-66, June 1966.

(128) Gohar, R., 'A Ball-and-Plate Machine for Measuring Elastohydrodynamic Oil Films', Proc. Inst. Mech. Engrs., 1967-68, 182, Pt. 3G, pp. 43-45.

(129) Gohar, R., 'Oil Film Thickness and Rolling Friction in Elastohydrodynamic Point Contact', to be published in ASLE.

(130) Lauder, W., 'Hydrodynamic Lubrication of Proximate Cylindrical Surfaces of Large Relative Curvature', Proc. Inst. Mech. Engrs., 1965-66, 180, Pt. 3B, paper 8.

(131) Whitaker, A.V., communication, Proc. Inst. Mech. Engrs., 1965-66, 180, Pt 3B, p.246.

(132) Hartung, H.A., 'The Entrance Space in Rolling Contacts', ASME paper 65-WA/LUB-5.

(133) Tipei, N., 'Boundary Conditions of a Viscous Flow Between Surfaces With Rolling and Sliding Motion', ASME Trans., Jan. 1968, 90, series F, no.1, pp. 254-261.

(134) Wedeven, L. and Cameron, A., 'The Observation of Elastohydrodynamic Lubrication in a Rolling Element Bearing Using Optical Interferometry', 10th International Symposium on Lubrication of Rolling Bearings, 27-31 Aug. 1968, Halle, Germany, to be published in shortened form in Schmierungstechnik.

(135) Wedeven, L. and Cameron, A., 'A Study of Elastohydrodynamic Lubrication in Rolling Bearings Using Optical Interference', Proc. Inst. Mech. Engrs., 1967-68, 182, Pt 3G, pp. 87-89.

(136) Cameron, A., Imperial College, private communication.

(137) Scarlett, N.A., 'Use of Greases in Rolling Bearings', Proc. Inst. Mech. Engrs., 1967-68, 182, Pt. 3A, paper 21.

(138) Wilson, A.R., Thornton Research Centre, private communication, 9 Feb., 1970.

(139) Schweitzer, P.H. and Szebehely, 'Gas Evolution in Liquids and Cavitation', J. Applied Phy., 1950, 21, no. 12, pp. 1218-24.

(140) Floberg, L., 'Sub-Cavity Pressures and Number of Oil Stresses in Cavitation Regions with Special Reference to the Infinite Journal Bearing', Acta Polytechnica Scandinavica, Mech. Engr. Series No. 37, 1968.

(141) Dowson, D., 'Investigation of Cavitation in Lubricating Films Supporting Small Loads', Inst. Mech Engrs., Proc. of the Conference on Lubrication and Wear, 1957, paper 49.

(142) Hingley, C.G. et al, 'A Study of the Geometry of Elastohydrodynamic Films in Point Contact', SKF Prog. Report, No.1, U.S. Navy Contract NOO 019-67-C-0206.

(143) Taylor, G.I., 'Cavitation in Hydrodynamic Lubrication', Cavitation in Real Liquids, Proceedings of the Symposium on Cavitation in Real Liquids, General Motors Research, 1962, Elsevier Publishing Co., 1964, pp. 80-101.

(144) Floberg, L., 'On Hydrodynamic Lubrication with Special Reference to Sub-Cavity Pressures and Number of Streamers in Cavitation Regions', Acta Polytechnica Scandinavica, Mech. Engr. Series, No.19, 1965.

(145) Scott, P.A.J., 'Conditiile La Limita Ale Ecuatiilor Lubrificatiei Hidrodinamice', Academia Republicii Socialiste Romania, Institutul de mecanica fluidelor "Traian Vuia", Bucuresti, 1968, also to be submitted as Ph.D. thesis, University of London.

PUBLICATIONS

Report 15

A STUDY OF ELASTOHYDRODYNAMIC LUBRICATION IN ROLLING BEARINGS USING OPTICAL INTERFERENCE

By L. D. Wedeven* and A. Cameron*

INTRODUCTION

THE optical interference technique used in previous work (1)–(3) has provided a convenient method for establishing oil-film thicknesses and shapes, and for studying the fundamental characteristics of elastohydrodynamic lubrication. These experiments have dealt with single balls or rollers rotated about an axis fixed in space and loaded against suitably reflecting transparent plates.

In rolling bearing practice the rolling elements are usually contained within a cage, and travel between a stationary and rotating race. The knowledge of the lubricating film characteristics in this situation is important for solving practical bearing problems.

This report describes a method of isolating the contact area of a given rolling element. A partially reflecting glass plate is used as the top race of a ball thrust bearing (Fig. 15.1). Load is applied to this stationary top race while the bottom race, supported by an air thrust bearing, is driven by an electric motor. The cage is furnished with a small mirror which operates an optical trigger. This triggers a high-intensity Xenon flash as the rolling element passes the point of observation.

This method along with the optical interference technique provides a stroboscopic observation of the lubricating film. Also, an instantaneous picture of the lubricating film can be obtained by using high-speed flash photography.

EQUIPMENT

Optics

The partially reflecting glass plate and the reflecting surface of the rolling element present a beam splitting device in which the variation of film thickness changes the optical path difference between the two reflecting beams, thus creating fringes produced by cancellation and reinforcement of the reflected beams. The interference fringes are observed through a microscope having a long working distance X5 objective and a wide field X10 eyepiece.

Illumination of the contact area is obtained through a collimator (Fig. 15.1) attached to the side of the microscope and below the objective. A thin glass reflector (E) directs the light normal to the contact area. An iris diaphragm (A) which is focused on to the object plane acts as a field stop and an extended source. It constitutes the first element of the collimator. The next element is a collecting lens (B), followed by another iris diaphragm (C) used as an aperture stop. The final element is a collimating lens (D) positioned from the aperture a distance equal to its focal length. The aperture stop determines the degree of collimation as well as the amount of illumination.

The above lens system provides excellent control of collimation and illumination for visual stroboscopic work. However, for high-speed flash photography the requirement of a short flash duration limits the luminous intensity of the source. The amount of available light in this case is more fully used by focusing the source on to the image plane with only a single lens. The effect of the resulting loss in collimation on fringe definition is not significant, since the relatively flat Hertzian area does not require a high degree of fringe definition. Furthermore, the definition of the fringes located outside the Hertzian area is already affected by the light reflecting from the curved surface of the ball.

The illuminating source is an Xenon-filled discharge lamp which has a colour temperature of approximately 6000°K and produces a continuous spectrum similar to daylight, except for a peak in the blue range. Interference colours from white light are limited to an oil-film thickness of about 10^{-4} in beyond which white becomes the dominant colour even though interference is still taking place. By filtering out all but single or multiple spectral band widths, greater thicknesses can be measured. Figs 15.2 and 15.3, for example, were taken with a Wratten interference filter type 77A.

TRIGGERING

The Xenon flash is synchronized with the cage rotation by means of a Dawe phototransistor pick-up. The cage is furnished with a small reflecting mirror positioned so that reflected light from a lamp and lens system energizes a

The MS. of this report was received at the Institution on 6th December 1967. 2

* *Imperial College of Science and Technology, London.*

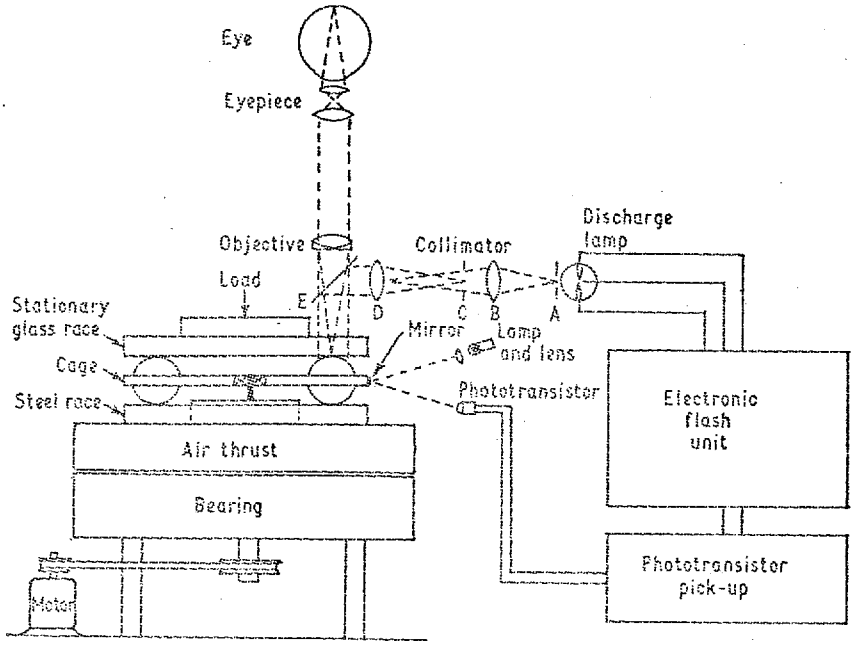


Fig. 15.1. Schematic diagram of apparatus for the stroboscopic optical interference observation of the elastohydrodynamic lubrication of a ball thrust bearing

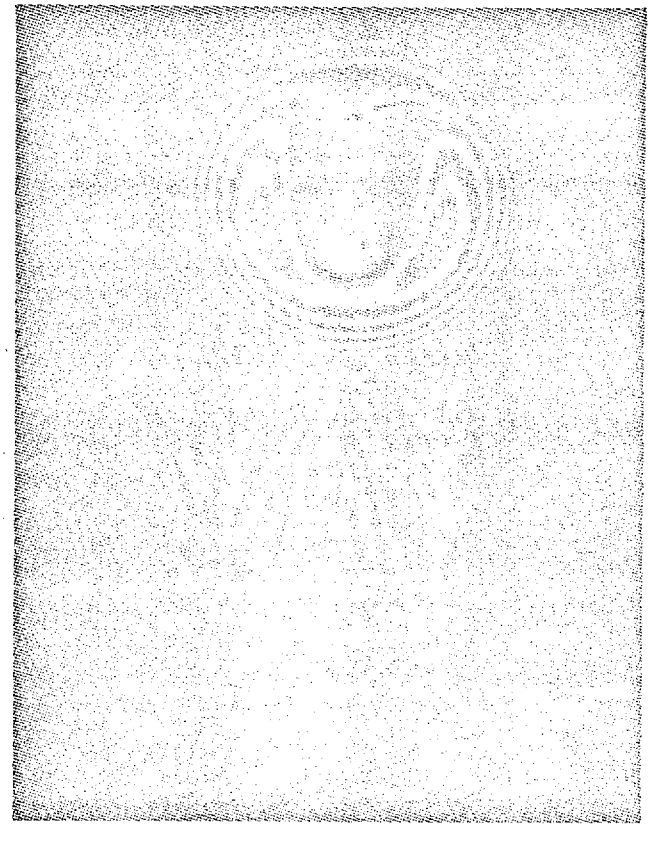


Fig. 15.2. High-speed photomicrograph showing oil-film shape and oil-film rupture

Fig. 15.3. High-speed photomicrograph of scratched ball. Unsymmetrical shape due to oil starvation

phototransistor when the rolling element passes beneath the point of observation. The output signal is applied to a two-transistor bi-stable circuit. An inductance contained within the switching circuit provides a pulse of 30–50 V.

ELECTRONIC FLASH UNIT

The Xenon discharge lamp is a Ferranti type CC10. The circuit provides a means of triggering the lamp from an external signal or an internal signal using a multivibrator as a frequency control. The internal signal operation provides a visible source for aligning and setting-up the equipment. The required voltage potential across the Xenon lamp is obtained by charging a capacitor connected in parallel with it. The lamp is triggered by a very high voltage impulse (8 kV) obtained from the secondary side of a firing transformer. The primary side is connected with a capacitor in conjunction with a high-peak current relay valve. An external signal from the phototransistor pick-up or an internal signal from the multivibrator applied to the relay valve commences the triggering operation for the Xenon flash.

HIGH-SPEED FLASH PHOTOGRAPHY

A microscope camera in conjunction with a synchronized flash from the discharge lamp provides a recording of the lubricating condition at an instant in time (Figs 15.2 and 15.3). High rolling element speeds where both a short flash duration and sufficient illumination are required are the most difficult to photograph. The procedure has been to use the maximum allowable discharge voltage (1·5 kV) with a low discharge capacitance (1μf) together with the single lens collimator providing energy just sufficient to expose a high-speed black and white film. Flash durations in the range of 5–10 μs can thus be obtained, enabling rolling element speeds in the order of 150 in/s to be photographed.

CONCLUSIONS

The experimental method presented provides a means of observing the elastohydrodynamic conditions of rolling elements in a situation where some of the practical and fundamental lubrication problems can be investigated. For example, in the ball thrust bearing an alteration of film shape due to gyroscopic spin can be observed. Furthermore, the use of high-speed flash photography facilitates the observance of the condition of lubrication at an instant in time. Thus, the influence of large surface irregularities (e.g. scratches on a ball) can be observed (Fig. 15.3). Also, in Figs 15.2 and 15.3 the oil-film rupture formation in the diverging section can clearly be seen. Under certain conditions this may indeed affect the lubrication of the following rolling element.

This stroboscopic optical interference method can be applied to many types of rolling bearings provided suitable transparent bearing races can be obtained. A flat glass plate conveniently serves as a bearing surface for a ball thrust bearing. It can also be applied to a specially designed taper roller thrust bearing that will operate on a flat surface, thus providing line contact in a pure rolling situation.

ACKNOWLEDGEMENTS

The authors wish to thank Ferranti Ltd and Mr Seabury of Imperial College for the basic design and construction of the electronic flash unit, and the Central Unit for Scientific Photography, Royal Aircraft Establishment, for their advice on electronic flash techniques.

The authors are grateful to Professor H. H. Hopkins, formerly of the Physics Department, Imperial College, for assistance in the design of a collimation system and also to Hoffman Bearing Company for their generous supply of bearings and bearing components.

REFERENCES

(1) GOHAR, R. and CAMERON, A. 'Theoretical and experimental studies of the oil film lubricated point contact', *Proc. Roy. Soc.* 1966 A, 291, 520–536.

(2) GOHAR, R. and CAMERON, A. 'The mapping of elastohydrodynamic contacts', *ASLE/ASME Conference*, Minneapolis, October 1966.

(3) FOORD, C. A., HAMMANN, W. C. and CAMERON, A. 'Evaluation of lubricants using optical elastohydrodynamics', *ASLE Conference* 1967 Chicago.

THE OBSERVATION OF ELASTOHYDRODYNAMIC LUBRICATION IN A ROLLING ELEMENT BEARING USING OPTICAL INTERFEROMETRY

by

L. Wedeven and A. Cameron

- the shortened version of a paper written for the 10th International Symposium on Lubrication of Rolling Bearings, 27-31 August, 1968, Halle, Germany; to be published in Schmierungstechnik.

ABSTRACT

One of the most important parameters associated with the performance and life of rolling bearings is film thickness. A method of investigating this parameter using optical interferometry is presented. A partially reflecting glass plate serves as one of the bearing surfaces in a ball thrust bearing. Reflected light from both the surface of the glass and the ball produce interference fringes from which the lubricant film thickness can be measured. The interference fringes are observed stroboscopically while the bearing is in operation. The condition of lubrication at an instant in time is obtained by using high-speed flash photography. In this way the conditions of the inlet and exit regions can be observed along with the film thickness measurement. The results show oil film rupture in the exit region and the effects of a starved inlet region on film thickness.

NOTATION

a	Radius of Hertzian contact zone (point contact)
$2b$	Width of Hertzian contact zone (line contact)
E_1, E_2	Elastic moduli of solids in contact
$1/E'$	$1/2\left(\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right)$
h	Film thickness
n	Fringe order
p	Pressure
R_1, R_2	Radius of rolling elements in contact
R	Effective radius of rolling elements $\frac{R_1 R_2}{R_1 \pm R_2}$
u_1, u_2	Surface velocity of solids in x-direction
u	$1/2\ (u_1 + u_2)$
w	Load per unit length of cylinder or total load on a sphere
x	Coordinate
α	Pressure Viscosity coefficient
η	Viscosity
η_o	Viscosity at inlet conditions (controlling viscosity)
θ	Angle of incidence
λ	Wavelength of light
μ	Refractive index
σ_1, σ_2	Poisson's ratio
$\emptyset$	Correction due to phase change on reflection

INTRODUCTION

In a typical rolling element contact situation the applied load deforms the mating surfaces causing an increase in load supporting area and thus a decrease in pressure. As the normal approach of the bearing surface in the inlet region proceeds, viscous shear prevents some of the lubricant from escaping the contact area. At the leading edge of the contact area the pressure rises sharply. This pressure, which is enhanced by the increase of viscosity with pressure, is primarily responsible for the development of an elastohydrodynamic (ehd) film between the bearing surfaces. It is evident that the solution of the ehd problem is quite complex since it must simultaneously satisfy the hydro-dynamic (Reynolds), elastic, and lubricant properties equations.

An approximate solution for the line contact problem was obtained by Gruhin (1) in 1949, and a full solution was later obtained by Dowson and Higginson (2). Typical pressure distributions are shown in Figure 1 where the dimensionless parameters are defined as:

$$U = \frac{\eta_o u}{E' R} \quad , \text{ Speed parameter}$$

$$W = \frac{w}{E' R} \quad , \text{ Load parameter}$$

$$G = \alpha E' \quad , \text{ Materials parameter}$$

$$P = p/E' \quad , \text{ dimensionless pressure}$$

$$X = x/b \quad , \text{ dimensionless x-coordinate}$$

From Figure 1 it is seen that the pressure distribution differs from the Hertizian distribution for dry contact. It is evident that the ehd pressure begins to rise ahead of the Hertzian pressure for dry contact. The distance at which the pressure rises ahead of the contact area increases with the speed parameter U. The significance lies in the fact that the overall film thickness is greatly influenced by the conditions of the lubricant in the inlet region. If there is not a sufficient lubricant

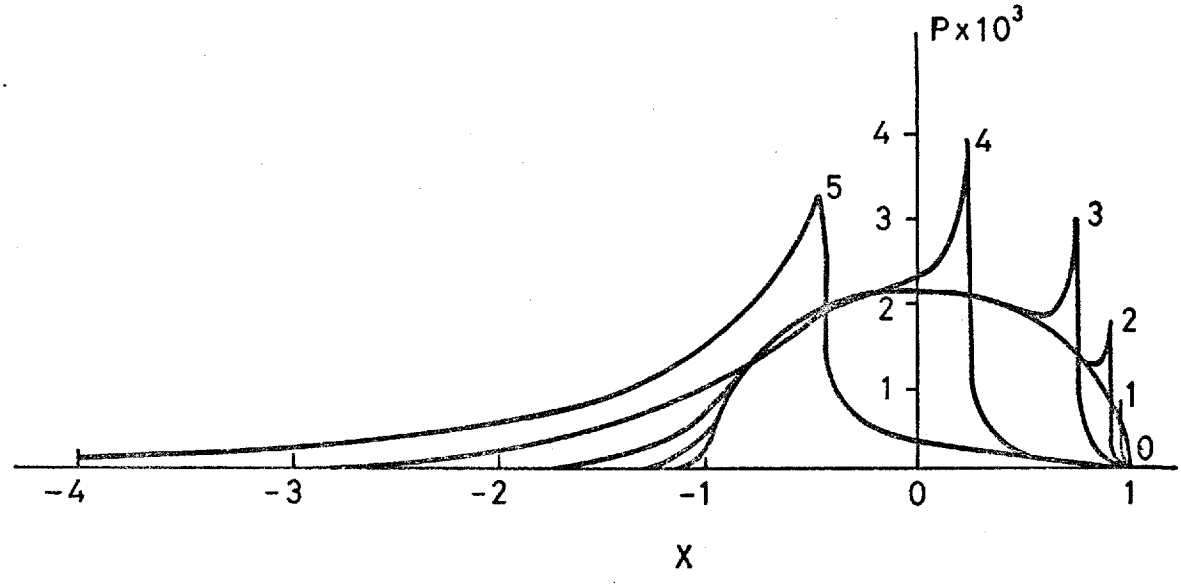


Fig.1 Pressure distributions for a compressible lubricant, from Dowson and Higginson(2): W = 3×10^{-5}, G = 5000, U = (0) 0 (dry contact), (1) 10^{-13}, (2) 10^{-12}, (3) 10^{-11}, (4) 10^{-10}, (5) 10^{-9}

supply ahead of the contact area the pressure will be forced to build up later resulting in a reduction of oil film thickness.

An important contribution provided by the theoretical solutions is the prediction of film thickness. The film thickness formula proposed by Dowson (3) for line contact is

$$H_{min} = 2.65 \frac{U^{.70} G^{.54}}{W^{.13}}$$

where U,G, and W are defined above and H_{min} is the dimensionless minimum film thickness ($H_{min} = {}^{h}/R$) taken at the constriction found near the rear of the contact region.

The ehd features of point contact are very similar to line contact. Evans (4) has recently obtained a theoretical film thickness formula for point contact using similar assumptions to those used by Gruhin (1) for line contact. The film thickness formula is

$$\frac{h_o}{R} = 1.73 \left(\frac{\alpha \eta_o u}{R}\right)^{.714} \left(\frac{w}{E' R^2}\right)^{-.048}$$

where h_o is the thickness of the film which is assumed to be parallel over the contact region. This formula agrees very well with experimental measurements obtained by interferometry.

Techniques for measuring film thicknesses such as by capacitance (5) and x-ray (6) methods have enhanced the knowledge and confidence of ehd. One of the most recent methods for studying ehd has been optical interferometry (7 - 11). The application of optical interferometry to the study of some of the fundamental and practical aspects associated with rolling element bearings is presented in this paper. Some preliminary results are given.

INTERFEROMETRY

Interferometry can be used for measuring small distances which are of the size often found in ehd contacts. The conditions for producing dark fringes in an interferometric system is given by

$$n = 2\mu h \cos \theta \qquad \text{.....(1)}$$

where n is the fringe order

λ is the wavelength of light

μ is the refractive index

h is the separation of the surfaces and

θ is the angle of incidence.

In the ball thrust bearing shown in Figure 2 the top race is made of glass. The bottom surface of the glass is coated with a thin (170 Å) chromium layer which reflects about 20% of the incident light. When an ehd film forms between the steel ball and glass race an interferometric system is formed from which part of the incident beam of light is reflected from the chromium layer and part from the surface of the ball. A variation of film thickness changes the optical path difference between the two reflected beams, and by cancellation and reinforcement of the reflected beams, light and dark fringes are formed. For normal incident illumination as shown in Figure 2 equation (1) becomes

$$h = \frac{n}{2\mu} - \emptyset \qquad \ldots\ldots(2)$$

where $\emptyset$ is a correction due to a phase change on refection.

MECHANICAL ARRANGEMENT

The mechanical arrangement is shown in Figure 2. A ball thrust bearing was chosen because a flat glass plate can conveniently serve as one of the bearing surfaces. The balls are 2.54 cm in diameter and have very good surface finish. They are separated by a fibre cage supported by small bearings in the centre. Three balls equally spaced have been used so that the load can be distributed evenly. Load is applied to the top race by a lever arm system, and the bottom race is supported by an air thrust bearing. The bottom race is driven by an electric motor while the top race is held stationary. The speed of the bottom race is measured using a magnetic transducer and displayed on a digital counter.

OPTICAL ARRANGEMENT

The interference fringes are observed through a microscope having a long working distance (29 mm) X5 objective and a X10 eyepiece. Illumination of the contact area is obtained through a collimator (Figure 2) attached to

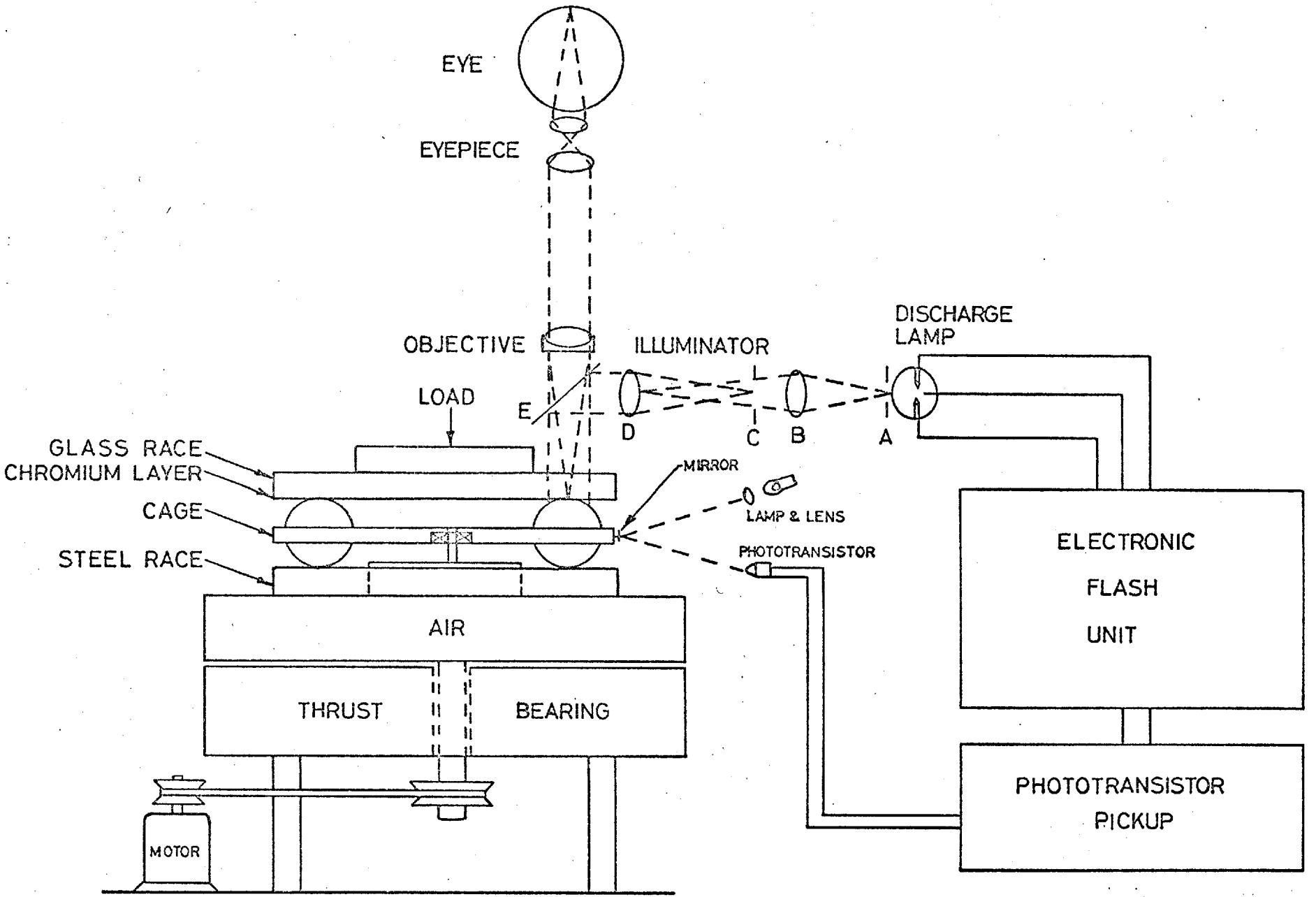


Fig.2 Schematic diagram of apparatus

the side of the microscope and below the objective. A thin glass reflector (E) directs the light normal to the contact area. An iris diaphragm (A) which is focused onto the object plane acts as a field stop and an extended source. It constitutes the first element of the collimator. The next element is a 30 mm collecting lens (E), followed by another iris diaphragm (C) used as an aperture stop. The final element is a 50 mm collimating lens (D) positioned from the aperture a distance equal to its focal length. The aperture stop (C) determines the degree of collimation as well as the amount of illumination.

The illuminating source is a xenon discharge lamp. It has a colour temperature of approximately 6000^{o}K and produces a continuous spectrum similar to daylight.

To isolate the contact area of a particular ball while the bearing is in operation the interference pattern is observed stroboscopically. The xenon flash is synchronized with the cage rotation by means of a photo-transistor. The cage is furnished with a small reflecting mirror positioned so that reflected light from a lamp and lens system energises a photo-transistor when the rolling element passes beneath the point of observation. The electrical signal from the photo-transistor then triggers the xenon flash. A high-speed photograph is taken with only one flash.

LIGHT SOURCE AND CALIBRATION OF FRINGES

Optical measurements of oil film thicknesses under various lubricating conditions (7,9) have been obtained by using either white light or monochromatic light. The thickness over which fringes remain visible depends primarily on the number and distribution of wavelengths emitted by the source. White light includes all wavelengths in the visible spectrum. It produces closely spaced fringes which provide a great amount of detail in film shape; however, clear fringes are no longer visible beyond about 10^{-4}cm in oil. Greater thicknesses can be measured with monochromatic light which consists of only a narrow range of wavelengths. Monochromatic fringes, however, give less detail in film shape.

The measurements described here were obtained by using two spectral wave bands. These were produced from the xenon source by using a filter which passes through only red and green light. This two-colour system provides a sufficient film thickness range as well as good detail in film shape.

In a monochromatic interferometric system the distance between each dark fringe is known. However, the absolute value of h corresponding to a particular fringe is often obscured by the unknown value of ϕ (see equation 2) due to the phase change on reflection.

In a multicoloured interferometric system, where the fringes are determined by their colour, further problems arise. Here the colours generally are not 'pure', but are formed by the addition, subtraction and intensity variation of the wavelengths present. The thickness h corresponding to each colour can be determined by measuring the diameters (2r) of the coloured fringes produced from the reflecting surfaces of a sphere of known radius (R) on a flat plate as in Figure 3. The thickness h can then be calculated from the known geometry by the relation

$$h = \frac{r^2}{2\mu R} \qquad \ldots\ldots(3)$$

where h, r and R are defined in Figure 3. Because of the very small thickness that h may acquire (5×10^{-6}cm) and because h is very sensitive to the measured value of r ($h \propto r^2$), two significant errors may occur. First, if a sphere and flat surface are allowed to contact each other as shown in Figure 3, a finite load that may be present will cause the surfaces to deform giving rise to a Hertzian contact radius 'a'. This error, introduced by a slight deformation of the surfaces, will tend to make the calculated values of h for each fringe higher than its actual value. Secondly, if there is nearly zero load applied to the system, the surfaces will be separated as a result of surface irregularities. This error will tend to make the calculated values of h lower than the actual values.

In an attempt to eliminate these problems, a 2.54 cm diameter steel ball (of the same kind used in the experiments) was loaded against the glass race with a balance sensitive to less than 1/10th gram. Using only the green band (produced by the superposition of another filter which eliminated the red band) the diameters (2r) of the dark fringes were measured for a series of loads between $\frac{1}{2}$ gram and 500 grams. The measured values of r for the first seven dark fringes are plotted against the calculated Hertzian radius 'a' for each load in Figure 4. At very small loads, 'a' should be very nearly independent of r so that the curve of 'a' vs r should approach the asymptote r = constant. It was noted that as the

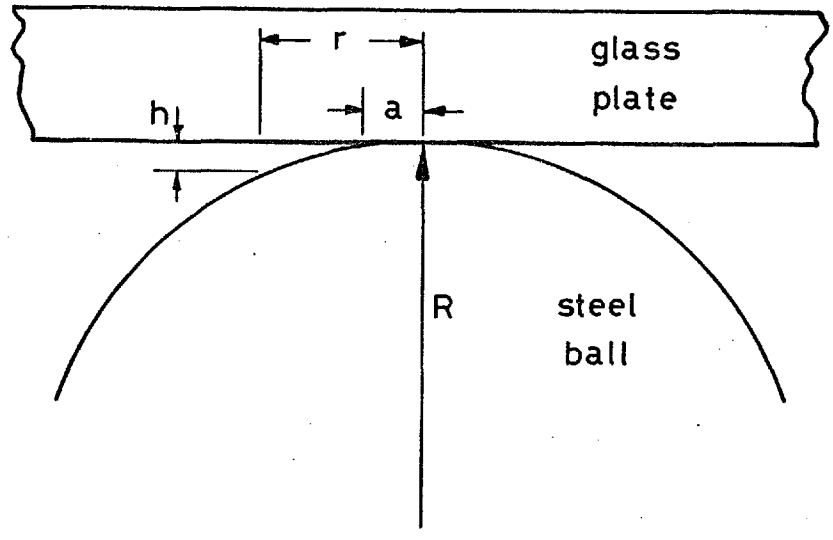


Fig.3 Calibration arrangement for calculating the thickness h corresponding to each fringe

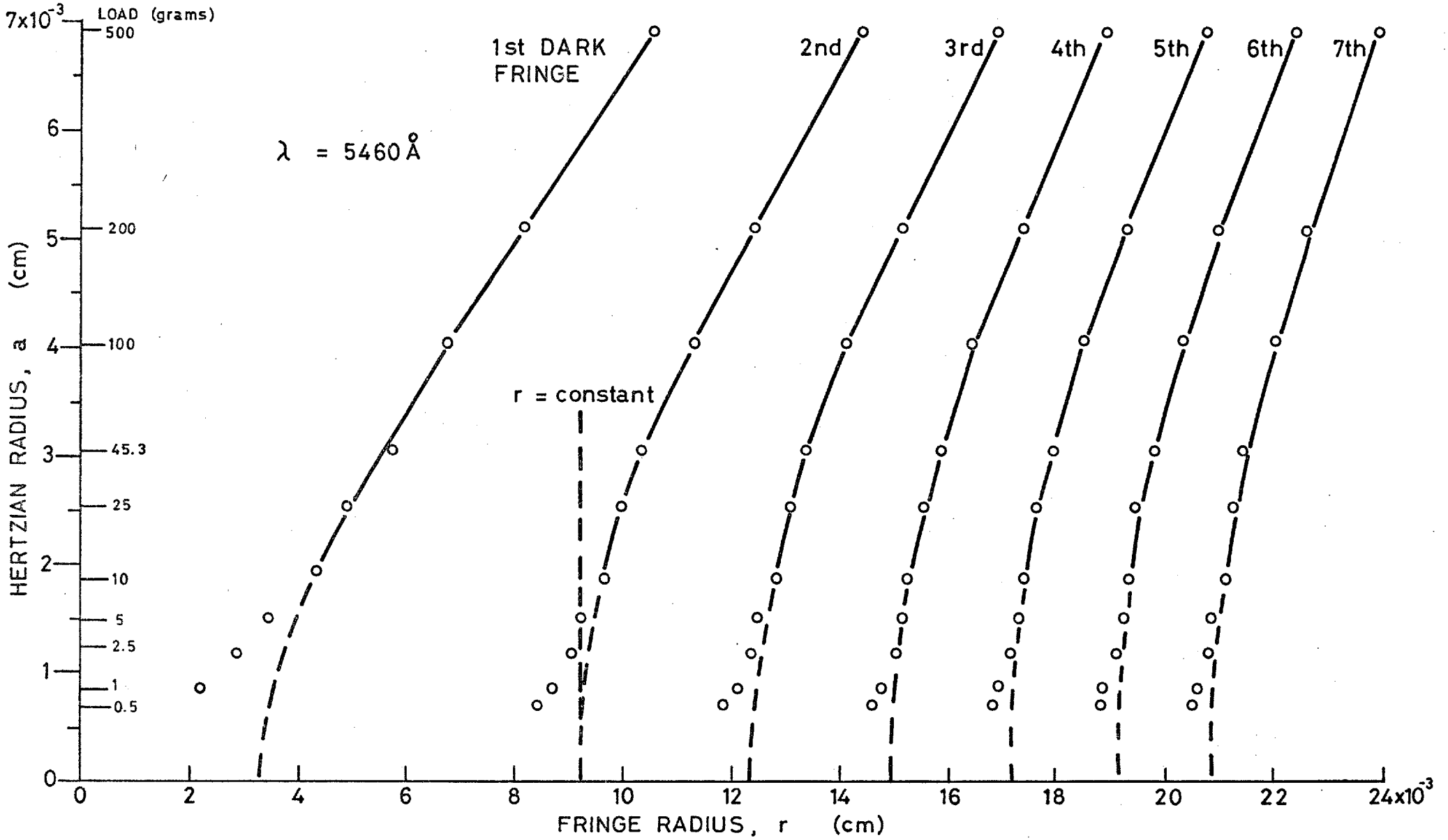


Fig.4 Calibration plot to determine r for zero load

load was decreased below 10 grams the central fringe corresponding to the Hertzian contact area changed from light grey to dark, indicating that separation due to surface irregularities was occurring. As can be seen from Figure 4, the measured values of r for loads below 10 grams deviate from the asymptote r = constant. With this in mind the seven fringe curves were extended to obtain the values of r for each fringe at zero load (i.e. as 'a'→0). In this way the values of r should not be influenced by surface roughness, surface deformation, or the correction $\emptyset$ due to phase change. The values of h corresponding to each black fringe were calculated from equation (3). Suitable corrections were then made for the two colour system.

The thickness h corresponding to each fringe depends on the refractive index μ of the medium. The refractive index for lubricating oils is about 1.5. The refractive index changes slightly under the high pressures encountered in ehd contacts. To make suitable corrections an estimate of the refractive index was obtained by using the Lorenz-Lorentz equation.

OIL FILM RUPTURE

In the entry region the normal approach of the surfaces and the viscous shear of the lubricant generate a pressure subjecting the lubricant to high compressive stresses. In the exit region the surfaces diverge, causing a decreasing pressure and thus subjecting the lubricant to tensile stresses. Since ordinary lubricants display little or no tensile strength the lubricant in the exit region ruptures as clearly shown in Figure 5.

Ordinary lubricants contain between 7 and 11% (by volume) of dissolved air at atmospheric pressure (12). The equilibrium amount of dissolved air in a lubricant is directly proportional to the absolute pressure and varies inversely with temperature. In a rolling element bearing where the lubricant may be agitated by moving parts a certain amount of entrained air may also be present. This entrained air is dispersed in the lubricant in the form of bubbles. Also, air in the form of microscopic or submicroscopic bubbles may be present at solid surfaces (12). When the lubricant pressure is reduced below the saturation pressure the dissolved air diffuses into the small air bubbles attached to the surfaces

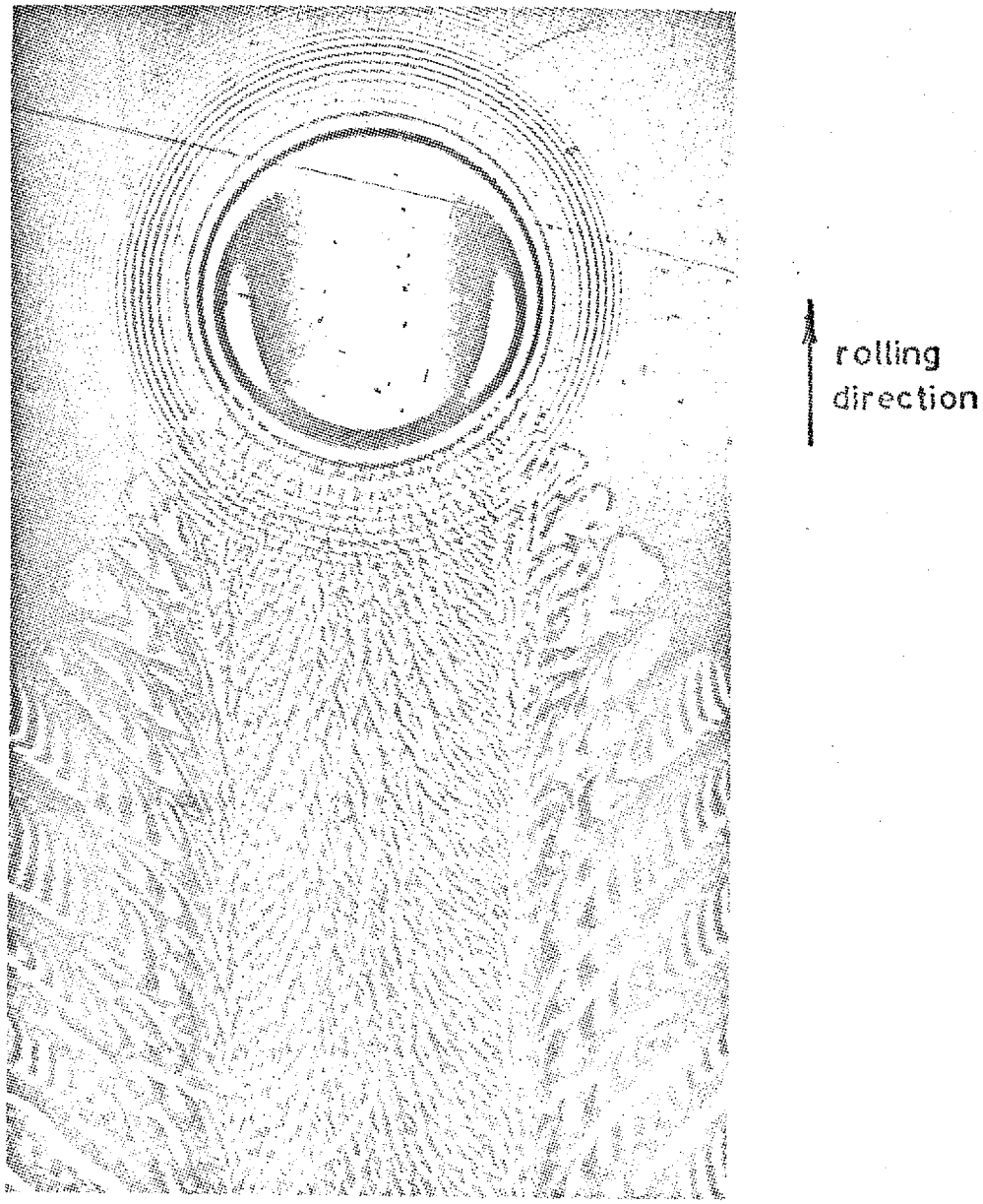


Fig.5 Oil film rupture

or entrained within the lubricant causing the bubbles to grow in size. This is usually referred to as "gaseous cavitation". If the pressure is reduced below the vapour pressure of the lubricant, the bubbles fill up with vapour giving rise to "vaporous cavitation". A small air bubble in the lubricant thus provides a convenient interface through which dissolved gas can diffuse and liquid lubricant can evaporate. If there is no liquid-air interface present, no diffusion or evaporation can take place, and the lubricant may thus acquire a very high tensile strength.

When the lubricant that passes through the contact area reaches the diverging section of the exit region it ruptures, forming a track of thin films of lubricant and air bubbles. The air bubbles appear to nucleate from the surface and when a scratch is present they preferentially nucleate along the length of the scratch. On both sides of the track, where a greater supply of lubricant is usually present, the air bubbles continue to grow. The thickness of these bubbles can be obtained from the interference fringes produced from the partially reflecting surfaces of the glass race and the air-oil interface, and the interference fringes which form between the ball and glass race. The air bubbles almost completely fill the gap, except for a thin sheet of lubricant attached to the surfaces.

OIL STARVATION

The development of an elastohydrodynamic film between two surfaces is very much influenced by the conditions of the lubricant in the inlet region. It is here, where the viscosity and pressure are relatively low, that the overall film thickness is primarily determined. The conventional ehd theory assumes the inlet region to have a sufficient supply of lubricant. Furthermore, the experimental programmes concerned with the fundamentals of ehd have always taken suffieient steps to provide an adequate lubricant supply. However, in many rolling element bearing applications such as instrument bearings where low stable friction is of prime importance the supply of lubricant may be very limited. Even in "well lubricated" bearings the quantity of lubricant in the vital inlet region may be severely restricted. If the inlet region is starved of oil the inlet pressure as shown in Figure (1) will be forced to build up later. A reduction in film thickness will therefore follow since the overall film thickness is greatly influenced by the conditions in the inlet region.

The importance of the quantity of lubricant in the inlet region can be seen in Figure 6. In this experiment the thrust bearing was run at constant speed and load. When the bearing is in operation, lubricant distributed over the inlet bearing surfaces merge forming an oil meniscus around the contact area as shown in Figure 6b. In Figure 6a where the meniscus is quite large, the inlet region is flooded with oil thus providing sufficient room for the inlet pressure to build up. A cross-sectional film thickness profile perpendicular to the direction of rolling is shown in Figure 7. As the oil meniscus becomes smaller the film thickness and shape become increasingly more sensitive to the inlet lubricant boundary as shown in Figures 6c and 6d. The film thickness profile of Figure 6d is also shown in Figure 7 and the reduction in film thickness due to oil starvation is obvious.

It is now quite evident that the quantity of lubricant in the inlet region may play a significant role in the determination of the ehd film thickness. The thin films of lubricant and air bubbles left in the track as shown in Figure 5 certainly do not enhance the inlet region of the following ball. This effect is clearly seen in the inlet region of Figure 6d. If oil starvation becomes too severe eventual asperity contact between ball and raceway will occur resulting in inferior bearing performance.

CONCLUSIONS

Optical interferometry provides a convenient tool for studying the fundamental aspects of elastohydrodynamic lubrication. Its application to rolling element bearings can be useful for investigating some of the practical and fundamental problems associated with the lubrication of rolling element bearings. This technique provides a means of accurately measuring the film thickness and shape of a rolling element in a dynamic situation. It also provides a means of observing how the environmental conditions of the inlet and exit regions can influence the oil film shape and thickness.

The experimental work presented is of a preliminary nature and is by no means complete. However, it can be concluded that the development of an ehd film in a rolling element bearing can be greatly influenced by the quantity of lubricant in the inlet region.

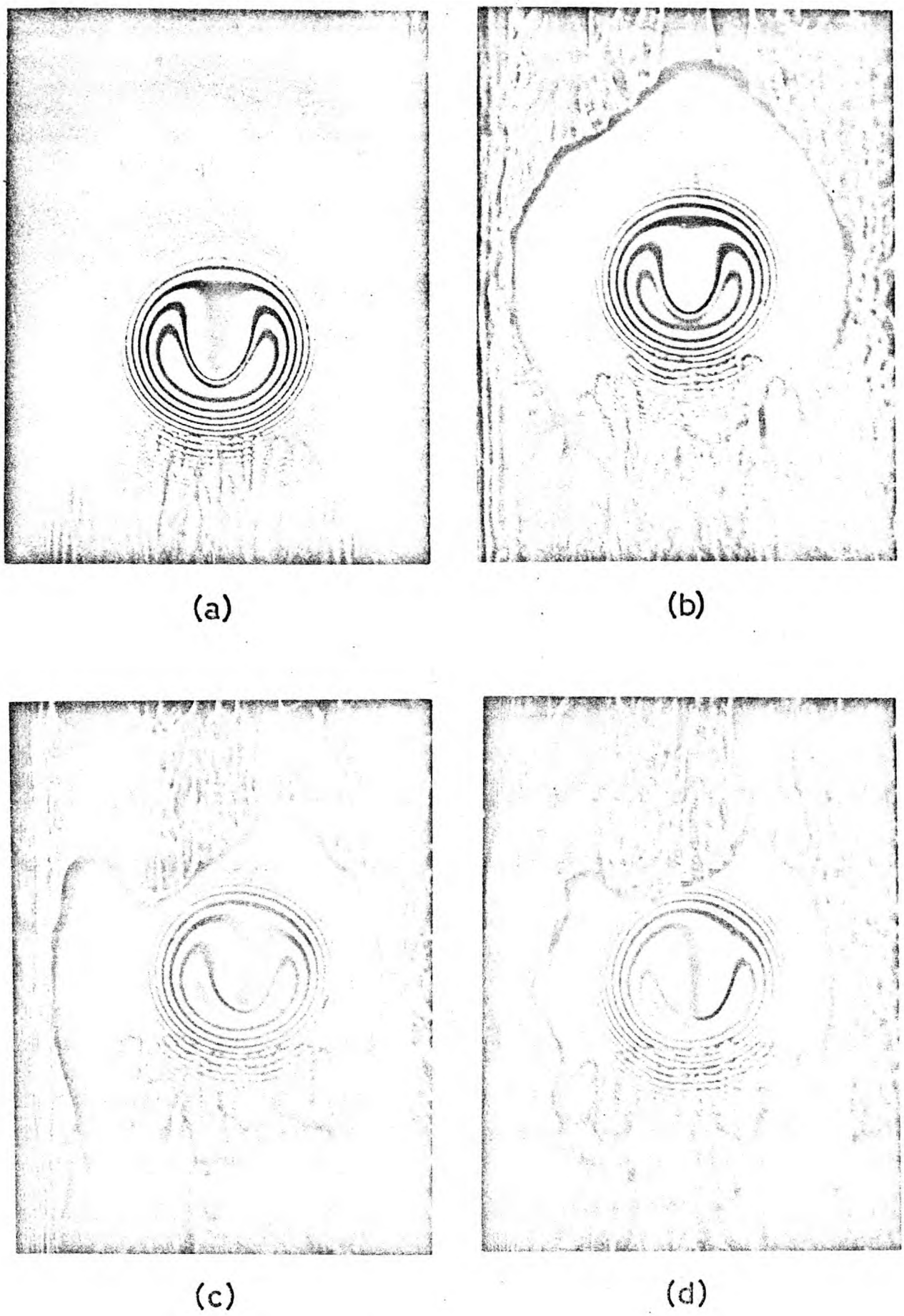


Fig.6 Effect of lubricant supply on film thickness

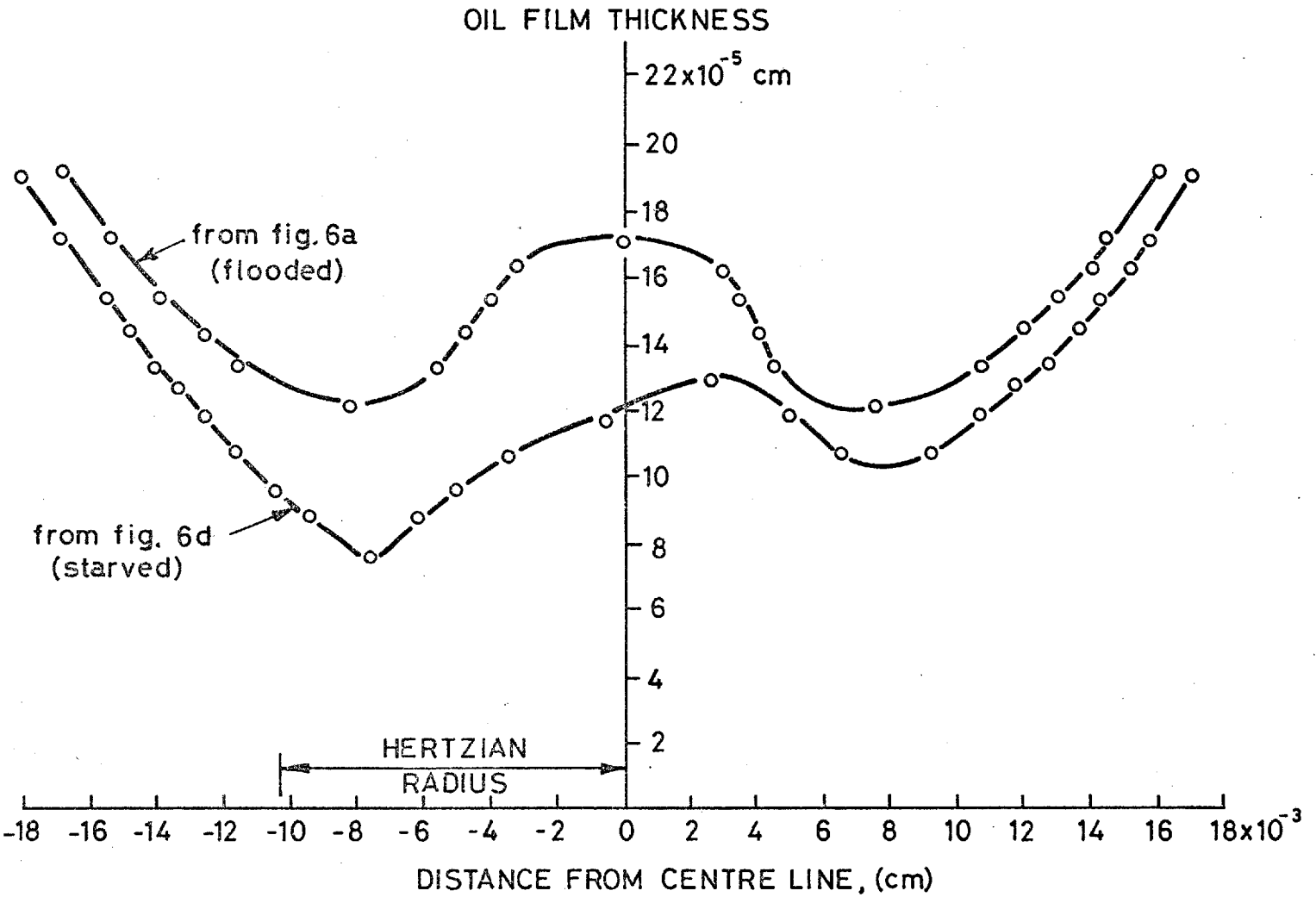


Fig.7 Effect of oil supply on film thickness and shape; cross section taken perpendicular to direction of rolling

REFERENCES

1. Grubin, A.N. and Vinogradova, I.E., "Investigation of Scientific and Industrial Research", Central Scientific Research Institute for Technology and Mechanical Engineering, 1949, Book No. 30, pp. 115-166. (D.S.I.R. Translation, No. 337).

2. Dowson, D. and Higginson, G.R. "A Numerical Solution to the Elasto-Hydrodynamic Problem", J. Mech. Engr. Sci., 1959, 1, No.1, pp. 6-15.

3. Dowson, D., "Elastohydrodynamics", Proc. Inst. Mech. Engrs., 1967-68, 182 (Pt 3A).

4. Evans, D., Imperial College of Science and Technology, to be published.

5. Dyson, A., Naylor, H. and Wilson, A.R., "The Measurement of Oil-Film Thickness in Elastohydrodynamic Contacts", Proc. Inst. Mech. Engrs., 1965-66, 180, Pt 3B, p. 119.

6. Sibley, L.B. and Orcutt, F.K. "Elastohydrodynamic Lubrication of Rolling Contact", ASLE Trans. 4, 134-149 (1961).

7. Gohar, R. and Cameron, A., "Theoretical and Experimental Studies of the Oil Film in Lubricated Point Contacts", Proc. Roy. Soc., Series A, 291, 520-536.

8. Gohar, R. and Cameron, A., "The Mapping of Elastohydrodynamic Contacts", ASLE/ASME Conference, Minneapolis, (Oct. 1966), ASLE Paper No. 66-LC-21.

9. Foord, C.A., Hammann, W.C., Cameron, A., "Evaluation of Lubricants Using Optical Elastohydrodynamics", ASLE Paper No. 67-LC-12, 1967.

10. Westlake, F.J. and Cameron, A., "A Study of Ultra-thin Lubricant Films Using an Optical Technique", Proc. Inst. Mech. Engrs., 1967-68, 182, Pt 3G, Report 12, pp. 75-78.

11. Wedeven, L. and Cameron, A., "A Study of Elasto-hydrodynamic Lubrication in Rolling Element Bearings Using Optical Interference", Proc. Inst. Mech. Engrs., 1967-68, 182, Pt 3G, pp. 87-89.

12. Schweitzer, P.H., Szebehely, V.G. "Gas Evolution in Liquids and Cavitation", Journal of Applied Physics, 21, No. 12, December 1950, 1218-1224.

THE INSTITUTION OF
MECHANICAL ENGINEERS

TRIBOLOGY GROUP
EDUCATION AND TRAINING GROUP

OPTICAL ELASTOHYDRODYNAMICS

C. A. FOORD, B Sc, Ph D
L. D. WEDEVEN, B S, M S M E
F. J. WESTLAKE, B Sc (Eng)
A. CAMERON, Sc D, D Sc (Eng), Ph D, C Eng, M I Mech E

Interferometric measurements of the elastohydrodynamic point contact shape are presented for rolling and sliding conditions. Optical design considerations are given in connection with illumination and reflectivity of the bearing surfaces.

Paper intended to be presented at an Ordinary Meeting of the Institution, in London, on Wednesday, 4th February 1970, at 6 p.m. Communications on this paper are invited for publication in the Proceedings. Contributors should read the instructions overleaf.

P28/70 *Advance copy subject to revision.*
The paper and discussion will be published in Proceedings 1969–70, Volume 184, Part 1

PUBLISHED BY THE INSTITUTION · 1 BIRDCAGE WALK, WESTMINSTER, LONDON SW1

INSTRUCTIONS FOR CONTRIBUTORS

State your degrees, honours and grade of membership, if any.

Draft your contribution in direct speech. Contributions should generally not exceed 500 words and two illustrations.

Provide two copies typed in double-spacing with wide margins, on one side of the paper only.

Provide one unlettered tracing or drawing in ink of each line illustration, and two prints bearing the captions. Dyeline prints are not suitable for reproduction.

Provide two prints, on glossy paper, of each photograph.

Contributions, which should be addressed to J. C. Mundy, Assistant Editor of Proceedings, must *reach* the Institution by 23rd March 1970 in order to meet the press date. Late contributions cannot be accepted.

TRIBOLOGY GROUP
EDUCATION AND TRAINING GROUP

OPTICAL ELASTOHYDRODYNAMICS

C. A. Foord, BSc, PhD* L. D. Wedeven, BS, MSME* F. J. Westlake, BSc(Eng)*
A. Cameron, ScD, DSc(Eng), PhD, CEng, MIMechE*

In Part 1 an extended series of pictures are given showing the shape of a point elastohydrodynamic contact. Extremely good fringe visibility is obtained by taking account of the reflectivity of the semi-reflecting chromium layer on the glass plate and the reflectivity of the steel balls. The shape of the contact is quite different from the Hertzian shape frequently assumed. Figures are given for pure rolling, pure sliding and skewed contact giving a slide/roll ratio of 0·16. Careful collimation permits the 30th fringe order to be observed.

In Part 2 detailed design considerations needed to give sharp single and multiple beam interference pictures are studied. Subjects considered are the qualities of the rolling pair, the degree of collimation and the design of collimators. The choice of reflecting layer and light sources is discussed. Methods of estimating phase change on reflection are described. A technique making possible film thicknesses down to 100 Å (0·4 μin) is given.

INTRODUCTION

THE first significant publication on this topic was Cameron and Gohar in 1965 (1)†. They loaded a lubricated rotating steel ball against a stationary plate of high refractive index glass and obtained interference patterns which were the first to show the now classic horseshoe constriction of elastohydrodynamic point contact. Their use of sliding contact and special glass imposed a severe restriction on loads, speeds and fringe quality. These limitations were later reduced by using sapphire and diamond as the transparent member as reported in a paper they presented at Minneapolis in 1966 (2). This paper also contained important pictures of the effect of roller blending. A great improvement in fringe quality was obtained by Foord *et al.* (3) who, instead of relying on differences in refractive index, used a 20 per cent reflectivity layer of chromium which allowed the transparent material to be selected for its mechanical properties and enabled them to use pure rolling and high speeds.

More recent developments are the use of high speed ciné photography to observe transient events, by Dowson and Jones (4), and Westlake and Cameron (5) who used a laser to observe even faster events. Wedeven and Cameron (6) used high speed flash photography which enabled them to study oil film rupture in detail. The fact that the horseshoe constriction can exist with isoviscous fluids was shown by Roberts and Tabor (7) using a rubber sphere at maximum Hertz stresses of 10 lbf/in².

The purpose of Part 1 is to describe a successful rig which has been used to obtain useful results already published (3), and to present additional results which cover a much greater speed and load range, as well as varying amounts of sliding. In Part 2 the design philosophy of optical interference systems for elastohydrodynamic studies is discussed in some detail, and techniques for improving the definition and accuracy beyond that achieved in Part 1 are described and illustrated.

The MS. of this paper was received at the Institution on 10th April 1969 and accepted for publication on 25th November 1969. 2

* *Department of Mechanical Engineering, Imperial College of Science and Technology, Exhibition Road, London, S.W.7.*

† *References are given in the Appendix.*

Notation

A	Absorption coefficient.
a	Radius of Hertz contact.
c	Specific heat.
$E_{1,2}$	Young's modulus of materials 1, 2.
E'	Reduced modulus, $\frac{1}{E'} \equiv \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}$.
H	Layer thickness of optical spacer.
k	Thermal conductivity, cal/(cm² s degC/cm).
m	Integer.
n	Refractive index of fluid.
p	Pressure.
p_{max}	Maximum Hertz pressure.
r	Radius of interference fringe.
R	Ball radius, reflection coefficient.
t	Thickness of fluid film.
U	Total rolling velocity $(U_1 + U_2)$.

V Fringe visibility.
W Load.
δ_m Fractional fringe width.
η Viscosity.
θ Angle of incident ray.
λ Wavelength.
$\nu_{1,2}$ Poisson's ratio materials 1, 2.
ϕ_{cr} Phase change at chrome–oil interface.
ρ Specific gravity.
τ Temperature, °C.

Definitions

Å Ångström (1 Å = 10^{-8} cm).
mW Milliwatt (laser power).

Part 1: Optical elastohydrodynamic tests

The apparatus is shown in Fig. 1. A specially coated glass plate is mounted on an air bearing and driven by a superfinished 1-in steel ball which is mounted in a polytetrafluorethylene (p.t.f.e.) cup*. The ball is loaded against the lower surface of the disc by an air cylinder, and rotated by a variable speed electric motor. Oil is drip fed on to the ball and drawn into the contact, where an elastohydrodynamic film is generated and its thickness measured by optical interference.

The optical system uses a microscope with a ×5 objective and ×10 eyepiece giving 50 times magnification in all, and a specially designed collimator, described in Part 2. As the oil films of interest are thin and relatively low magnifications are adequate, the critical points in achieving good fringe quality are the intensities of the interfering beams, and vibration-free mechanical design to give stationary fringe patterns.

The fringe visibility is governed by the intensity distribution within the fringe pattern. If the collimation and coherence are adequate the intensity distribution due to two interfering waves is:

$$I = I_1+I_2+2\sqrt{I_1I_2}\cos\left(\frac{2\pi}{\lambda}P\right)$$

where I_1 and I_2 are the individual intensities of the two beams, λ the wavelength, and P the path difference. Thus

$$I_{\min}^{\max} = I_1+I_2\pm2\sqrt{I_1I_2}$$

The visibility V is given by

$$V = \frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}} = 2(I_2/I_1)^{1/2}/(1+I_2/I_1)$$

Hence for 100 per cent visibility the intensities I_2 and I_1 should be equal.

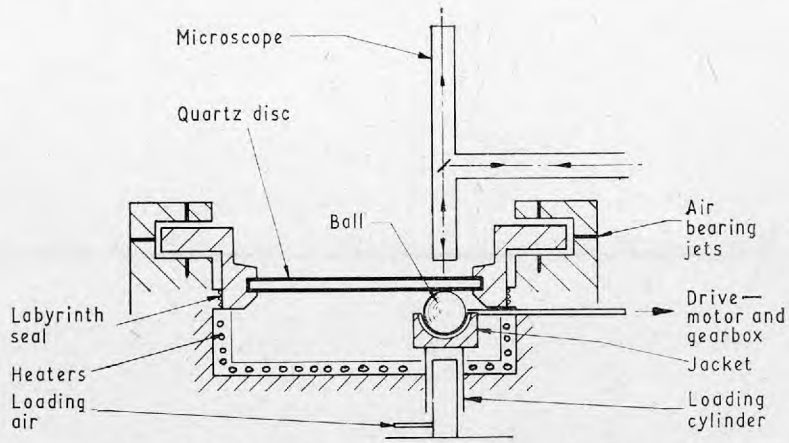
The reflectivity at an interface is

$$R = \{(n_1-n_2)/(n_1+n_2)\}^2$$

where n_1 and n_2 are the refractive indices of the two media. As oil and glass have similar refractive indices no reflection will occur with oil present. To overcome this problem Gohar and Cameron (1) used a special high refractive index glass ($n = 1{\cdot}94$) and this gives a glass/oil reflectivity of 1·3 per cent (assuming $n_{oil} = 1{\cdot}54$ for high pressure oil). The reflectivity of the steel balls is about 60 per cent, which means that the visibility calculated as above is 29 per cent, and explains their poor fringe quality.

The disadvantages of the high refractive index glass are overcome by using a disc of $\frac{1}{2}$-in thick plate glass coated with a 20 per cent reflectivity chromium film which, as shown in Part 2, gives good fringe visibility. Also it is about one-fifth of the cost of the high refractive index disc. More than 25 fluids have been tested on one track of the

* *Glacier DU material.*



4-in diameter elastohydrodynamic machine; maximum working temperature 60°F.

Fig. 1. Sketch of apparatus

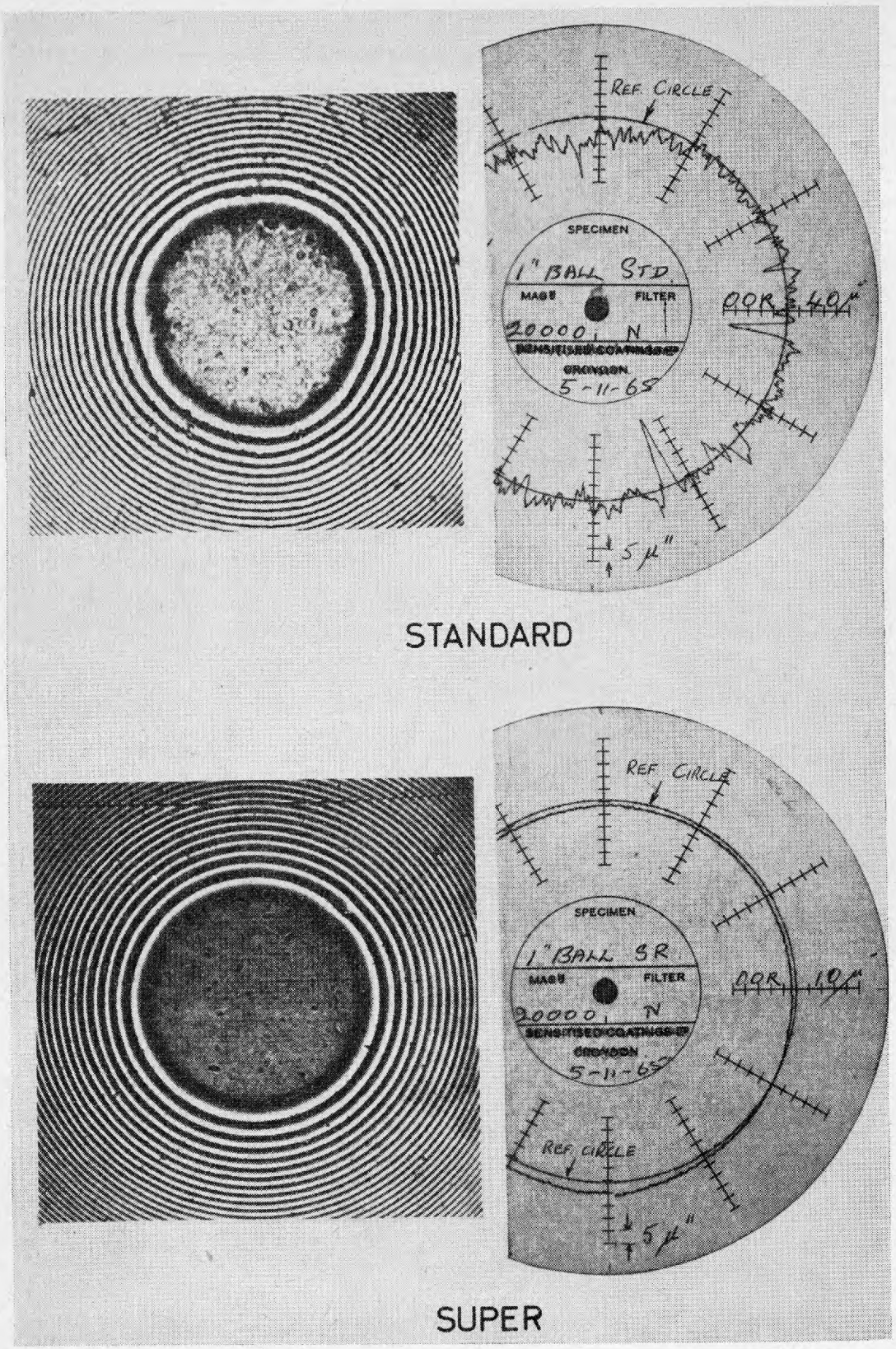


Fig. 2. Talysurf traces and stationary Newton's rings for standard and super-finished balls

disc, and some sliding tests in which the disc is stationary have been carried out, without much damage to the chromium.

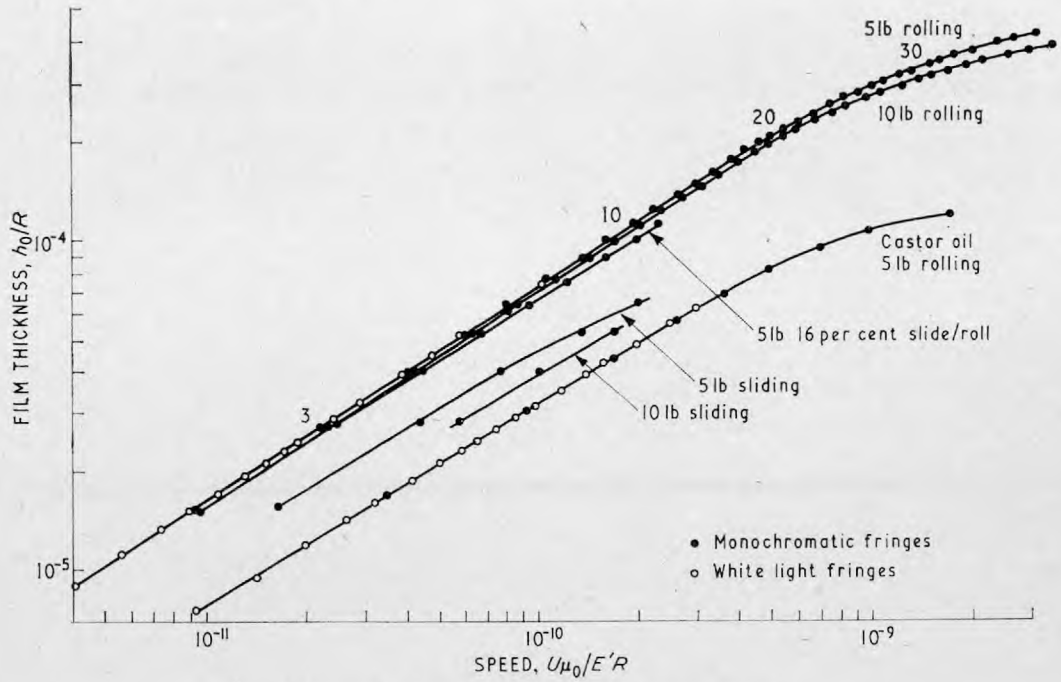
The ball is located in a p.t.f.e. cup which is mounted on a 1-in piston which rides freely in a cylinder. The cylinder is fed with compressed air, thus loading the ball. The reasons for choosing this system are that the load is very easily controlled by a pressure regulator, and that the system has low friction and inertia which enables the ball to follow any unevenness in the disc. The piston is hollow to keep its weight down. The centre line of the ball is about 0·010 in behind the centre line of the piston, so that when the ball rotates the cup will 'caster' behind the piston and take up a stable position.

To reduce frictional heating and remove the danger of embedded debris scoring the ball, much of the lining of the cup is scraped away leaving four pads to support the ball. Most of the heat generated by friction in the p.t.f.e. cup is carried away by water from a thermostatted bath which is circulated below the cup. A fine chromel–alumel thermocouple is attached to the cup and projects into the oil film ahead of the contact. This method of temperature measurement is robust and was found to be adequately sensitive.

The main disadvantage of this loading system is that heat is generated by shearing of the oil film in the cup. Other systems have been considered but so far the balance of advantage lies with the present one. In particular, any system which involves air in contact with the ball presents problems of the oil being sprayed everywhere and wasted, and also necessitates control of the air temperature. Hydrostatic systems require an oil feed circuit and therefore use far more oil than the 5 ml used by the present system; they are also difficult to clean between tests. The only other class of system is one in which a ball is mounted on, or made integral with, a shaft, and this shaft mounted in suitable bearings. Such a system overcomes the disadvantages of the others but would be extremely difficult and expensive to produce to the required standard of accuracy and ball surface finish, especially as the ball has to be replaced fairly frequently. A ball has been used for routine testing of as many as 25 fluids; but with very thin fluids, high loads or sliding tests the life may only be 2 or 3 test runs. The present balls are round to within 4 μin, which means that the run-out will be of this order when they are rotated in the present mounting system. Fig. 2 shows Talysurf traces for standard and super-finished balls (this super ball is from an earlier batch and is 10 μin ovality on radius). Stationary Newton's rings for each ball are also shown.

The ball is driven by a variable-speed electric motor, through a Hooke's joint flexible coupling and a drive shaft which is cemented on to the ball with an epoxy resin.

Owing to the great range of fluid viscosities a motor speed from 1 to 6000 rev/min is required. This is achieved by using a variable-speed motor which will run satisfactorily between 100 and 6000 rev/min and either a direct drive or a 10/1 or 100/1 reduction gear. Provided that care is taken to screen the wires and earth the motor and rig to minimize interference, speed can be very conveniently measured to within 1 per cent over the entire range, by standard digital counters.



Lubricant 5-phenyl-4-ether unless otherwise stated.

Fig. 3. Dimensionless film thickness–viscosity speed parameter

The slip between ball and plate has been measured for a wide range of conditions as described in reference (3) and found to be less than 1 per cent under all conditions.

TEST PROGRAMME

The apparatus was used with white light illumination (3) to test a wide range of fluids at a single load over a limited speed range. In this paper the use of a viscous oil enables the apparatus to be run at much higher dimensionless speeds before temperature rise becomes a problem (Fig. 3), when the thicker films are generated monochromatic light (mercury green in this instance) must be used.

The black and white photographs (Figs 4 and 7) are taken with a special purpose very high speed film, but better results have since been obtained using a widely available high speed panchromatic film.

Tests in which the slide/roll ratio (defined here as $(u_1-u_2)/(u_1+u_2)$) is unity are carried out by rotating the ball against a stationary disc. The friction of the sliding contacts is determined by measuring the torque required to prevent the rotation of the air bearing.

If the axis of rotation of the ball is moved through an angle θ the rolling velocity becomes $u_1\sqrt{1+3\cos^2\theta}$ and a sliding velocity of $u_1 \sin\theta$ is introduced. The rolling velocity is the vector sum, and the sliding velocity the vector difference, of the surface velocities of the ball and disc as shown in Fig. 5; u_1 is the surface velocity of the ball, and u_2, the velocity of the disc, is equal to $u_1 \cos\theta$.

A test was run using 5-phenyl-4-ether (5P4E) as lubricant with an angle θ of 18° which gives a slide/roll ratio of 0·16.

Cameron (8) also used an angled drive to produce sliding.

RESULTS

The film thickness over a wide range of speed is plotted in dimensionless form in Fig. 3. The curves for rolling with polyphenyl ether (5P4E), which is a very thick fluid (almost 2000 cP viscosity), begin to deviate from a straight line at about the 20th fringe point although no appreciable temperature rise was noted until after the 30th fringe. This deviation may be due to viscoelastic behaviour, or due to a temperature rise in the film itself, which cannot be measured. In all the tests of reference (3) deviation from a straight line was associated with a change in temperature which could be detected. In this case, deviation in the absence of temperature rise suggests that non-Newtonian behaviour may be significant; a further possibility is that of oil starvation (Wedeven (9)). The castor oil curve starts to deviate slightly from the straight line around the 7th black and curves much more rapidly after the 9th black when the temperature rises appreciably.

The photographs in Fig. 4 show the film shapes at 5 and 10 lb loads at the 3rd, 10th, 20th, and 30th black points, which are marked on Fig. 3. They show that at constant load the point of minimum film thickness moves from the side lobes to the back as speed increases; with a 5 lb load this occurs at the 15th black fringe and with a 10 lb load at the 21st black fringe. Profiles in the direction of rolling (x co-ordinate) are plotted in Fig. 6 for the conditions of Fig. 4. Because of the difficulty of estimating the pressure distribution and therefore the refractive index change in the thicker film cases, all the profiles of Fig. 6 are plotted using the refractive index at atmospheric pressure. The increase in density of this fluid at the maximum Hertz pressure of 66 000 lbf/in² is 16 per cent which gives a refractive index increase of 10 per cent calculated by the Lorenz–Lorentz relation. As the density increase of most fluids, except silicones and fluorocarbons, levels off at about this value at high pressures, it is unlikely that refractive index change will be greater even in the region of any pressure spike. After making full allowance for this change the constriction at the back of the contact is always greater than 20 per cent of the film thickness. The viscous constriction must impose very severe stresses on the material, which implies that beyond a certain value an increase in film thickness may do more harm than good.

Table 1. Fluid properties

Fluid	Viscosity at τ, cP	Test temperature τ, °F	A.S.-T.M. slope	Density at τ, g/ml	Refractive index at τ	α_τ, in²/lbf
5-Phenyl-4-ether	1950	77	0·90	1·19	1·63	4·0
Castor oil	720	77	0·72	0·955	1·482	1·09

A point of interest is that the film thickness at the back of the contact, which under high-speed low-load conditions is the minimum, is nearly independent of load; whereas the film thickness in the side lobes, which is the minimum under low-speed high-load conditions, decreases rapidly with load as shown in reference (3) and is only 33 per cent of the thickness in the centre for the example

$$\frac{U\mu_0}{E'R} = 4{\cdot}38\times10^{-11}, \quad \frac{W}{E'R^2} = 2{\cdot}3\times10^{-6}$$

For most lubricants the empirical formula

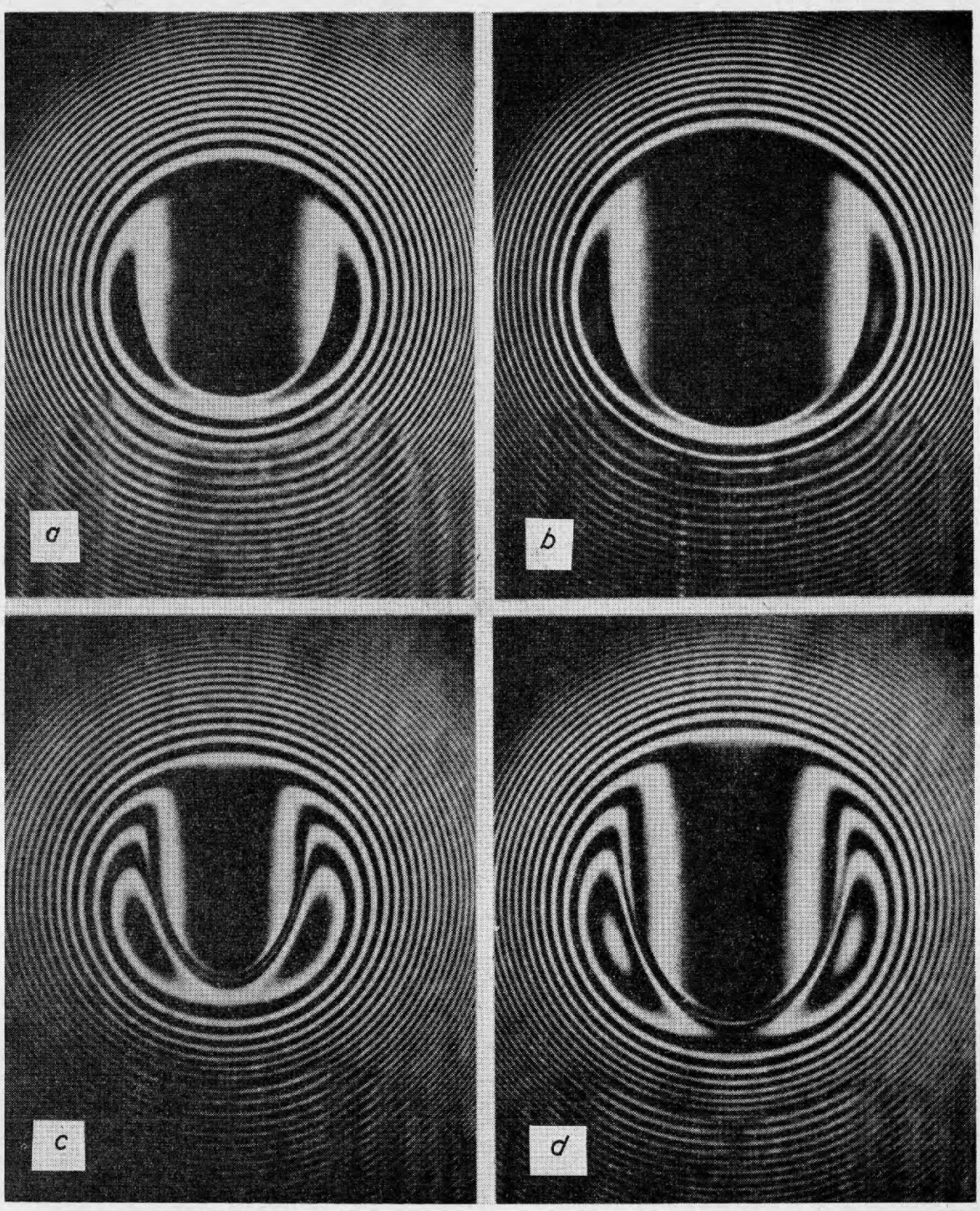
$$\left(\frac{h}{R}\right) = 0{\cdot}86\left(\frac{U\mu_0}{E'R}\right)^{2/3}\left(\frac{W}{E'R^2}\right)^{-0{\cdot}05}(\alpha E')^{0{\cdot}6}$$

is fairly accurate for the film thickness in the flat central area of the contact, although the measurements in reference (3) show the index of speed to vary between 0·55 and 0·7 for different fluids. The index is generally higher than the 0·55 of the crossed cylinder measurements of Archard and Kirk (10); the reason for this is that a 90° crossed cylinder machine has a slide/roll ratio of 1.

For the loads used in practical applications the minimum film thickness should be taken to be $\frac{1}{3}$ of the value calculated by the above formula. For contact between non-spherical bodies the equivalent radius is

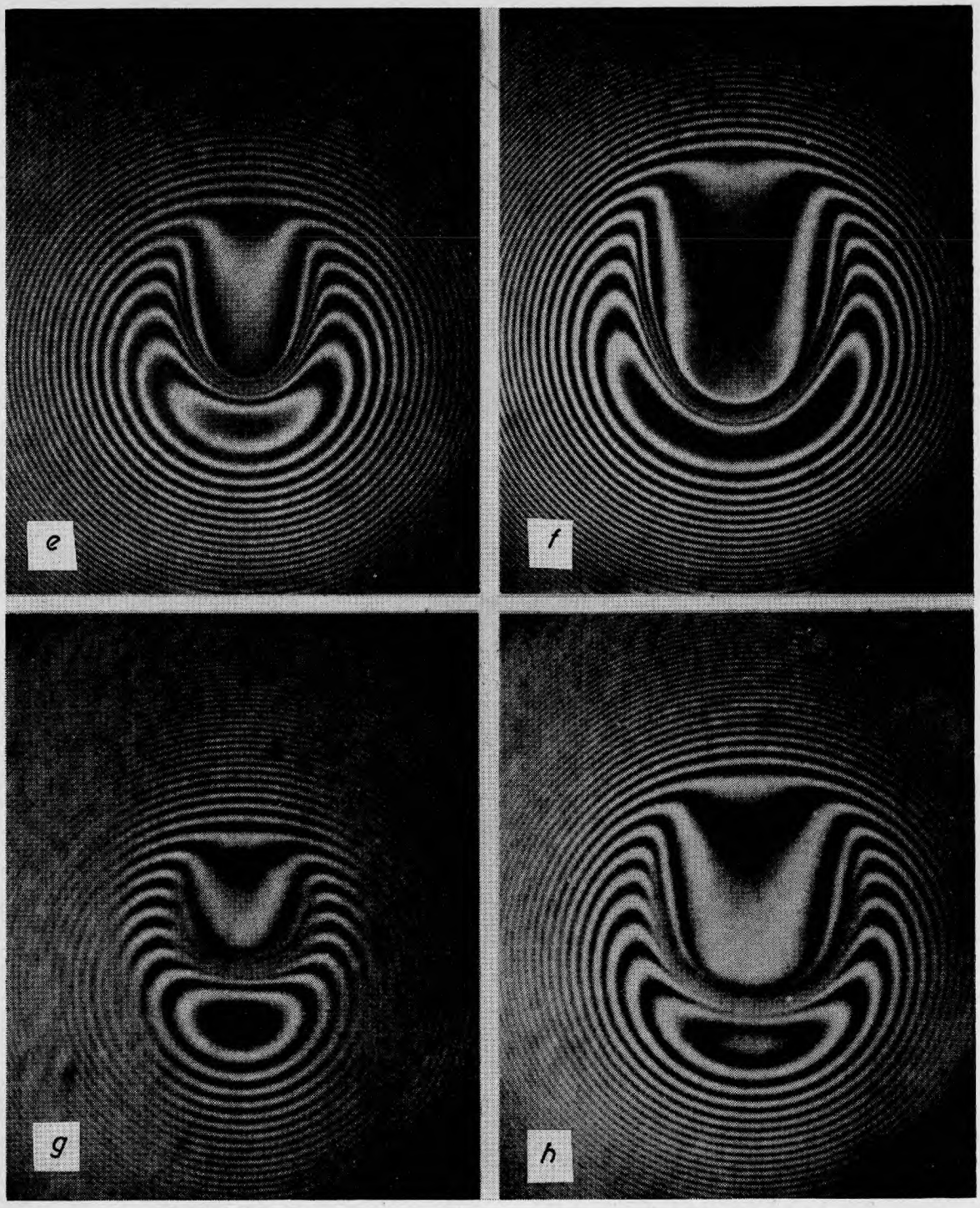
$$1/R = 2/\Sigma R_i$$

where ΣR_i is the sum of the principal curvature of the contacting bodies. The curvature sum of ball bearing



a 5 lb 3rd black.
b 10 lb 3rd black.
c 5 lb 10th black.
d 10 lb 10th black.

Fig. 4. Rolling contact,



e 5 lb 20th black.
f 10 lb 20th black.
g 5 lb 30th black.
h 10 lb 30th black.

lubricant 5-phenyl-4-ether

*

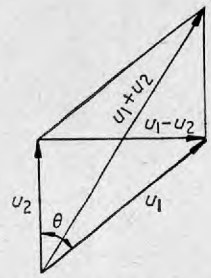


Fig. 5. Vector diagram for angled drive for standard and super balls

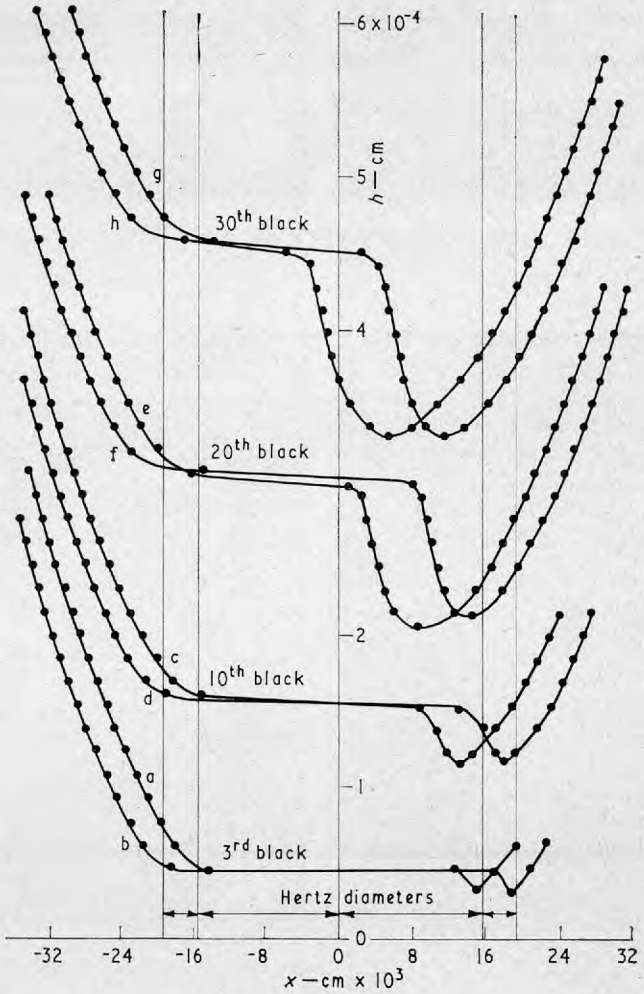
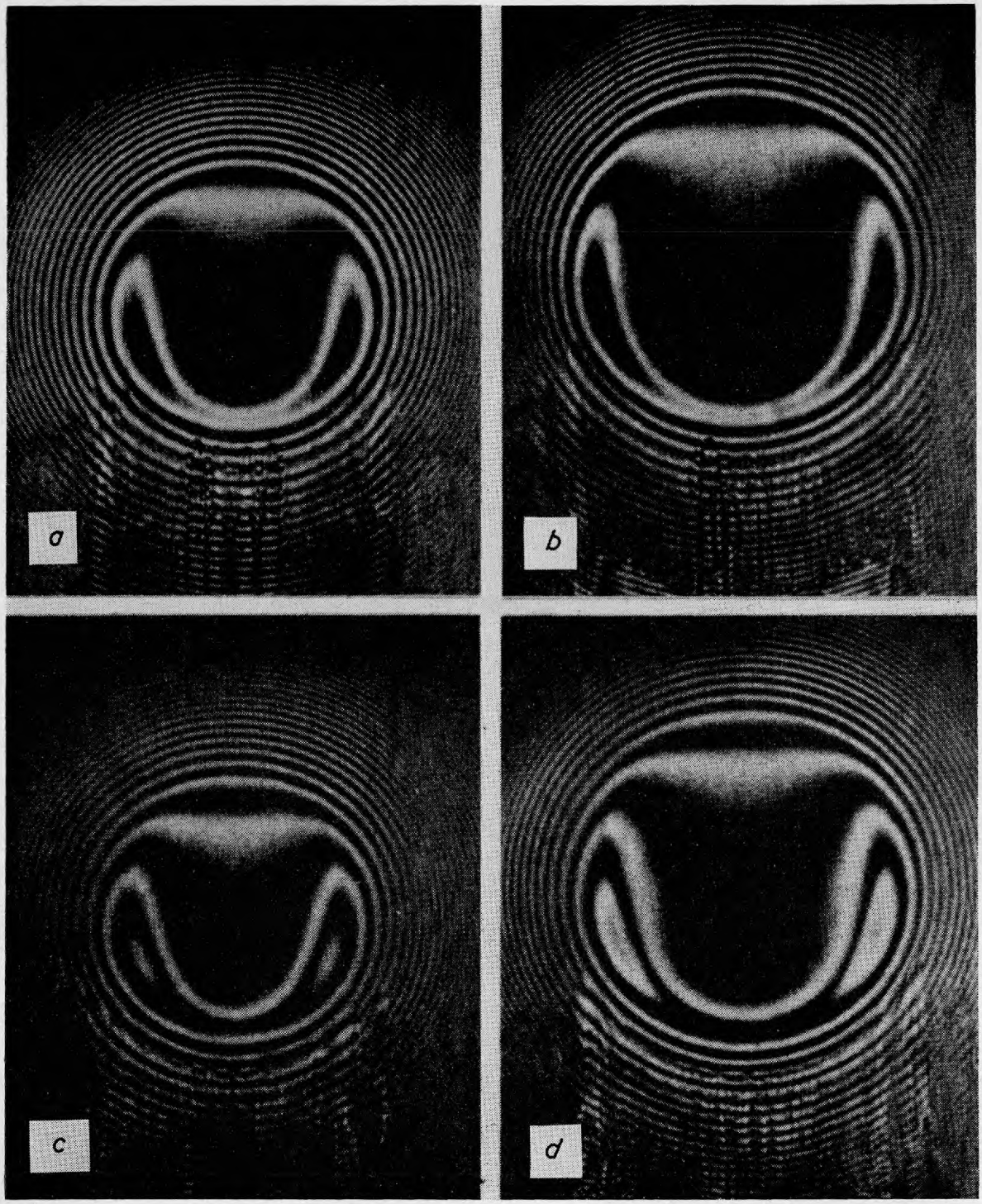
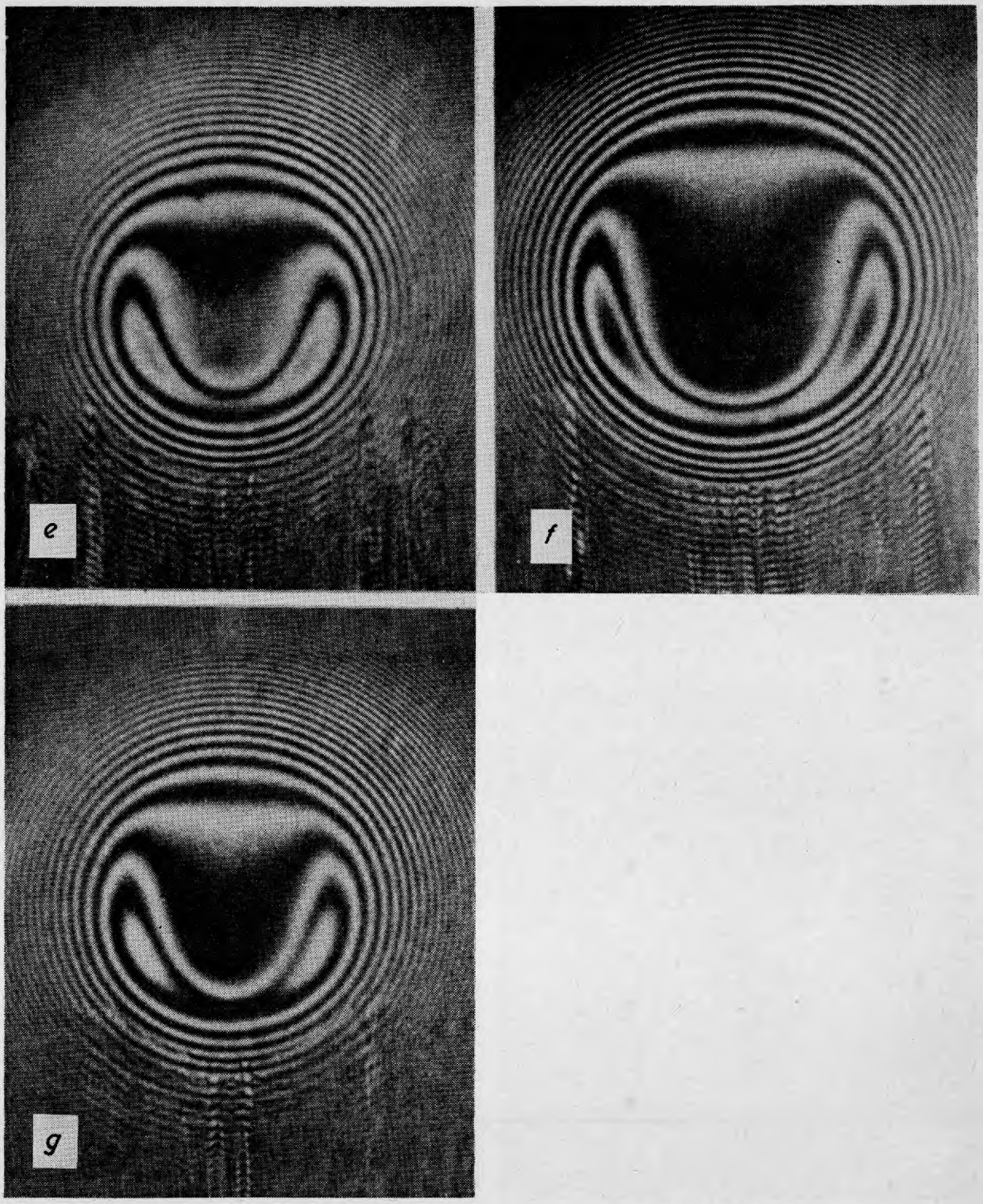


Fig. 6. Profiles in rolling



a 5 lb 3rd black.
b 10 lb 3rd black.
c 5 lb 4th black.
d 10 lb 4th black.

Fig. 7. Sliding contact, 5-phenyl-4-ether *(Contd overleaf)*



e 5 lb 5th black.
g 5 lb 6th black.

f 10 lb 5th black.

Fig. 7—contd

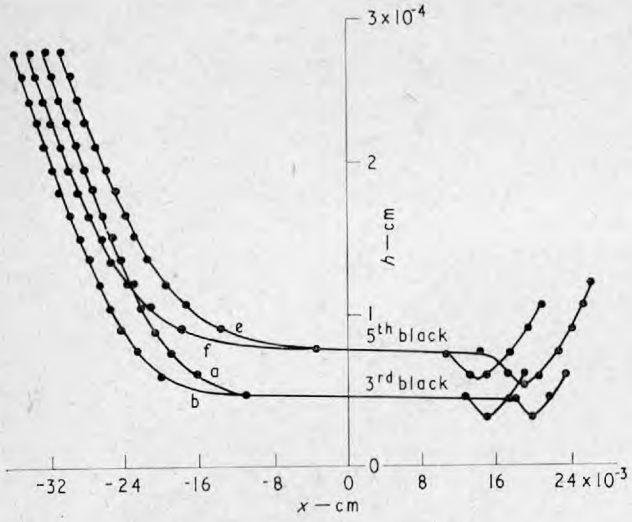


Fig. 8. Profiles in sliding

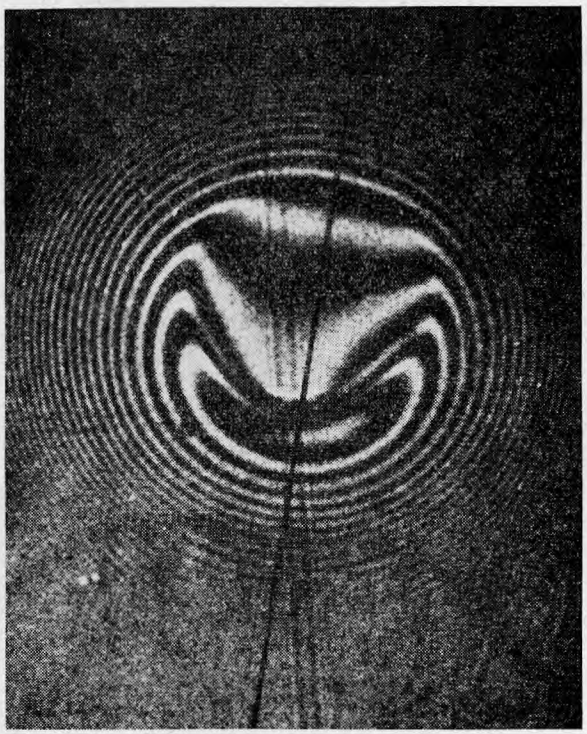


Fig. 9. Angled drive

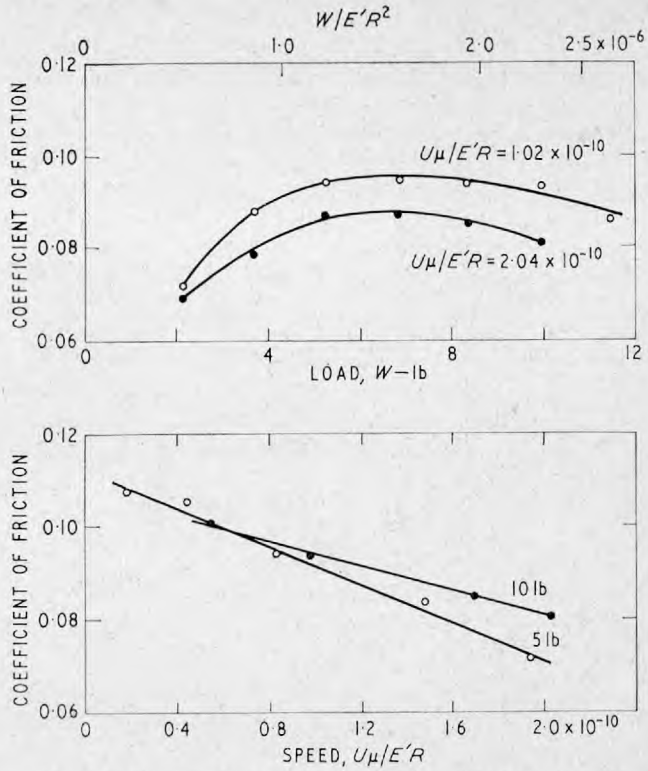


Fig. 10. Friction with 5-phenyl-4-ether, 100 per cent sliding

contacts can be found in standard works on ball bearing analysis such as reference (11).

Sliding tests in which the disc is held stationary against the rotating ball are given in Fig. 3. The thicknesses at the same dimensionless speed parameter are considerably lower than rolling values, being even lower for the 10 lb than the 5 lb load, and are almost certainly due to viscous heating in the inlet sweep. This heating made it much more difficult to obtain repeatability in sliding than in rolling. With the angled drive (16 per cent slide/roll) a film only slightly thinner than pure rolling is formed because the viscous heating effect is relatively small. Repeatability is easier to obtain than in 100 per cent sliding, partly because there is less heating but also because the rotation of the disc means that the contact is always fed entirely with fresh cool oil, whereas when the disc is stationary the contact receives a mixture of warm oil from the disc and fresh oil from the ball.

The photographs in Fig. 7 and profiles in Fig. 8 show the shape of the oil films in 100 per cent sliding. The main difference is that in sliding the entry to the Hertzian flat is more gently curved than the rolling case. The constrictions are almost identical in the two instances.

In pure rolling it is assumed that the individual deformations of the ball and disc are in the ratio of their reduced Young's moduli (i.e. glass being deformed 2·73 times as much as steel) but in sliding the surface shear stress, which must integrate up to give a traction of about 10 per cent of the normal load, acts in opposite directions on the two bodies, which means that each must be deformed differently.

Fig. 9 is a photograph of the 16 per cent slide/roll angled contact. The ruled black line shows the direction of the rolling vector. The inlet zone of the contact is symmetrical about this vector but the rear constriction is displaced to one side. This displacement is in the direction in which the sliding vector acts on the glass, and shows that glass is deformed more than the steel, which is as expected.

The coefficients of friction in 100 per cent sliding are plotted against speed and load in Fig. 10. The values of about 0·1 are higher than those usually expected for sliding point contact, but this is probably because of the very high viscosity and pressure–viscosity coefficient α_r of the 5-phenyl-4-ether lubricant. The film thickness in the centre is 15 μin, with no metallic contact.

CONCLUSIONS (PART 1)

A good theoretical method of predicting the film shape and thickness in rolling point contact is urgently needed, but in the meantime the optical interference technique provides a fairly simple and accurate method for measuring them, as shown in this paper. Knowledge of the stress distribution in point contact would be extremely valuable, especially as the measured shapes (Figs 4 and 6) suggest that the stress distribution at very high speeds may be more damaging than the more nearly Hertzian values at lower speeds. It is hoped that optical elastohydrodynamic contour maps will help solve this problem.

Part 2: Optical design

In order to get very fine fringes considerable attention to detail is necessary. The various factors are all small in themselves but in sum are important. The experience gathered over some 5 years' work is collected together in Part 2. Each element will be studied in turn.

ROLLING PAIR

Ball

In most work this is a conventional steel ball bearing ball, with a very highly finished surface, better than 1 μin c.l.a. Rougher surfaces give tolerable fringes, but for top fringe quality the smoother surface is needed. It has been found that very careful and prolonged lapping produces the required surface smoothness and reflectivity, which is about 75 per cent in air, which reduces to 60 per cent in oil. For very high loads a tungsten carbide ball is used.

No better method has yet been found for holding the ball than the p.t.f.e. ('Teflon') cup described earlier.

The whole test can be carried out with very little fluid; 5 ml is ample for determinations at several temperatures.

Transparent plate

The glass normally used is optical crown, free of striae and flat to one fringe per 5 in^2. The nominal Hertz stress glass can carry under lubricated conditions is something over 100 000 lbf/in^2. A sapphire plate can be used if higher stresses are needed.

In elastohydrodynamic systems, thermal properties of the surfaces are very important. Hence when studying oil film formation by optical systems, there must be thermal similarity between the transparent plate and the steel ball.

This is not very easy to achieve with glass and steel, but with sapphire and steel it is satisfactory, as Table 2 shows.

It may be noted that the thermal conductivity of sapphire decreases with temperature. The simple exponential relation shown here was obtained from data given by Union Carbide Corporation.

The conductivity of steel is only 20 per cent higher than sapphire at 0°C and the diffusivity of the two are equal at 37°C. This similarity is important.

Table 2. Thermal properties

Material	k Thermal conductivity, cal/(cm^2 s degC/cm)	c Specific heat	ρ Specific gravity	$R/c\rho$ Thermal diffusivity
Carbon steel	0·12	0·12	7·8	0·12
Crown glass	0·0024	0·20	2·4	0·005
Synthetic sapphire	0·1 exp (−4 $\times 10^{-3}\tau$) ($0 < \tau < 100$°C) τ°C	0·18	4·0	0·14 exp (−4 $\times 10^{-3}$°C)

DEGREE OF COLLIMATION

All fringes are of finite thickness so their position can be read to only a certain degree of accuracy. For two-beam interferometry this is about $\frac{1}{20}$ of the fringe width, while for multiple beam it is, say, $\frac{1}{60}$.

For all these techniques the illumination is by 'normal' incidence. But the beam is not perfectly collimated. Tolansky (reference (**12**), p. 160) considers the effects of such incomplete parallelism and shows that if a fringe accuracy δm can be tolerated, then $\delta m\lambda/t = \theta^2$ where θ is the $\frac{1}{2}$ cone angle of collimation. The permitted value of θ is seen to be proportional to $t^{-1/2}$ and Table 3 shows values for practical situations.

Table 3. Collimation angle

λ, Å	t, Å	δm	θ, degrees
4000	5000	1/20	11
		1/60	6½
	20 000	1/20	5¾
		1/20	3¼

In white light (coloured fringes), it has been shown (**13**) that it is possible to discriminate to 10 Å at 6000 Å, so the maximum value of θ is $2\frac{1}{2}°$. It is interesting that coloured fringes need more stringent collimation than monochromatic fringes.

COLLIMATORS

For normal illumination the incident and reflected rays are coincident. The incoming light beam will therefore have to go through a semi-reflecting mirror or beam splitting prism. It is preferable to put this below the object lens, firstly, to avoid spurious reflections (very serious with laser work) and, secondly, to avoid using the object lens as a collimating lens, as well as for its proper purpose. A long working distance object lens is therefore demanded in order to accommodate the prism.

A further advantage of a below-the-lens system is that it can be specially designed to give control both of intensity of the light and of the object area illuminated. A collimator designed for the Lubrication Laboratory by Professor H. H. Hopkins is shown in Fig. 11. The field stop is imaged by the field and aperture lenses, at almost unit magnification, on to the interference plane. This enables the area needing illumination to be decided exactly by controlling the field stop diameter. The aperture stop, which is placed at the focus of the aperture lens, controls the degree of collimation as well as the intensity.

In this design perfect collimation is not achieved. However, the angle of the cone of rays emerging is low enough to satisfy the most stringent requirements of colour interferometry as the pictures in reference (**3**) (which were obtained with this collimator) amply confirm.

The above-the-lens system described by Tolansky (**12**) is shown in Fig. 12. The source is imaged by a condenser lens on to the aperture stop. The aperture lens is arranged so that the image of the aperture stop is placed at the back focus, P, of the objective lens via a semi-reflecting mirror, M. The object lens therefore produces collimated light.

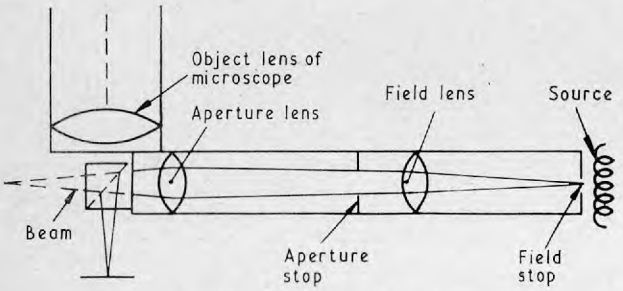


Fig. 11. Hopkins collimator

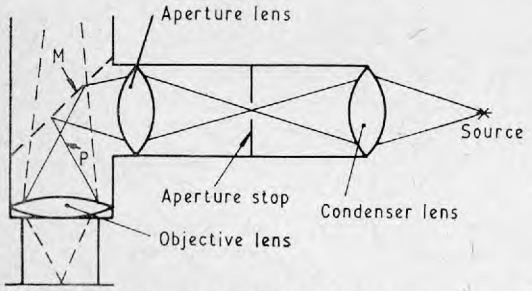


Fig. 12. Tolansky collimator

CHOICE OF REFLECTING LAYER

The reflectivity of the semi-reflecting layer used determines the intensity distribution of the fringes. If the reflectivity is low, rather broad 'two-beam' fringes are produced. However, much narrower 'multiple-beam' fringes can be obtained if the reflectivity is high. The amount of absorption in the semi-reflecting layer must be low for multiple-beam fringes in order to produce fringes of good visibility.

For lubrication studies a metal layer of chromium is often used. It is usually deposited on to the glass when hot (140°C) (**16**) and so gives a very abrasion resistant surface. Chromium also has the advantage that the oil 'sees' chromium, thus the surface lubricating properties are the same, or nearly so, as a metal.

The optimum reflectivity R, of the chrome layer is $\sim$ 20 per cent. This layer also has a $\sim$ 20 per cent absorption A of the incident light. The steel ball reflects 60 per cent. The two interfering beams therefore have the same intensity (18 per cent), as Fig. 13 shows. The second and third reflected beams are weak so effectively only two beams interfere. Generally, the intensity distribution between the light and dark fringes for two-beam interferometry is $\cos^2$ (Fig. 14), which results in rather

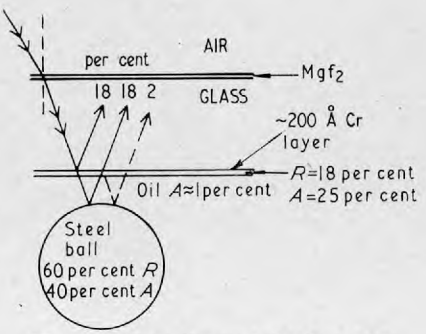


Fig. 13. Two-beam interference

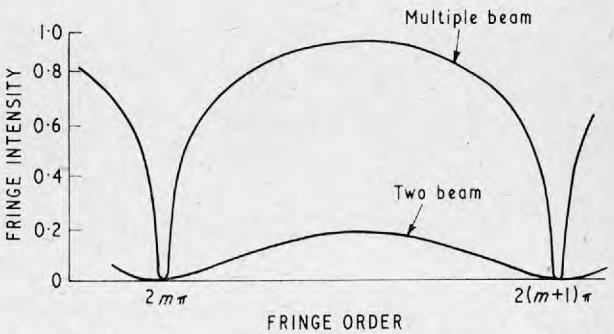


Fig. 14. Fringe intensities for two-beam and multiple-beam interferometry

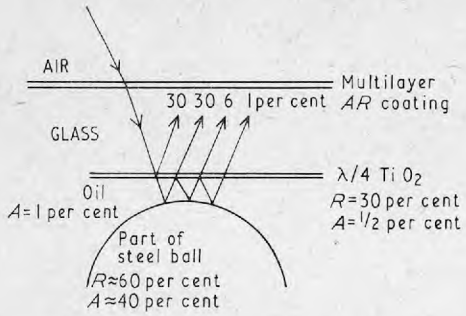


Fig. 15. Multiple-beam interference

broad fringes. A thin chromium layer, however, gives a modified intensity distribution which is discussed below.

If a dielectric, such as TiO_2, is used as the semi-reflecting layer, then the absorption is very low ($\sim\frac{1}{2}$ per cent) and a 30 per cent reflectivity can be achieved. The thickness of the layer must be a quarter wavelength of the light used. The reflected rays are shown in Fig. 15. With TiO_2 several reflections are achieved giving multiple beam fringes. The dark fringes are very much sharper, though of course the fringe separation is the same in the two-beam and multiple-beam modes. The intensity distribution is shown in Fig. 16.

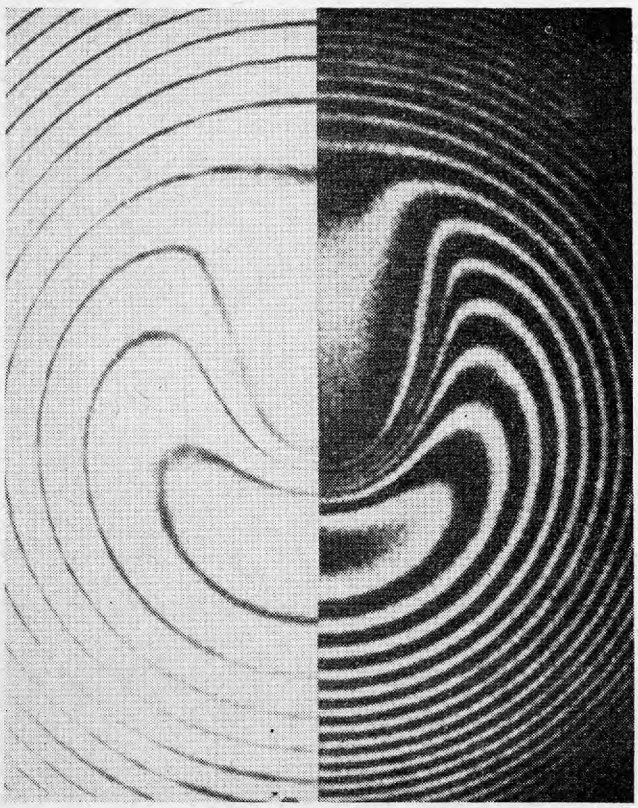


Fig. 16. Comparison of two-beam and multiple-beam interferographs

Multi-layer dielectric stacks have been used in place of a single TiO_2 layer to give higher reflectivities and considerable success has been achieved with stacks of up to four layers. However, above this number the layers break down under load.

The difference between two-beam and multiple-beam fringes can be appreciated from Fig. 16, which shows the film formed by a rolling ball. In both instances the load was 5 lb and the fluid 5-phenyl-4-ether. On the right-hand side the semi-reflecting layer was chrome and the source the 5461 Å mercury band (taken from half of Fig. 4*e*). On the left-hand side the layer was TiO_2 and light was from a helium–neon laser ($\lambda = 6328$ Å), giving a different fringe spacing.

LIGHT SOURCES

There are two broad divisions of light sources, single band (monochromatic) and continuous spectrum (chromatic). The former gives dark and light fringes, whereas the latter gives various coloured fringes.

A very common monochromatic light source is a mercury discharge lamp, with a filter passing the 5461 Å band. The new high pressure mercury lamps are very intense and compact. The arc is only $\frac{1}{4}\times\frac{1}{2}$ mm in extent. The band width of such high pressure lamps is widened considerably, up to some 200 Å. This tends to broaden the fringes and reduce the contrast at large values of t.

For precise work, especially with multiple beams, a helium–neon laser is almost ideal. The beam is already beautifully collimated, very intense and the red line at 6328 Å is only $\frac{1}{10}$ Å wide. For normal observation a power of 1–2 mW is adequate, while for high speed photographic studies at 5000 frames/second, 14 mW was found to give good exposures.

Laser light can provide clear fringes over large distances owing to its long coherence length and its good collimation. This is very valuable for measuring large film thicknesses, but it also gives interference patterns between the various glass–air interfaces in the optical system. All such interfaces must therefore be carefully coated with anti-reflection layers specially designed for the particular wavelength of the laser, to avoid as far as possible spurious patterns.

The most usual chromatic sources are white light incandescent lamps of the quartz–iodine type. Xenon discharge lamps also produce white light. With the continuous spectrum of white light, each wavelength independently interferes with itself; fringes are more closely spaced with the shorter than with the longer wavelengths. The fringes thus formed overlap one another, creating unsaturated spectral colours. The interference colours vary in sequence, thus enabling the fringe *order* to be determined without counting fringes. The colours are impure and the thickness corresponding to each must be found by calibration, which is discussed below. The thickness over which fringes are visible is limited. In particular, after about 8 or 10 fringes, with white light (wavelengths

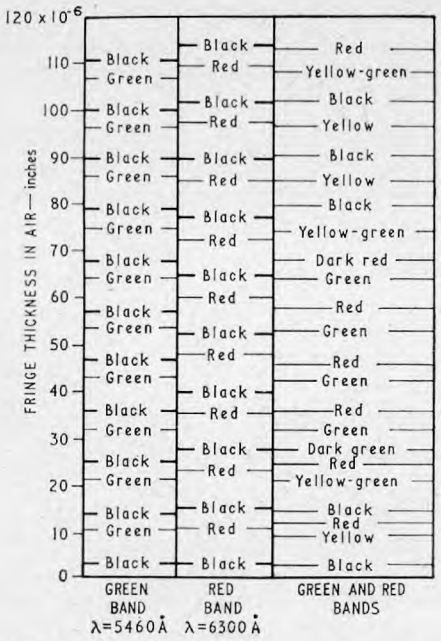


Fig. 17. Fringes of two separate wavebands, diagrammatic representation

between 4000 and 7500 Å) so many colours are present at one point that the resultant hue is practically white, even though interference is still taking place. It is for this reason that white light fringes can only be used for fluid thicknesses up to about 10^4 Å (40 μin).

Greater thicknesses can be measured by using a set of discrete wavelengths. The sequence of colours produced considerably simplifies the recognition of the fringe order. The mechanism by which this occurs is shown in Fig. 17. By superimposing various filters on a xenon discharge lamp, the individual green and red monochromatic systems were obtained. These are shown in the first two columns of Fig. 17. The green fringes are more closely spaced than the red fringes because of the shorter wavelength. When both wavebands are used simultaneously the chromatic fringes, as indicated in the third column, are obtained. The colour of the chromatic fringes is determined by the intensity of the green and red wavebands at each point. At zero thickness both green and red fringes are in phase relative to each other. At about 45 μin the green waveband has gained a half wavelength on the red. Hence, a light fringe of one falls on a dark fringe of the other. The corresponding chromatic fringes are therefore green and red. At about 90 μin the green band has gained a whole wavelength on the red. Thus, the light and dark fringes of each wavelength are again in phase. The corresponding chromatic fringes then become yellow and black.

The apparatus with which this system was used is described in references (**6**) and (**9**). A black and white print of such a two-waveband system is shown in Fig. 18. By using high speed flash photography the oil film rupture formation can clearly be seen. It also enables the film thickness to be evaluated as a function of a transient inlet lubricant boundary under lubricant starvation conditions. Preliminary results have been given in reference (**9**).

At the time of writing studies are being made using an argon laser which gives several discrete wavelengths between 4579 Å and 5145 Å. It would seem an ideal source. However, there are difficulties in designing suitable anti-reflection coatings and high reflecting stacks which will work efficiently for the spread of wavelength involved. The krypton laser will also give several discrete wavelengths with better spacing and providing the coatings can be designed it would seem to hold even greater promise.

CALIBRATION, PHASE CHANGE

Generally, when light is reflected from an interface of 'rare' to 'dense' medium a phase change of π occurs. For thin metal films such as chromium the phase change depends on the thickness of the metal film. Thus, a correction θ_{cr} must be added to the basic equation for interference fringes which is

$$m\lambda = 2nt \cos \theta + \phi_{cr} \quad . \quad . \quad . \quad (1)$$

Another result of this phase effect is that the intensity distribution is no longer $\cos^2$, but becomes asymmetric. This is shown in Fig. 17 for the monochromatic systems where the light fringes are not equally spaced between the

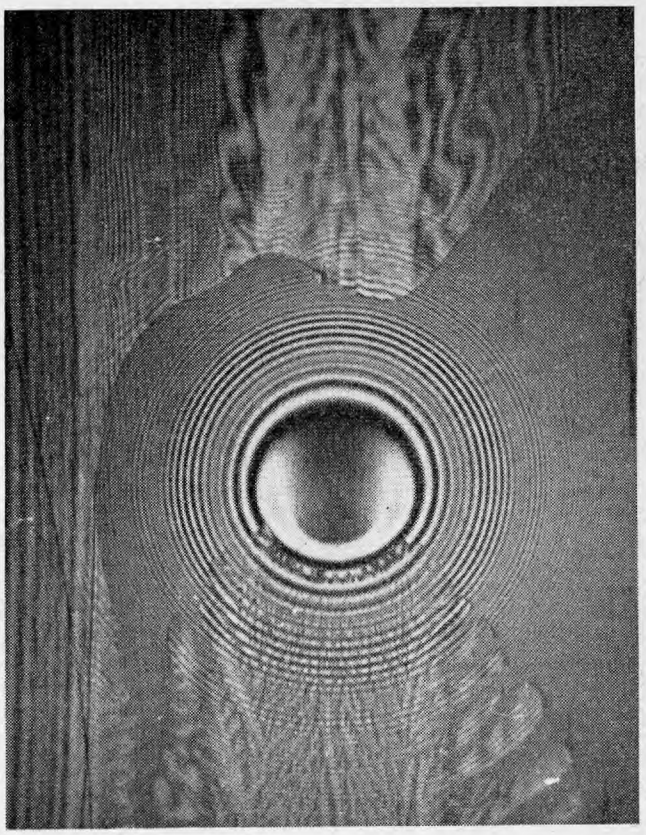


Fig. 18. High speed photograph of elastohydrodynamic contact using two-colour chromatic system

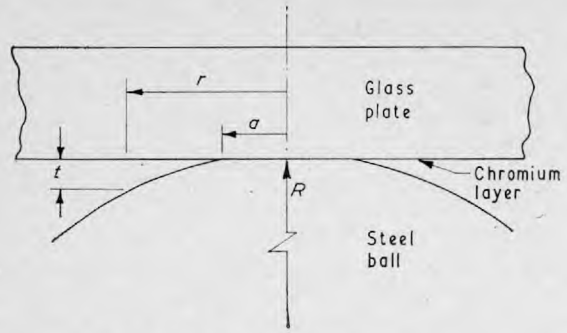


Fig. 19. Ball and plate contact

dark fringes. Thus, the absolute thickness t corresponding to each light fringe is not precisely known. The explanation of this effect is given by Holden (**15**).

The absolute thickness t corresponding to each fringe in a chromatic as well as a monochromatic system can be obtained by observing a system of known geometry, which most conveniently is a steel ball loaded against a glass plate. This system of calibration is used here and is shown in Fig. 19. The ball and the plate were the same as in the experimental machine. The fringes produced by this system of calibration consist of a series of concentric rings.

If an ideal condition of 'zero' load can be obtained, then the thickness t corresponding to each fringe can be calculated from the measured values of the fringe radius r by the simple relation

$$t = r^2/2R \quad . \quad . \quad . \quad . \quad . \quad (2)$$

This was attempted by measuring the diameters ($2r$) of the fringes for loads decreasing from 500 g towards zero. The fringe radius r was then plotted against the calculated Hertzian radius a. The first seven dark fringes of a monochromatic green ($\lambda = 5460$Å) system are shown in Fig. 20. As the load approaches zero, a should become a negligible proportion of the fringe radius r for each fringe. Thus, the curve of a versus r should approach the asymptote $r =$ constant. The experimental points deviate from this condition for loads below 10 g ($a = 1 \times 10^{-3}$ in). It is probable that this is due to the surfaces being held apart by microasperities which, at high loads, were squeezed flat. It was also noted that the fringe within the central Hertz area gradually changed from light grey to black as the load was reduced below 10 g, thus indicating that separation had taken place.

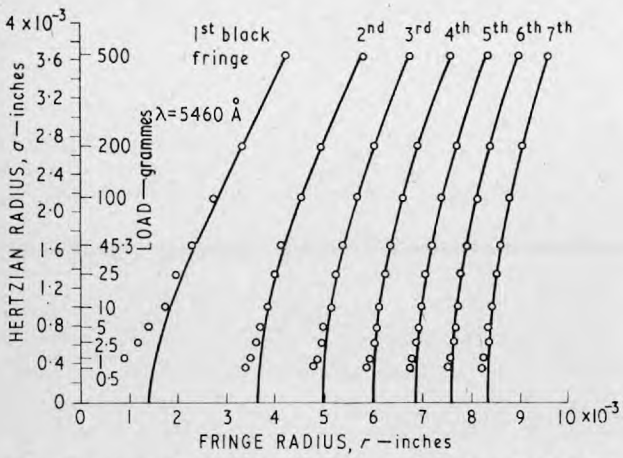


Fig. 20. Hertzian radius and load against fringe radius for the first seven dark fringes λ = 5460 Å

The theoretical shape of the curves a versus r can be determined from the Hertz equation for the gap outside the main contact region:

$$t = \left(\frac{1-\nu_1^2}{2E_1}+\frac{1-\nu_2^2}{2E_2}\right) a p_{\max} \left[-\left(2-\frac{r^2}{a^2}\right)\cos^{-1}\frac{a}{r}+\left(\frac{r^2}{a^2}-1\right)^{1/2}\right] \quad (3)$$

The measured values of r at 200 g were used as a reference. The theoretical curves are shown in Fig. 20.

The fringe radius r at zero load (i.e. $a = 0$) was obtained for each fringe by using the theoretical curves. The absolute thickness t corresponding to each fringe was then obtained from equation (2). The data for the chromatic and monochromatic systems of Fig. 17 was obtained in this manner.

It may be appreciated that the above method is fundamentally equivalent to using equation (3) directly with the measured values of r for a load of 200 g.

Knowing the absolute value of t, the phase change ϕ_{cr} can be obtained from equation (1). Using the monochromatic system of green light ($\lambda = 5460$ Å), ϕ_{cr} was found to be $-2\pi/5$. It is quite reasonable to assume that the phase change at the ball interface is π and that ϕ_{cr} is between 0 and π. Since the first fringe to occur is dark, ϕ_{cr} must therefore be negative.

CHANGE IN REFRACTIVE INDEX

The 'optical' thickness can be determined quite accurately in a lubricated contact. The 'actual' thickness, however, will depend on a knowledge of the refractive index of the medium. The refractive index changes about 10 per cent for most lubricants between atmospheric pressure and 100 000 lbf/in^2. The rate of change decreases for increasing pressures.

To determine the refractive index change a knowledge of the pressure is required. For thin elastohydrodynamic films the maximum Hertzian pressure is very similar to the pressure in the centre of the contact region. This pressure is usually taken as a reference.

The change in density with pressure can be determined by Hartung's empirical formula for hydrocarbons as mentioned by Chu and Cameron (**16**), and the change in refractive index with density can be evaluated from the Lorenz–Lorentz relation. Poulter *et al.* (**17**) found that the Lorenz–Lorentz relation was accurate to within 1 per cent for a paraffinic oil up to a pressure of 200 000 lbf/in^2.

Work is now being conducted to allow the refractive index in the contact region to be determined by an optical method. This will enable more accurate measurements of thickness to be obtained.

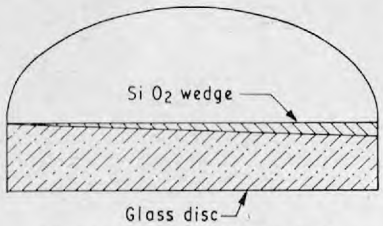


Fig. 21. Taper film of SiO_2

EXTENSION OF THE LOWER LIMIT

For the two-beam interferometry, when a phase change of exactly 0 or π occurs at each interface, the first fringe change (from light to dark or vice versa) occurs at $\lambda/4$ film thickness. This corresponds to a film thickness t of about 1000 Å. In lubrication much interest occurs with t below this value. In reference (5) a method was devised, the so-called 'optical spacer' method, which allows the absolute position of the first and all subsequent fringes to be moved. The basis of the technique is to replace the first few hundred ångströms of fluid with a controlled thickness of silicon dioxide. The SiO_2 has a refractive index close to that of most lubricants, so no perceptible partial reflection occurs at the SiO_2 to oil interface. To make the first dark fringe in multiple beam occur at a fluid separation of t then the required depth H of SiO_2 is given by

$$H = (\lambda/2n) - t \quad . \quad . \quad . \quad . \quad (4)$$

In reference (5) fluid thicknesses down to 100 Å were studied. An interesting advance is to deposit a variable thickness spacer using a source of SiO_2 displaced from the central axis of the glass disc. If the evaporation source is put below the edge of the disc, a wedge shaped layer is deposited (Fig. 21). In this way H can vary from nearly zero to some way above $(\lambda/2n)$. The value of H is determined by multiple-beam techniques and checked using ellipsometry (**18**). The refractive index of the layer is obtained by an immersion method. A motorized camera records 10–20 interferographs per revolution of the disc and by super-position, a very detailed fringe pattern is obtained.

CONCLUSIONS (PART 2)

By careful design very good interference fringes can be obtained in a lubricated contact. Good surface finish and matching reflectivities of the bearing surfaces are important for high quality interference fringes. Chromium and TiO_2 have been successfully used as semi-reflecting materials on the transparent bearing surface.

The narrow multiple-beam fringes enable very small changes in thickness to be determined, and by using the 'optical spacer' technique, very thin lubricant films can be measured.

Fringes of a chromatic two-colour source can simplify the recognition of fringe order considerably. The fringe system can be calibrated by observing the circular fringes produced by a ball loaded against a glass plate. The phase change on reflection associated with some thin metal reflecting layers can also be determined by this method.

ACKNOWLEDGEMENTS

The authors would like to thank the Science Research Council and Rolls-Royce Ltd, Derby, for grants to purchase some of the apparatus used here and for maintenance grants.

We would also like to thank Professor H. H. Hopkins for the design of the collimator, and the Hoffmann Bearing Co. of Chelmsford, for supplying the high quality steel balls used for Parts 1 and 2.

APPENDIX

REFERENCES

(1) CAMERON, A. and GOHAR, R. 'Theoretical and experimental studies of the oil film in lubricated point contact', *Proc. R. Soc. A.* 1966 **291**, 520.

(2) GOHAR, R. and CAMERON, A. 'The mapping of elastohydrodynamic contacts', *A.S.L.E./A.S.M.E. Lubric. Conf.* Minneapolis 1966, Preprint 66-LC-21.

(3) FOORD, C. A., HAMMANN, W. C. and CAMERON, A. 'Evaluation of lubricants using optical elastohydrodynamics', *A.S.L.E. Trans.* 1968, **11**, 31.

(4) DOWSON, D. and JONES, D. A. 'An optical-interference method of measurement of time-dependent elastohydrodynamic film profiles', Rep. 3 *Symp. Exptl Methods in Tribology, Proc. Instn mech. Engrs.* 1967–68 **182** (Pt 3G), 49.

(5) WESTLAKE, F. J. and CAMERON, A. 'A study of ultra-thin lubricant films using an optical technique', Rep. 12 *ibid.*, 75.

(6) WEDEVEN, L. D. and CAMERON, A. 'A study of elastohydrodynamic lubrication in rolling bearings using optical interference', Rep. 15 *ibid.*, 87.

(7) ROBERTS, A. D. and TABOR, D. Short communication 'Fluid film lubrication of rubber—an interferometric study', *Wear* 1968 **11**, 163.

(8) CAMERON, A. 'Hydrodynamic theory in gear lubrication', *J. Inst. Petrol.* 1952 **38**, 614.

(9) WEDEVEN, L. D. and CAMERON, A. 'The observation of elastohydrodynamic lubrication in a rolling element bearing using optical interferometry,' *10th Int. Symp. on Lubrication of Rolling Bearings* 1968, Halle, Germany (to be published in *Schmierungstechnik*).

(10) ARCHARD, J. F. and KIRK, M. T. 'Lubrication at point contacts', *Proc. R. Soc. A.* 1961 **261**, 532; also 'Film thicknesses for a range of lubricants under severe stress', *J. mech. Engng Sci.* 1964 **6** (1), 101.

(11) HARRIS, T. A. *Rolling bearing analysis* 1966 (Wiley, New York and London).

(12) TOLANSKY, S. *An introduction to interferometry* 1955 (Longmans, London).

(13) WRIGHT, W. D. *Research in normal and defective vision* 1946 (Henry Kimpton, London).

(14) THETFORD, A. University of Reading, private communication, 1967 (July).

(15) HOLDEN, J. 'Multiple-beam interferometry: intensity distribution in the reflected system', *Proc. phys. Soc.* 1949 **62**, 405.

(16) CHU, P. S. Y. and CAMERON, A. 'Compressibility and thermal expansion of oils', *J. Inst. Petrol.* 1963 **49**, 140.

(17) POULTER, T. C., RITCHEY, C. and BENZ, C. A. 'The effect of pressure on the index of refraction of paraffin oil and glycerine', *Phys. Rev.* 1932 **41** (Aug. 1), 366.

(18) *Ellipsometry: Proc. Symp. Ellipsometry and its use in the measurement of surfaces and thin films*, Nat. Bur. Standards, Wash., Miscellaneous Publ. 256, 1963.

MADE AND PRINTED IN GREAT BRITAIN BY WILLIAM CLOWES AND SONS, LIMITED, LONDON AND BECCLES