Rolling Contact Fatigue

George Plint
MA, CEng, FIMechE
Tribology Trust Silver Medal 2017

Phoenix Tribology Ltd

info@phoenix-tribology.com

PLINT Tribology Products
from Phoenix Tribology Ltd

Introduction

The background to this guidance note was a request from a client to attempt to explain why tests on nominally identical test bearings, in two different test machines, gave radically different bearing lives. In this case, the bearing in question was a 7206 angular contact ball bearing and tests were run at nominally identical test loads. Other test parameters were not disclosed.

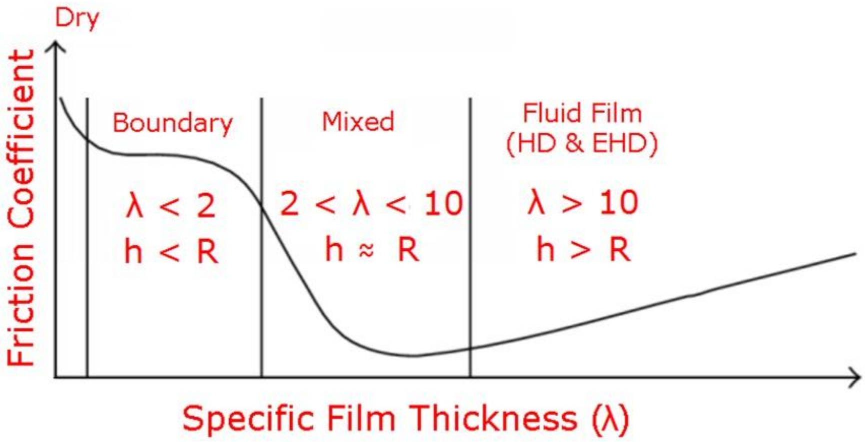
The results showed two orders of magnitude variation in fatigue life, between the two test set-ups. We concluded that this must have been caused by significant differences in test bearing lubrication, giving rise to significant variations in contact morphology.

My aim here is not to give a lengthy discourse of rolling contact fatigue but to give some basic guidance, possible benchmark calculations and assorted caveats, to assist those interested with experimental design. Having covered some of the basics, the presentation will conclude with a review of different test geometries and how these should best be used to greatest advantage.

Rolling Contact Fatigue

Rolling contact fatigue involves the initiation and propagation of cracks, which can lead to surface pitting, spalling, and ultimately failure of the bearing surfaces. While load is a primary factor in fatigue life, lubricant, lubrication regime and material properties, all influence rolling contact fatigue.

Rolling Contact Fatigue



PLINT Tribology Products
from Phoenix Tribology Ltd

Rolling contact fatigue can occur through both surface and subsurface mechanisms. Surface-initiated rolling contact fatigue, usually referred to as micro-pitting, is associated with asperity stress interactions, hence rolling contacts, of sufficient surface roughness, operating under mixed lubrication.

Micro-pitting

Surface-initiated rolling contact fatigue:

- asperity stress interactions under mixed lubrication
- influenced by contact pressure, material hardness, surface roughness and lubricant film thickness
- occurs when ratios (film thickness/surface roughness) of less than are generated
- may occur at between 10^5 and 10^6 fatigue cycles

PLINT Tribology Products from Phoenix Tribology Ltd

Micro-pitting is influenced by contact pressure, material hardness, surface roughness and lubricant film thickness. It can occur when λ ratios (film thickness/surface roughness) of less than 5 are generated. Micro-pitting may occur at between 10^5 and 10^6 fatigue cycles.

Macro-pitting

Surface-initiated rolling contact fatigue:

- occurs as a result of high cycle, sub-surface stresses, generated during rolling motion
- cracks originate beneath surface of material, often due to plastic deformation, or at stress raisers such as material flaws or inclusions
- usually occurs under conditions of fully developed elasto-hydrodynamic lubrication
- may occur at between 10^9 and 10^{10} fatigue cycles

PLINT Tribology Products
from Phoenix Tribology Ltd

Subsurface initiated rolling contact fatigue, otherwise termed macro-pitting, occurs as a result of high cycle, sub-surface stresses, generated during rolling motion. Cracks originate beneath the surface of the material, often due to plastic deformation, or at stress raisers such as material flaws or inclusions.

Subsurface originated macro-pitting usually occurs under conditions of fully developed elasto-hydrodynamic lubrication and at between 10^9 to 10^{10} fatigue cycles. Higher applied loads reduce rolling contact fatigue life, because of increased stress amplitudes, giving rise to more rapid fatigue damage and crack initiation and propagation.

Because of the difference in the scale of damage caused by macro-pitting, compared with micro-pitting, it invariably marks the end of component life, whereas micro-pitting may just result in a reduction in overall bearing life.

Micro-pitting as pre-cursor for early onset Macro-pitting

Surface fatigue in lubricated contacts: Mapping the failure modes of micro-pitting versus macro-pitting - B. Wainwright and A. Kadiric, Tribology Group, Imperial College London - International Journal of Fatigue Volume 197, 2025

"The key observation here is that macro-pitting exclusively occurred at higher pressures (2 GPa and 2.5 GPa); the lowest test pressure of 1.5 GPa never resulted in macro-pitting, despite significant micro-pitting taking place. This strongly indicates that the risk of micro-pitting transitioning to a more critical failure mode of macro-pitting, is primarily driven by the level of the applied contact pressure, with high pressures making the eventual occurrence of macro-pitting inevitable."

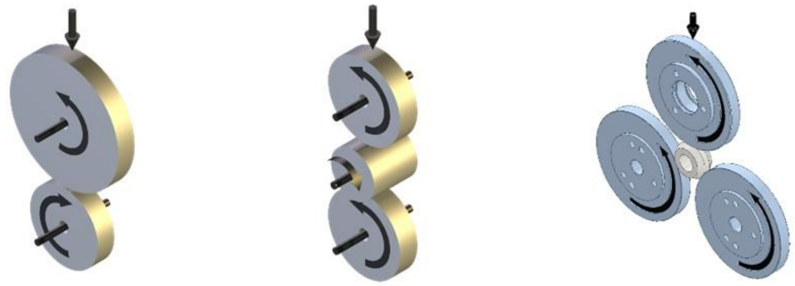
PLINT Tribology Products
from Phoenix Tribology Ltd

It will be apparent that, as a rough rule of thumb, tests that end with macro-pitting, typically last about 100 times the number of cycles as tests that produce micro-pitting. There is, however, one significant caveat to this: micro-pitting may act as a pre-cursor and accelerator for early onset macro-pitting. To quote:

> Surface fatigue in lubricated contacts: Mapping the failure modes of micro-pitting versus macro-pitting - B. Wainwright and A. Kadiric, Tribology Group, Imperial College London - International Journal of Fatigue Volume 197, 2025
>
> "The key observation here is that macro-pitting exclusively occurred at higher pressures (2 GPa and 2.5 GPa); the lowest test pressure of 1.5 GPa never resulted in macro-pitting, despite significant micro-pitting taking place. This strongly indicates that the risk of micro-pitting transitioning to a more critical failure mode of macro-pitting, is primarily driven by the level of the applied contact pressure, with high pressures making the eventual occurrence of macro-pitting inevitable."

Test Geometries – Bearings versus Roller Test Machines

Idealised Test Models



PLINT Tribology Products
from Phoenix Tribology Ltd

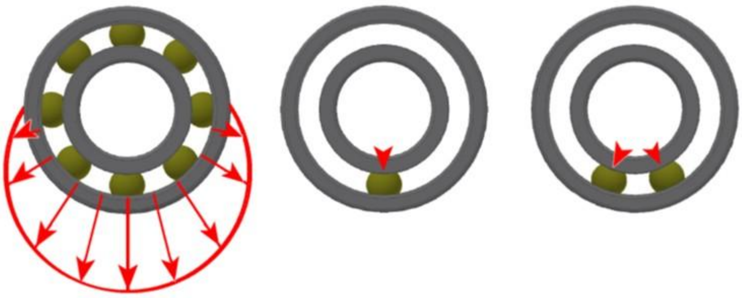
Idealised test models based on two, three or four roller type test geometries are amenable to analysis of hertz pressure and lubricant film thickness, plus provide no difficulty when it comes to defining and measuring the number of fatigue cycles, with or without contact slip. Hence defining a test in terms of contact pressure and rotational speed, with the result the number of fatigue cycles to failure, is readily achievable.

When it comes to rolling element bearings, everything becomes much more difficult, and, in the case of axially loaded angular contact bearings, spectacularly so. As a result of this, and going back to the early days of Palmgren and Lundberg, the tendency has been not to define bearing test conditions and resulting bearing life in terms of contact pressure and number of fatigue cycles, but to define the test conditions as applied load on the bearing and rotational speed, with L10 life simply the number of hours run, at that load and speed, until failure occurs.

So, there is something of a difference between how we define test conditions and fatigue life in our idealised and fully deterministic two, three or four roller machines and how we do it when it comes to bearings.

Test Geometries – Bearings versus Roller Test Machines

Radially Loaded Bearings



PLINT Tribology Products
from Phoenix Tribology Ltd

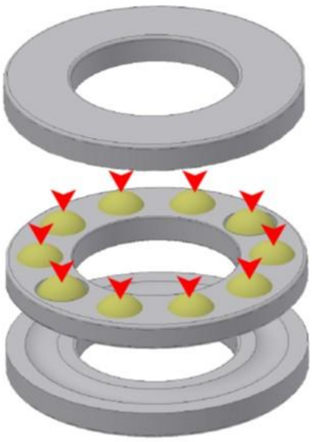
Radial load is a force acting perpendicularly to the axis of a rotating. Load is transferred from the inner raceway, through the rolling elements, to the outer raceway, with force distributed over multiple rolling elements simultaneously.

The load, hence contact pressure, on a given ball, depends on its rotational position within the bearing. A rough estimate of the maximum contact pressure experienced by a single ball can be made by assuming that at some point in the cycle, the total load is carried by just one ball. This will result in a contact pressure that will exceed that likely to occur in practice.

It will be apparent that with multiple rolling elements supporting the load on a radially loaded bearing, that the load on each bearing element, hence the contact pressure, is statically indeterminate.

Test Geometries – Bearings versus Roller Test Machines

Axially Loaded Bearings



PLINT Tribology Products
from Phoenix Tribology Ltd

In an axially loaded bearing it is usually assumed that the load applied to the bearing is shared equally by all the rolling elements and does not vary with rotational position. The assumption of uniform load distribution facilities meaningful analysis of contact pressure.

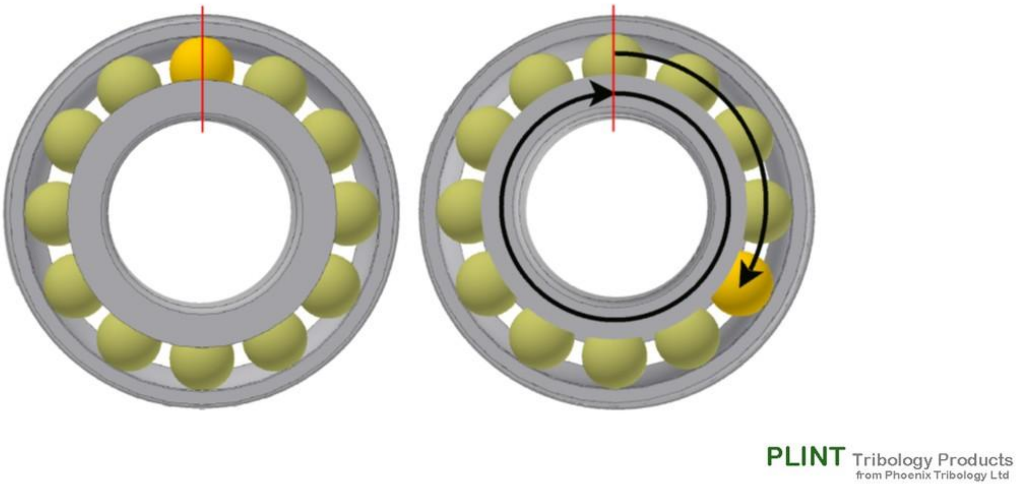
Reduced Compliments and Modified Cages

If a test is to be based on a standard bearing, it makes sense to choose bearings with an even number of rolling elements, for example, a 7209 angular contact ball bearing usually has twelve balls. This means that this bearing can either be used for tests with the full compliment of balls, or with the number reduced to six, four or three evenly spaced balls.

In the case of a thrust bearing such as 51208, the number of balls is typically thirteen, making it impossible to reduce the number of balls while maintaining uniform distribution. In this case, the simplest solution is to make a modified twelve ball cage, allowing tests to be run with six, four of three balls.

Motion & Contact Pressure - Axially Loaded Angular Contact Bearing

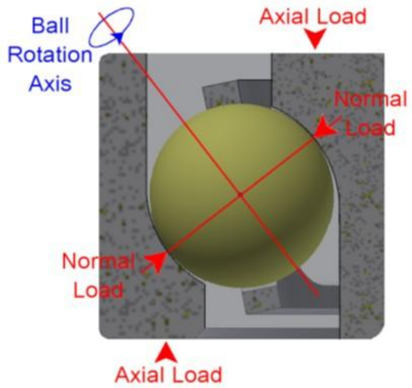
Fatigue Cycles per Bearing Rotation



The first thing to note from this illustration is that the speed of the inner race relative to the balls and ball separator (not shown) is greater than that of the outer race relative to the separator. It follows that a given point on the inner race will be subjected to a greater number of stress cycles per revolution than the outer race.

Motion & Contact Pressure - Axially Loaded Angular Contact Bearing

Bearing Contact Angle



PLINT Tribology Products
from Phoenix Tribology Ltd

Assuming:

1. that the speed of rotation and/or the mass of the balls is such that the balls do not generate significant centrifugal force,
2. that the inner and outer race contact angles are approximately equal,
3. that the balls are in pure rolling, rotating about an axis perpendicular to and in the plane of loading,

we could assume that normal loads on each ball are equal and opposite. If we know every detail of the contact geometry, we should be able to estimate the contact pressure on both contacts, noting that although the contact loads may be the same, the resulting contact pressures will be different, because of the differences in contact geometry between inner and outer race.

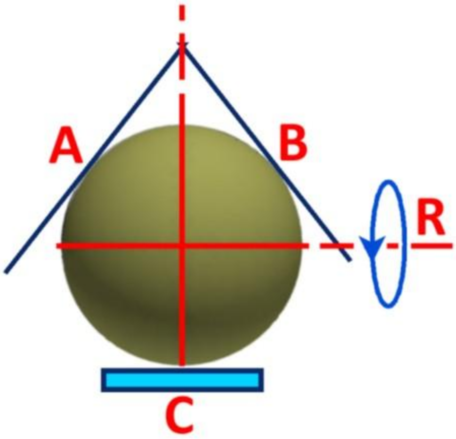
However, although Assumption (1) may be true at modest speeds, Assumptions (2) and (3) require further consideration.

Bearing Contact Angle

It should be clear that with an axially loaded angular contact bearing, the static geometry is complicated but becomes even more complicated with the introduction of rotational motion, deflection through application of load, and thermal expansion, all of which change the bearing contact angles, thus changing the contact pressures.

Motion & Contact Pressure - Axially Loaded Angular Contact Bearing

Ball Spin



PLINT Tribology Products
from Phoenix Tribology Ltd

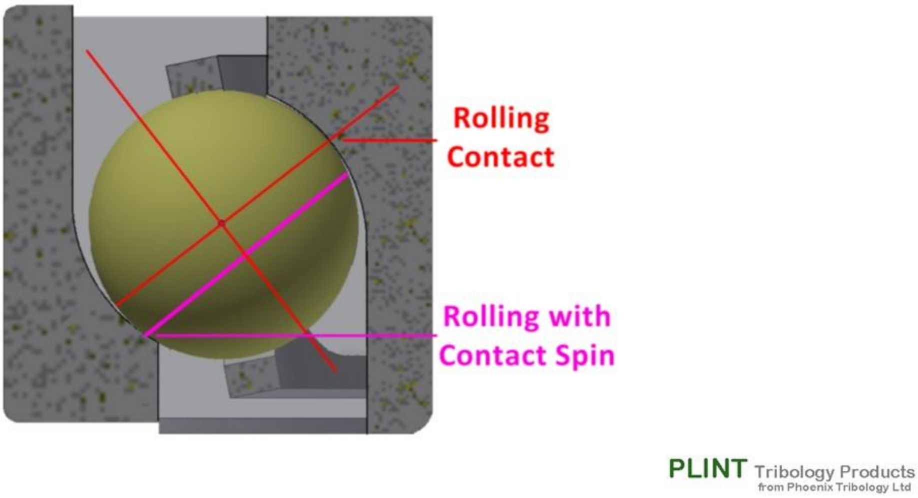
As well as the rolling component of velocity, the balls are subject to contact spin, which provides the largest component of rolling resistance, hence the largest contribution to frictional losses and heat generation.

If the ball illustrated is running in a straight line on a flat surface, in other words, rotating about the axis R, then contact C will have pure rolling but points of contact at A and B will be subject to contact spin. In this case, failure will occur at either contact A or B before C, hence as a test assembly, this will only work as a device for testing rolling contact fatigue in a contact with spin, not pure rolling.

If the ball is running around a circular track instead of in a straight line, there will be spin in all three contacts. Spin is defined as rotation within the hertzian contact. This is what develops traction forces within a pure rolling contact. However, because the traction developed by the spin in all three contacts must be balanced somehow, there is bound also to be slip in each contact.

Motion & Contact Pressure - Axially Loaded Angular Contact Bearing

Ball Spin



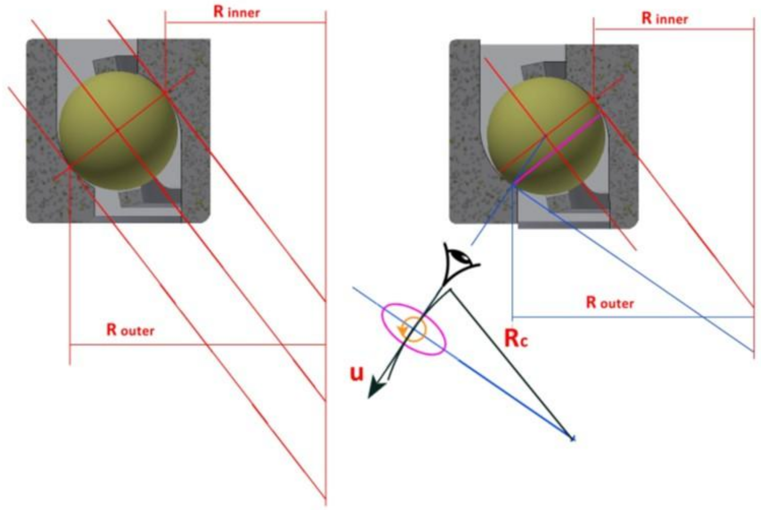
In an angular contact ball bearing because the ball rolling axis is inclined to the direction of the bearing axis there must be spin at one or both of these contacts. It is possible for rolling with spin to occur at one race contact and rolling without spin at the other.

Except at very high speeds, the tighter curvature of the inner race results in a higher spin moment than at the outer races, resulting in pure rolling at the inner race and rolling with spin at the outer race. This is termed "inner race ball control".

At very high speeds, centrifugal force contributes to an increase in load on the outer race and a decrease in load on the inner race, with a result that "ball control" may transfer from the inner to the outer race.

Motion & Contact Pressure - Axially Loaded Angular Contact Bearing

Ball Spin



PLINT Tribology Products
from Phoenix Tribology Ltd

This illustration shows, on the left, an idealised contact with the tangents at the ball race contacts and the axis of rotation of the ball all parallel. On the right-hand side, the illustration shows what happens if there is, for whatever reason, deflection in the bearing. For simplicity, I have only shown the point of contact moving on the outer race, however, in practice, R_{inner} would increase as R_{outer} decreases. What actually happens will depend on frictional conditions, lubrication, distribution of loads, both applied, centrifugal and gyroscopic etc.

Lubrication

Rolling element bearings:

- typically operate within a piezo-viscous-elastic regime
- lubricant film thickness plays significant role is determining rolling contact fatigue outcomes
- thicker lubricant films prevent asperity contact and distribute load more evenly across the contact
- thinner lubricant films lead to increased asperity contact leading to increased friction, localised stress concentrations, wear and, potentially, micro-pitting

PLINT Tribology Products from Phoenix Tribology Ltd

Lubricant Film Thickness

Rolling element bearings typically operate within a piezo-viscous-elastic regime. It will be apparent that lubricant film thickness plays a significant role is determining rolling contact fatigue outcomes. Thicker lubricant films prevent asperity contact and distribute load more evenly across the contact, minimizing stress concentrations that might otherwise initiate fatigue cracks. Thinner lubricant films lead to increased asperity contact, leading to increased friction, localised stress concentrations, wear and, potentially, micro-pitting.

A well-lubricated contact, with a sufficient lubricant film, generally leads to a longer fatigue life. This might suggest a preference for using higher viscosity lubricants, however, at higher speeds, this may result in increased lubricant shear and the bearing running hot, with a reduction in lubricant viscosity and hence film thickness.

Satisfactory lubrication requires sufficient lubricant to be delivered to the in-running sides of the rolling contacts to ensure that inlets are fully flooded. If this is not possible, the bearing may run starved of lubrication, resulting in premature failure, possibly through seizure.

In a rolling element bearing, pressures within the hertzian contact are sufficient to cause elastic deformation of the rolling elements. The lubricant (provided it has been properly entrained and carried into the contact) is subjected to these high pressures and this causes a massive increase in its effective viscosity. This is the pressure-viscosity effect and it can cause the lubricant's effective viscosity to approach that of glass. The fact that the lubricant is also effectively incompressible means that within the EHD contact the film thickness is close to being constant.

So, under EHD, the solid elements of the bearing are separated by an extremely stiff, glass-like, film of lubricant. This prevents the surfaces coming into contact and wearing or micro-pitting but obviously does not prevent the transmission of load across the contact, thus not preventing pressure generated rolling contact fatigue and eventual macro-pitting failure. It is useful to note that the thickness of the elasto-hydrodynamic lubricant film is pretty insensitive to applied load, implying that the harder it is squeezed, the stiffer it gets.

Electrohydrodynamic Lubrication

Assuming that adequate lubrication can be made available in our test rig, it makes sense to make an estimate of the lubricant film thickness, hence the λ ratio, for the proposed test. This will potentially allow the test to be designed to produce either micro-pitting or macro-pitting.

Modulus of elasticity	E	Nm^{-2}
Effective elastic modulus	E'	Nm^{-2}
Normal load applied	W	N
Central film thickness	h	m
Reduced radius	R	m
Entrainment velocity	u	ms^{-1}
Pressure viscosity coefficient of fluid	α	m^2N^{-1}
Viscosity at atmospheric pressure & test temperature	η_0	Pas
Poisson's ratio	ν	

Where R is the reduced radius and E' is the reduced modulus of elasticity.

The reduced radius R of two ellipsoidal bodies in contact is defined by the equation:

$$\frac{1}{R} = \frac{1}{R_x} + \frac{1}{R_y}$$

Where R_x and R_y are the effective radii in two directions at right angles.

If r_{ax} and r_{bx} are the respective radii of the two bodies a and b in the x direction:

$$\frac{1}{R_x} = \frac{1}{r_{ax}} + \frac{1}{r_{bx}}$$

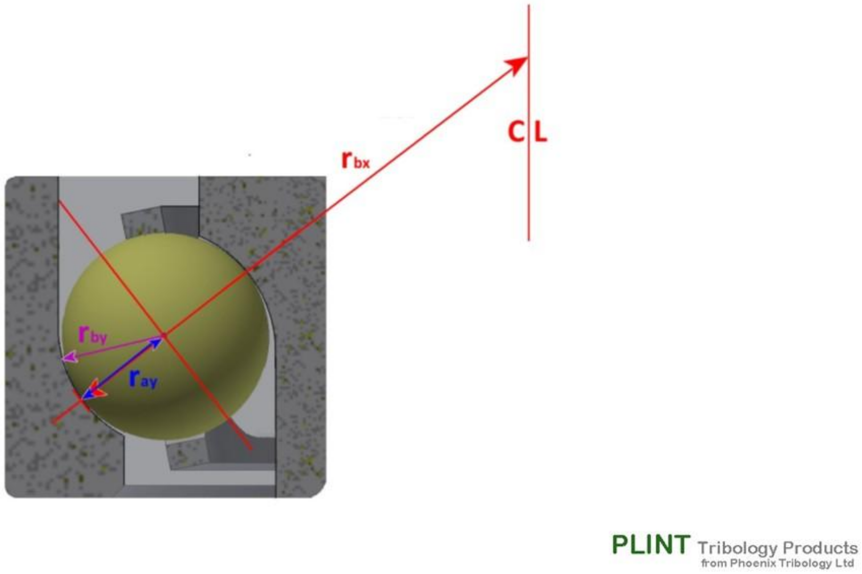
Similarly, if r_{ay} and r_{by} are the respective radii in the y direction:

$$\frac{1}{R_y} = \frac{1}{r_{ay}} + \frac{1}{r_{by}}$$

Then:

$$\frac{1}{R} = \frac{1}{r_{ax}} + \frac{1}{r_{bx}} + \frac{1}{r_{ay}} + \frac{1}{r_{by}}$$

Lubricant Film Thickness



For an angular contact bearing:

$$r_{ax} = r_{ay} = radius\ of\ ball$$

$$r_{bx} = outer\ race\ running\ radius$$

$$r_{by} = outer\ race\ curvature\ radius$$

Note: in this case, r_{bx} and r_{by} will be negative as the surface is concave, not convex.

The effective elastic modulus E' for EHL is defined by:

$$\frac{1}{E'} = \frac{1}{2}[\frac{1-v_d^2}{E_d} + \frac{1-v_b^2}{E_b}]$$

Note: a potential source of confusion is the difference between E^*, the contact modulus, and E', the effective elastic modulus. In calculating contact pressure using the Hertzian contact equations, we use E^*, whereas in the various EHL equations, we use E'. E^* is defined as:

$$\frac{1}{E^*} \equiv \frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2}$$

The pressure viscosity coefficient is defined as:

$$\eta_p = \eta_0 e^{\alpha p}$$

The film thickness equation, based on dimension analysis, is:

$$\frac{h}{R} = k\left[\frac{\eta_0 u}{E'R}\right]^a \left[\frac{W}{E'R^2}\right]^b [\alpha E']^c$$

Note: the entrainment velocity u will be entrainment velocity between ball and outer race.

Numerous different values for the constant k and the indices a, b and c have been proposed, including by:

- Hamrock & Dowson
- Westlake and Cameron (corrected by Jackson)
- Ranger and Cameron
- Foord and Cameron

For Hamrock & Dowson:

$$k = 1.900$$

$$a = 0.670$$

$$b = -0.067$$

$$c = 0.530$$

$$\frac{h}{R} = 1.900\left[\frac{\eta_0 u}{E'R}\right]^{0.670}\left[\frac{W}{E'R^2}\right]^{-0.067}[\alpha E']^{0.530}$$

Regardless of the equation we decide to use, there is one significant caveat and that is that the pressure-viscosity coefficient α, as defined, is assumed to remain constant, regardless of the shear rate or temperature. In practice, this can only apply to Newtonian fluids at ambient temperature. However, we can console ourselves with the thought that we only wish to *estimate* film thickness for the purposes of our experiment.

Lubricant Film Thickness

Ball			**EHD Equation Parameters**	Hamrock & Dowson	
Rax	0.005	m	k	1.900	
Ray	0.005	m	a	0.670	
E1	2.10E+11	N m^-2	b	-0.067	
Sigma1	0.30		c	0.530	
Ball Surface Speed	4.00	ms^-1			
			Preliminary Calculations		
Outer Race			R (reduced radius)	5.1724E-03	m
Rbx	-0.025	m	E' (effective elastic modulus)	2.31E+11	N m^-2
Rby	-0.006	m	U (entrainment velocity)	2.000	m s^-1
E2	2.10E+11	N m^-2	Dynamic Viscosity @ 40C	0.0602	Pas
Sigma2	0.30		Dynamic Viscosity @ 100C	0.0077	Pas
Race Surface Speed	0	ms^-1			
Lubricant			**Film Thickness Calculations**		
Viscosity @40C	68.00	centiStokes	Velocity-Viscosity Parameter	2.01E-07	
Density @40C	886	kg m^-3	Load Parameter	2.09E+00	
Viscosity @100C	8.70	centiStokes	Pressure-Viscosity Parameter	1.05E+02	
Density @100C	886	kg m^-3			
Pressure Viscosity Coefficient	2.80E-08	Pa^-1			
Test Conditions					
Normal Load on Ball	100	N	Film Thickness	0.432	Microns
Dynamic Viscosity	0.0602	Pas			

PLINT Tribology Products
from Phoenix Tribology Ltd

There are numerous on-line calculators available, alternatively, you may decide to make your own.

Lubricant Film Thickness

lubricants

MDPI

Review

Elastohydrodynamic Lubrication

James A. Greenwood

Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK; jag@eng.cam.ac.uk

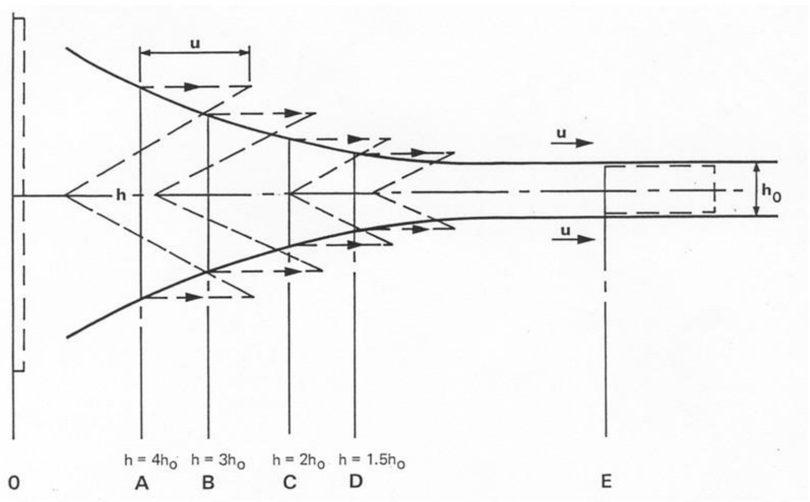
Received: 30 December 2019; Accepted: 26 April 2020; Published: 6 May 2020

Abstract: The development of EHL theory from its tentative beginnings is outlined, with an account of how Ertel explained its relation to Hertz contact theory. The problems caused by the failure of the early numerical analysts to understand that the film thickness depends on only two variables are emphasised, and answers of the form $H = F(P, S)$ given. Early methods of measuring the film thickness are described, but these became archaic with the development of optical EHL. The behaviour of surface roughness as it passes through the high pressure region and suffers elastic deformation is described, and the implication for the traditional Λ-ratio noted. In contrast, the understanding of traction is far from satisfactory. The oil in the high pressure region must become non-Newtonian: the early explanation that the viscosity reduction is the effect of temperature proved inadequate. There must be some form of shear thinning (perhaps according to the Eyring theory), but also a limiting shear stress under which the lubricant shears as an elastic solid. It seems that detailed, and difficult, measurements of the high pressure, high shear-rate behaviour of individual oils are needed before traction curves can be predicted.

PLINT Tribology Products from Phoenix Tribology Ltd

For a thoroughly comprehensive summary of the history the development of EHD film thickness equations, I would strongly recommend reading Jim Greenwood's paper from 2020, which is available online. That is the same Jim Greenwood as co-authored the important Green-Williamson paper: "Contact of nominally flat surfaces" (1966).

Lubricant Shear



PLINT Tribology Products
from Phoenix Tribology Ltd

This figure shows the velocity profile in a lubricant approaching an elasto-hydrodynamic rolling contact (of the type generated in a two-roller machine contact). The fluid in contact with the moving surfaces is assumed to travel with the surface at velocity u. Considerations of continuity suggest successive velocity profiles as the fluid approaches the contact of the form sketched.

The film thickness h in an elasto-hydrodynamic contact is largely determined by processes taking place in the approach or inlet zone to the contact zone proper. In this region lubricant entering the contact has been subjected to intense shear, if at only moderate pressure, then, on its passage into the contact, further shear, at a lower rate, but increasing pressure.

If the response of the lubricant to this combination of shear and pressure is non-Newtonian, that is if shear stress ceases to be proportional to shear rate, it may be expected that its behaviour in the approach region will differ from that of a Newtonian fluid of the same nominal characteristics.

It is this act of shearing that gives rise to loss of apparent viscosity in polymer thickened oils and limits the usefulness of viscosity index improvers. It is perhaps no surprise that we have a lubricant shear stability test (KRL) which involves running a loaded taper roller bearing under load, immersed in a small volume of test oil, for a number of hours; this is followed by measurement of the kinematic viscosity of the lubricant sample.

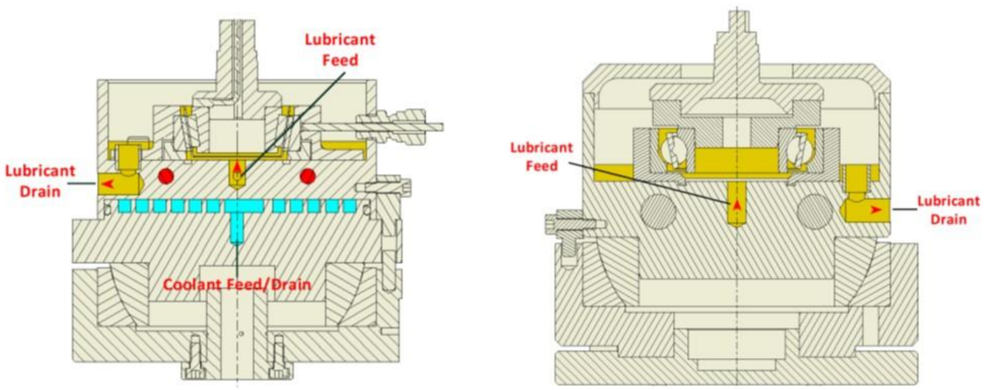
The specified test durations are:

- 4 hours, giving 348,000 revolutions
- 8 hours, giving 696,000 revolutions
- 20 hours, giving 1,740,000 revolutions

In other words, in an immersed test with just a small volume of lubricant, said lubricant is degraded long before we achieve the number of cycles necessary to produce macro-pitting.

Lubricant Test versus Material Test

Material Test:



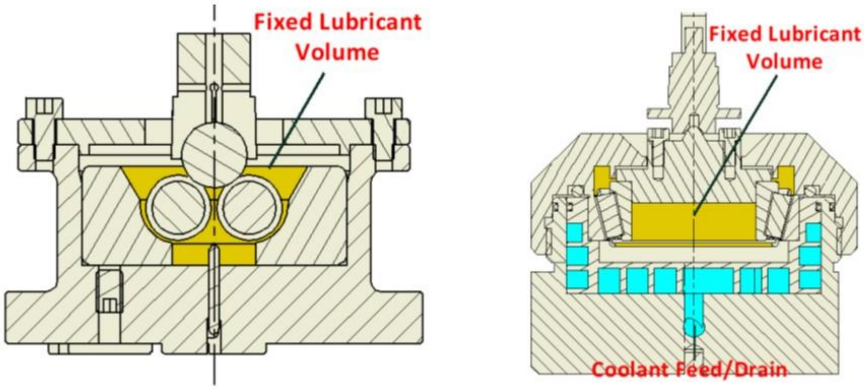
PLINT Tribology Products from Phoenix Tribology Ltd

Bearing in mind the influence of lubricants on rolling contact fatigue behaviour, plus the influence of shear and temperature on lubricant performance and life, it follows that a bearing test designed to evaluate the performance of a material will ideally be run with minimal change in the lubricant properties, and a bearing test designed to evaluate a lubricant will ideally be run in such a way as to maximise the change in the lubricant properties, in the shortest time possible.

It follows that for a test of the material properties or bearing design, lubricant should be circulated through the test assembly from a large volume source, which includes filtering, such that changes to the overall lubricant properties are limited to those associated with temperature-viscosity effects within the test assembly, and not other properties associated with exposing the lubricant to temperature and mechanical shear.

Lubricant Test versus Material Test

Lubricant Test:

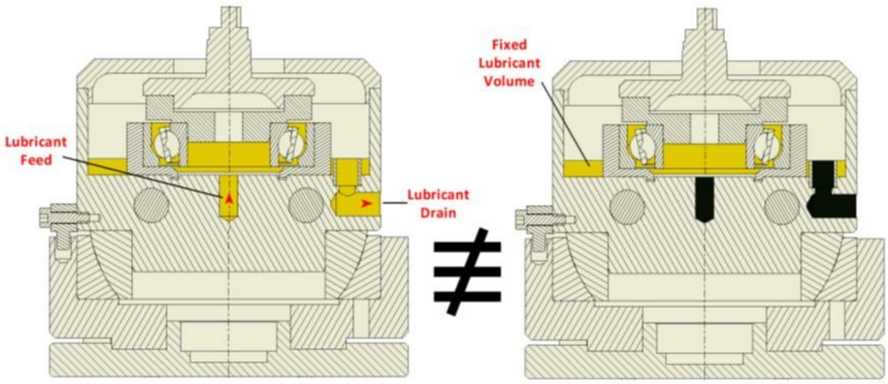


PLINT Tribology Products
from Phoenix Tribology Ltd

Bearing tests to evaluate lubricant performance invariably involve subjecting a specified small volume of lubricant to both temperature and mechanical shear, to promote lubricant degradation and explore its resulting effects on material or bearing performance.

To put this more simply and graphically:

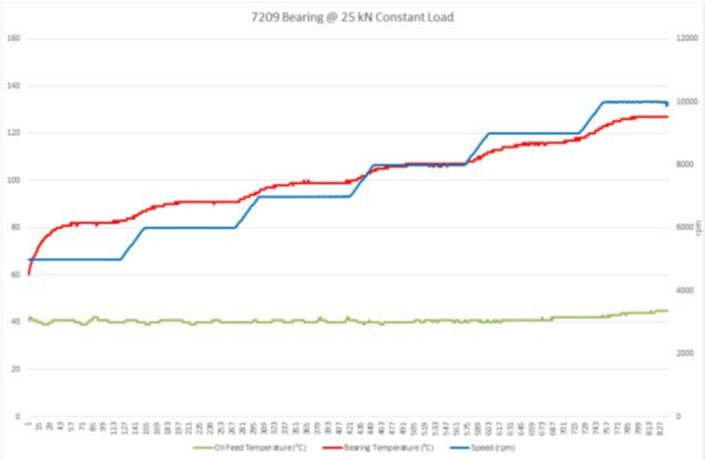
Lubricant Test versus Material Test



PLINT Tribology Products from Phoenix Tribology Ltd

Note that for higher frictional loss bearing types, such as taper roller bearings, it may be necessary to include cooling passages in the test assembly.

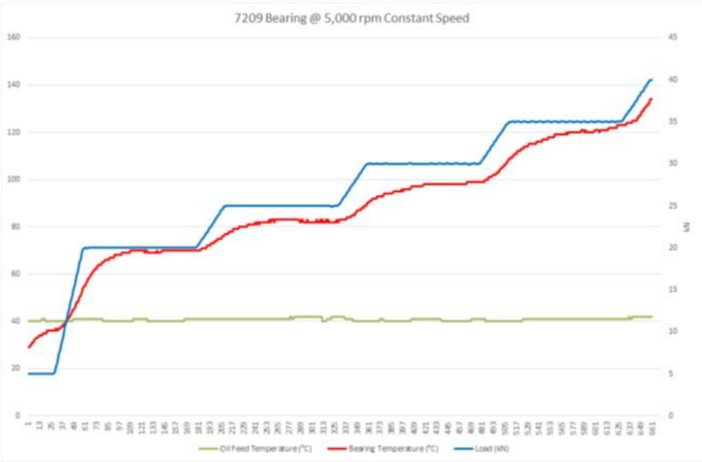
Lubricant Cooling Function



PLINT Tribology Products
from Phoenix Tribology Ltd

The graphs show the effects varying the test speed on an axially loaded 7209 bearing, under constant load and varying the load at constant speed. In both cases, increase of speed at constant load and increase of load at constant speed both leads, as one might have predicted, to increase in bearing temperature, hence reduction in lubricant viscosity and elasto-hydrodynamic film thickness. In both cases, the test bearings seized shortly after the bearing temperature exceeded 120°C. In other words, another form of bearing failure was precipitated, which was neither micro nor macro-pitting.

Lubricant Cooling Function



PLINT Tribology Products
from Phoenix Tribology Ltd

A corollary to this observation is that with a controlled lubricant supply, it may be possible to achieve a lower stable bearing test temperature with a higher lubricant inlet temperature, hence lower viscosity, and lower flow rate, hence lower churning losses, than with a lower inlet temperature and/or higher flow rate. This is an analogous problem to that created when over-filling a grease lubricated bearing; the optimum fill is typically about 25% of the bearing free space.

Measuring Lubricant Film Thickness

- Optical Film Thickness Measurement
- Capacitance Film Thickness Measurement
- Ultrasonic Film Thickness Measurement

PLINT Tribology Products
from Phoenix Tribology Ltd

Optical Film Thickness Measurement

An attractive and informative method, optical elasto-hydrodynamic film thickness measurement involves loading of a steel ball against a glass disc, with a semi-reflexive coating, however, the method suffers from one serious limitation: the strength of the coated glass. Because of this, contact pressures cannot exceed about 0.8 GPa and, in order to preserve the coating, are typically limited to 0.5 GPa.

There is a significant caveat arising from this: it takes something of an act of faith to assume that the liquid tested in an optical elasto-hydrodynamic film thickness measuring device, at 0.5 GPa, is going to behave in the same way at a contact pressure of, say, 2.5 GPa, in a contact between a bearing steel ball and a bearing steel bearing race. There is not much we can do about this problem!

Whereas optical film thickness measurement produces an accurate representation of the lubricant film thickness profile in the contact, all other methods only result in an average film thickness for the contact as a whole.

Capacitance Film Thickness Measurement

Two capacitance film thickness measurement methods are available, one in which the space between the two surfaces is treated as a capacitor and the other in which a fully isolated capacitance probe is inserted into one surface, to measure the gap between the sensor and the counter-surface.

The first method can potentially be used in systems where both surfaces are moving, for example in a two-roller type contact, however, although technically feasible, this method is technically challenging.

The second method requires the surface carrying the capacitance probe to be stationary, hence, for example, the outer race of a bearing. A small hole is drilled in the bearing race and the probe is inserted flush with the race surface. As a rolling element passes the probe a change in capacitance, which is a function of lubricant film thickness, can be detected. This is, of course, a transient measurement.

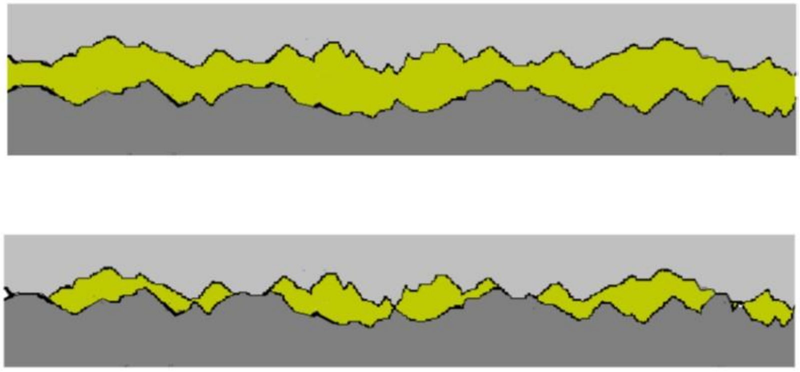
Both methods require knowledge of the dielectric properties of the oil and both are limited to producing only an average film thickness measurement.

Ultrasonic Film Thickness Measurement

Ultrasonic sensors can be fitted to a stationary component in a rolling element bearing, typically the bearing outer race. Unlike the capacitance probe, there is no requirement to drill a hole in the bearing race. Ultrasound measurement utilises the fact that sound waves are reflected from surface interfaces, hence at the boundary between the bearing inner race and the lubricant and between the lubricant and the surface of the rolling element.

Reflection, resonance and phase-based methods can be used to determine the lubricant film thickness. Although a viable measurement technique for systems in which a steady state lubricant film thickness occurs, for example in a plain journal bearing, we have some doubts about how practical such measurements are when measuring transient film thickness as must occur when a rolling element passes a point adjacent to the sensor.

Detecting Asperity Contact



If we have two surfaces separated by a thick, non-conducting, fluid film, we effectively have an open circuit "switch", so nominally an infinite resistance. A contact resistance signal will thus be the open circuit voltage of the system.

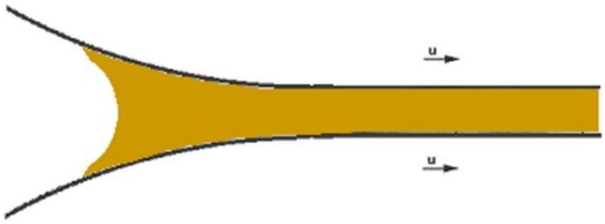
If the fluid is a mineral oil, it will have very high electrical resistance; oils are good insulators. If the surfaces are brought together but remain out of contact at an asperity level, the "switch" will still be open circuit, so the contact resistance signal will still be high. In effect, once there is no metal-metal contact in the "switch", it does not make any difference whether the surfaces are separated by 100 microns, 100 mm or 100 m!

As the surfaces are brought together, perhaps through a reduction in entrainment velocity or lubricant inlet viscosity, electrical contact is made through the film, initially occurring at a single asperity level, at which point the "switch" is momentarily closed, producing a short circuit between the surfaces. As a result of this, the voltage across the contact falls to zero, while the asperities remain in contact.

It will be obvious that each asperity is in effect an individual "switch", so just one asperity coming into contact will be sufficient to produce a short circuit. However, we may have many more than one asperity in contact, in other words, we may have multiple "switches" in parallel. There is no way of telling from the contact resistance measurement how many closed "switches" there are; it could be one, it could be one hundred.

So, whereas a contact resistance signal cannot be used for determining lubricant film thickness or the number of asperity contacts, it can be used for indicating whether there are or are not any asperity contacts, hence whether the contact is running with full elasto-hydrodynamic separation or under a mixed lubrication regime.

Starved Lubrication



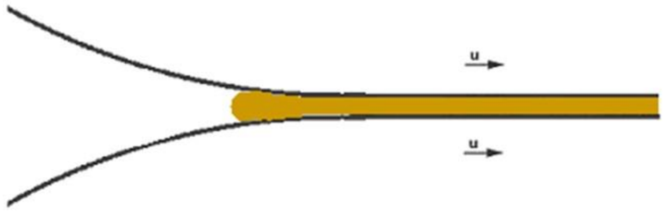
$$Entrainment\ Velocity = \frac{1}{2}(U_1 + U_2)$$

PLINT Tribology Products
from Phoenix Tribology Ltd

To generate elasto-hydrodynamic separation of the surfaces, any system must fulfil three requirements. Firstly, a lubricant with sufficient viscosity and pressure-viscosity coefficient is required. Secondly, there must be a mechanism for delivering the lubricant to where it is required and in sufficient quantity to flood the inlet to the contact. Thirdly, there must be a sufficient entrainment velocity to carry the lubricant into the contact, in order for it to generate the necessary elasto-hydrodynamic separation.

The sketch illustrates what happens in an elasto-hydrodynamic contact between surfaces running with the same surface velocity, with a fully flooded inlet and a lubricant of sufficient viscosity.

Starved Lubrication



PLINT Tribology Products
from Phoenix Tribology Ltd

As indicated previously, the system is weakly dependent on load, so that once an elasto-hydrodynamic film, of sufficient thickness to separate the surfaces, is generated, increasing the load has little impact on the level of separation.

This is what happens if:

1. There is insufficient lubricant to flood the inlet - starved lubrication.
2. The lubricant viscosity is too low – potentially caused by too high temperature at the inlet.

In this case, we cannot prevent intimate contact between the mating surfaces at an asperity level, so we may end up with seizure or micro-pitting, depending on the test conditions.

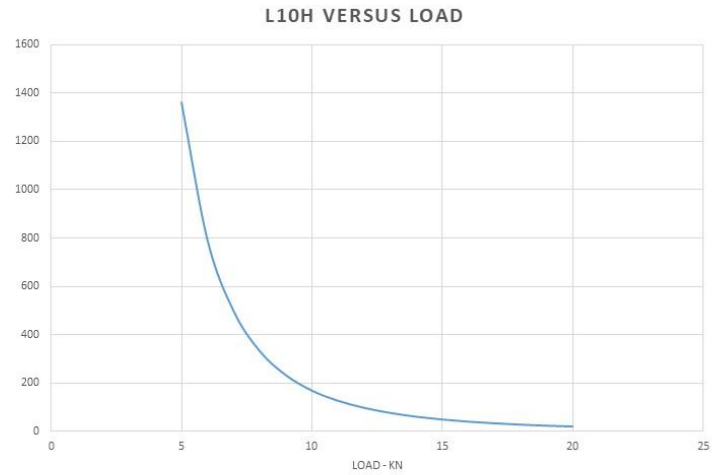
Churning Losses

- Increase in frictional torque may be caused by churning or shearing of the lubricant, sliding contact between rolling elements and the cage, skidding in lightly loaded bearings etc, all of which cause the bearing to heat up and possibly fail
- Measuring frictional torque in bearing useful if one wishes to calculate the power losses associated with a given bearing, otherwise, measuring the bearing outer race temperature is a useful and less complicated way of determining if something is going wrong

PLINT Tribology Products
from Phoenix Tribology Ltd

Frictional torque in a correctly lubricated rolling element bearing, starts more or less at zero, then increases with increasing speed of rotation. Rolling element bearings are, by definition, very low friction devices and measurement of frictional torque is not of very general interest; it only becomes of interest in more specialist circumstances, usually associated with operating bearings at very high speeds. In this case, the increase in frictional torque may be caused by churning or shearing of the lubricant, sliding contact between rolling elements and the cage, skidding in lightly loaded bearings etc, all of which cause the bearing to heat up and possibly fail. Measuring the frictional torque in the bearing may be useful if one wishes to calculate the power losses associated with a given bearing, otherwise, measuring the bearing outer race temperature is a useful and less complicated way of determining if something is going wrong.

Load and Fatigue Life Relationship

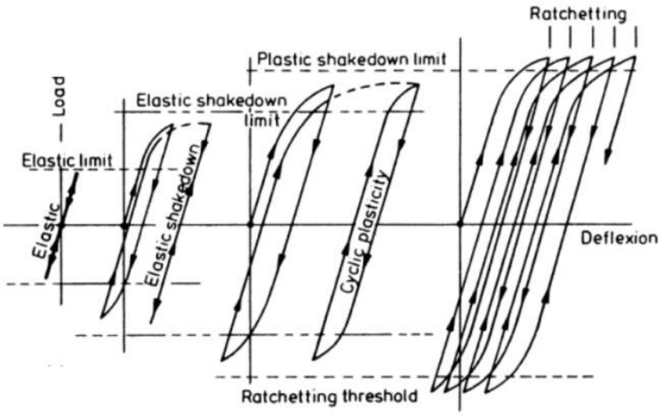


PLINT Tribology Products
from Phoenix Tribology Ltd

Higher loads lead to a shorter time before crack initiation and propagation, thus reducing the bearing's fatigue life. The following graphs have been generated using the SKF on-line bearing life calculator for an axially loaded 7206 angular contact bearing running at 6,000 rpm at a temperature of 70°C.

The key thing to note from this is that increasing the load from 5 kN to 10 kN reduces the L10 life from 1360 hours to 170 hours, in other words, by a factor of 8. Increasing the load from 10 kN to 15 kN reduces the L10 life to 51 hours, in other words, by a factor of 3.3. Bearing in mind the risk of precipitating other failure mechanisms through applying excessive load, reference to this curve provides some guidance with regard to choosing the optimum test conditions for rolling contact fatigue tests: not too long and not too short!

Importance of Running-in



The influence of repeated loading, residual stresses and

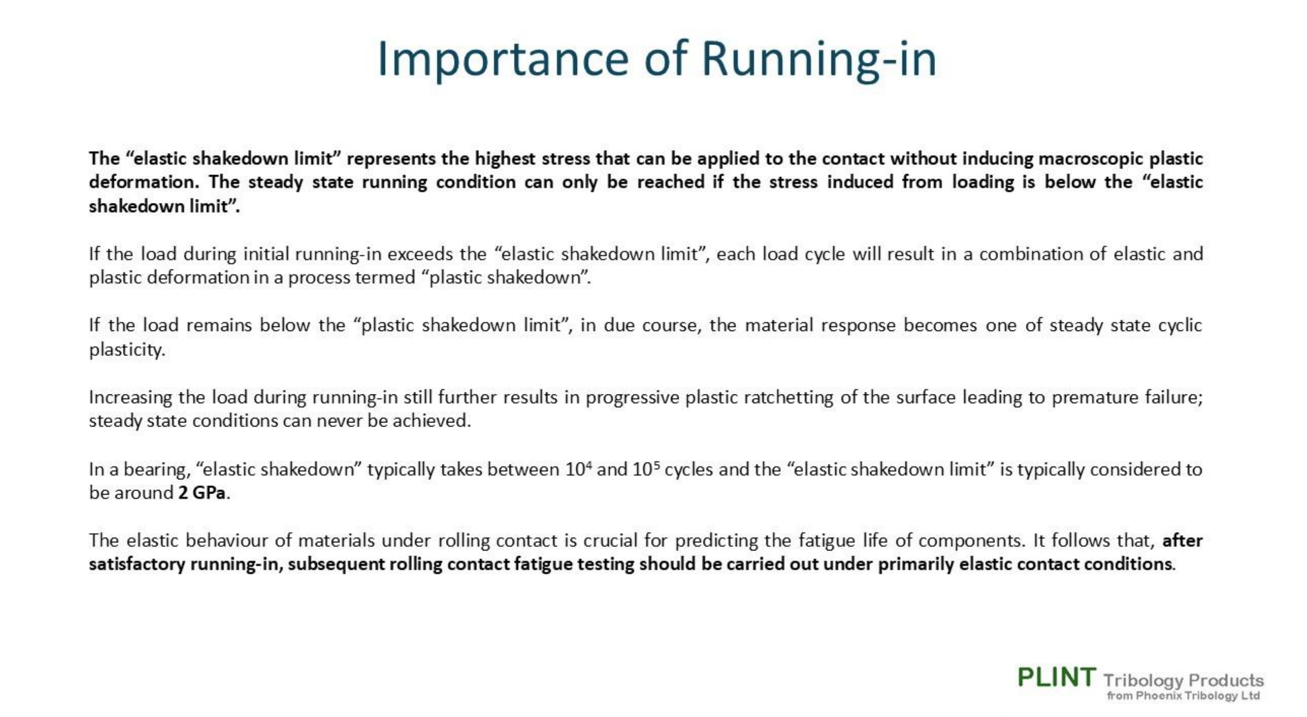
shakedown on the behaviour of tribological contacts

J. Williams - Tribology International 38 (9) (2005) 786-797

PLINT Tribology Products
from Phoenix Tribology Ltd

In rolling contacts and rolling element bearings, "shakedown" refers to the initial running-in stage where, providing the contact pressure is not too high, microplastic deformation and work hardening occur. This leads to a buildup of residual stress and a subsequent increase in the contact's ability to withstand cyclic loads. After this shakedown stage, the bearing material can behave elastically, under loads above those associated with the initial, pre-shakedown, elastic limit. The material has become harder and stronger.

Residual stresses are stresses that remain in the material even when the external load is removed. After “shakedown”, the load experienced by the material is subjected to both contact stresses and the residual stresses introduced during the “shakedown” process. At this point, further plastic deformation ceases to be possible, and the material achieves a steady state running condition, which is a necessary pre-condition for any subsequent rolling contact fatigue test. This running-in process is referred to as “elastic shakedown”.



The “elastic shakedown limit” represents the highest stress that can be applied to the contact without inducing macroscopic plastic deformation. The steady state running condition can only be reached if the stress induced from loading is below the “elastic shakedown limit”.

The influence of repeated loading, residual stresses and shakedown on the behaviour of tribological contacts

J. Williams

Tribology International 38 (9) (2005) 786-797

If the load during initial running-in exceeds the “elastic shakedown limit”, each load cycle will result in a combination of elastic and plastic deformation in a process termed “plastic shakedown”. If the load remains below the “plastic shakedown limit”, in due course, the material response becomes one of steady state cyclic plasticity. Increasing the load during running-in still further results in progressive plastic ratchetting of the surface leading to premature failure; steady state conditions can never be achieved.

In a bearing, “elastic shakedown” typically takes between 10^4 and 10^5 cycles and the “elastic shakedown limit” is typically considered to be around 2 GPa.

The elastic behaviour of materials under rolling contact is crucial for predicting the fatigue life of components. It follows that, **after satisfactory running-in, subsequent rolling contact fatigue testing should be carried out under primarily elastic contact conditions**.

Detecting Pitting Failure

- Macro-pitting - vibration sensor
- Micro-pitting - optical examination

Photograph: © NSK Europe

PLINT Tribology Products
from Phoenix Tribology Ltd

Detecting macro-pitting is easy using a simple vibration sensor or microphone “buzzer”. This is because when macro-pitting occurs it results in a large amount of noise and machine vibration.

Detecting micro-pitting is a much more challenging proposition, as the levels of surface damage are typically too small to induce significant mechanical vibration, indeed damage may be limited to nothing more than visible dulling or greying of the surface. In this case, optical examination is the primary method of establishing if micro-pitting has occurred, however, this clearly requires the test surfaces to be accessible for observation.

Another yet unproven method of detection, that does not require direct observation of the surfaces, is to use low ohmic value electrical contact resistance measurement to detect changes in conductivity of the contact. The assumption here is that the presence of micro-pits may increase retention of oil on the bearing surface, resulting in a small and detectable change in conductivity.

Analysis of Vibration Signals

- Frequency Analysis
- Order Analysis
- Wavelet Analysis

PLINT Tribology Products
from Phoenix Tribology Ltd

For determining failure by macro-pitting, we need nothing more than the un-calibrated output from a basic vibration sensor or microphone. The output can be a time smoothed and damped signal; the magnitude of the signal is meaningless and is only used as a trigger or alarm device, not for making an analytical measurement.

If we want to measure vibration in a meaningful way, we must use a properly calibrated accelerometer, with high-speed data acquisition of the dynamic signal. Having acquired the signal, for it to be of any use at all, we need to process and analyse the output, in order to convert the raw data into meaningful content.

Frequency Analysis

Frequency analysis is the most commonly used method for analysing a vibration signal. The most basic type of frequency analysis is an FFT, which converts a signal from the time domain into the frequency domain, allowing a power spectrum to be generated, showing the energy contained in specific frequencies of the overall signal. This is useful for analysing stationary signals, whose frequency components do not change over time; it is not suitable for signals whose frequencies vary over time. If we have something like a bearing rotating at a given speed, it will generate a signature frequency or frequencies; we might sensibly decide to focus on this specific frequency, for monitoring and analysis purposes.

Order Analysis

Order analysis is used to analyse systems where the excitation frequency varies, for example, changes in running speed of a motor, so it is geared specifically towards the analysis of rotating machinery and how vibration levels and frequencies change, as the rotational speed of the machine changes.

Wavelet Analysis

Wavelet analysis is used for characterizing machine vibration signatures with narrow band-width frequencies. In essence, we pick a given frequency that we wish to monitor and use that as our measured parameter. This is a bit like saying: “I am not interested in the overall volume of the music; I am only interested in monitoring the volume of noise at Middle C (256 Hz if you are a mathematician and 261 Hz if you are a musician!).

Then there are more complicated forms of analysis, such as model based analysis.

Hence, we really need to decide what it is we want or think we need to measure, when it comes to vibration measurement and analysis, then choose an appropriate bandwidth sensor, then choose an appropriate analysis system. It is not enough just to data log some parameter; we really need to know what the signal means and how it is to be analysed and interpreted.

Rolling Test Configurations

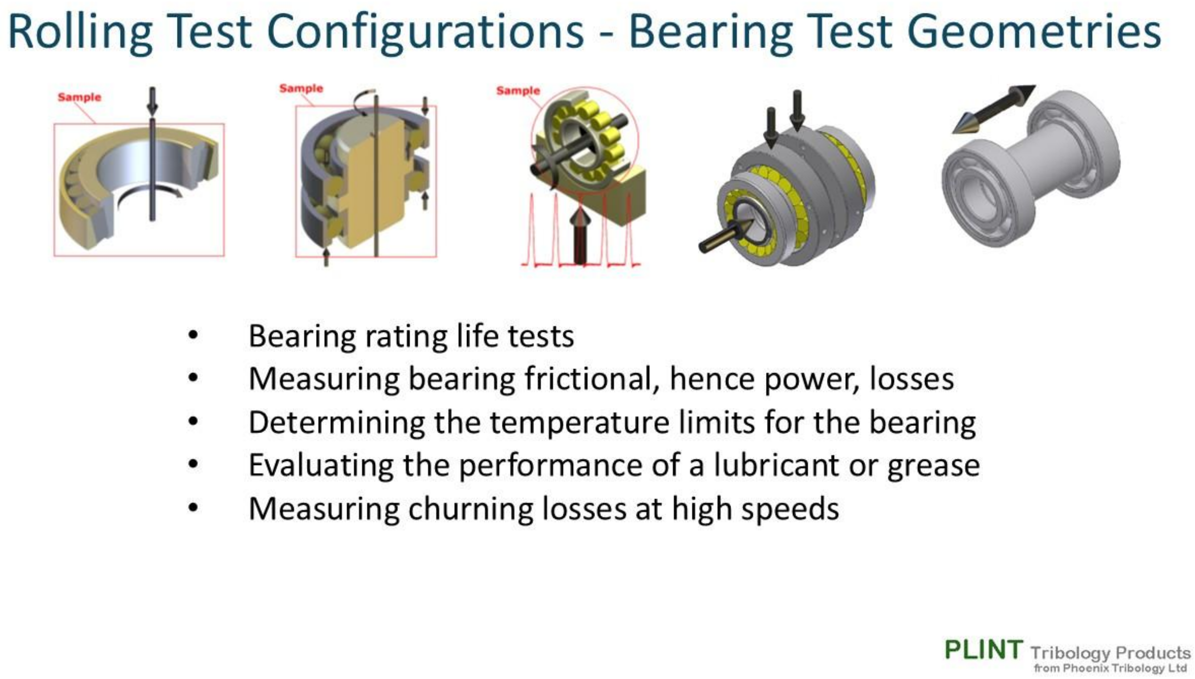
- Bearing rating life
- Friction and power losses
- Bearing temperature limits
- Bearing churning losses at high speeds
- Lubricant/grease evaluation
- Rolling contact fatigue life of candidate materials

PLINT Tribology Products
from Phoenix Tribology Ltd

Having rehearsed many issues, we can now summarise how to use the many different test geometries available for investigating rolling contact fatigue and other aspects of bearing life. Here are some of the things we may wish to investigate:

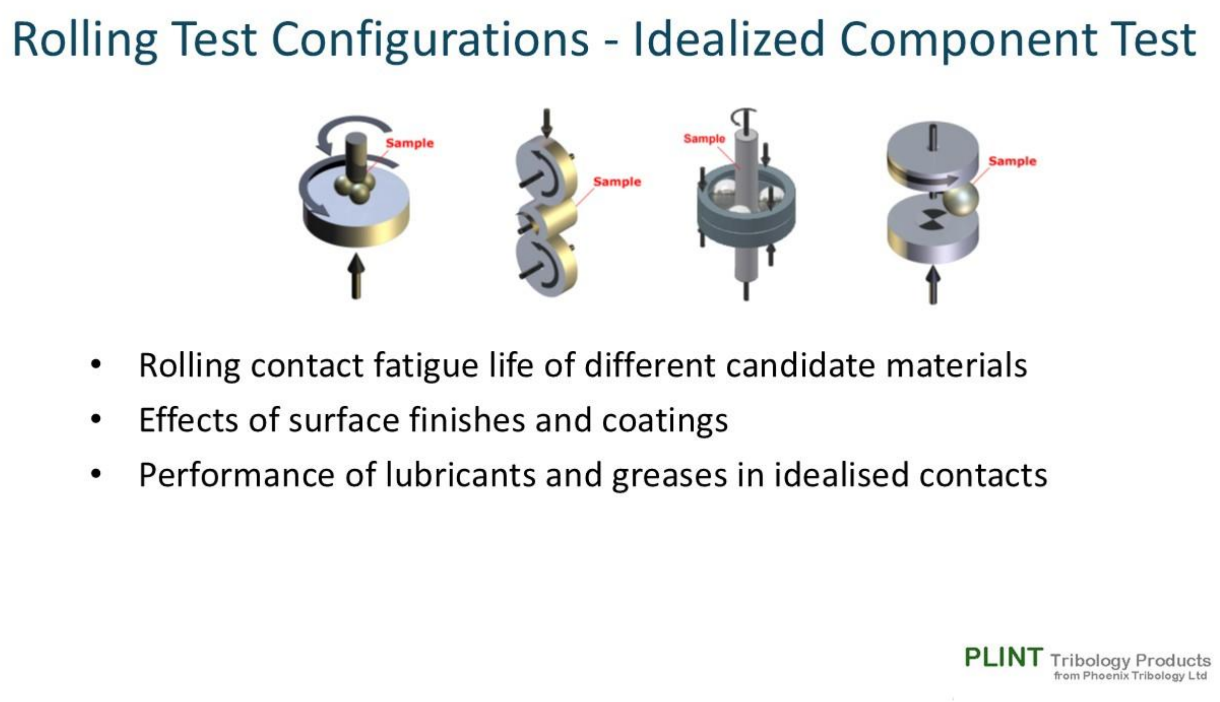
- Bearing rating life
- Friction and power losses
- Bearing temperature limits
- Bearing churning losses at high speeds
- Lubricant/grease evaluation
- Rolling contact fatigue life of candidate materials

Pure rolling test configurations can be divided into three basic categories:



These tests use standard rolling element bearings as the test component and can be used for:

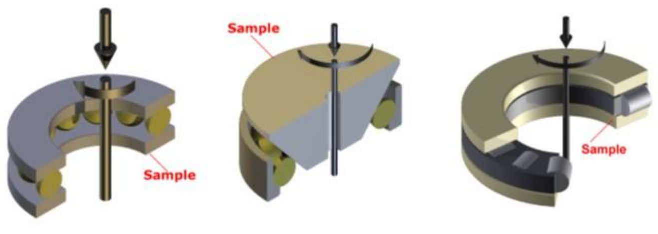
- Bearing rating life tests
- Measuring bearing frictional, hence power, losses
- Determining the temperature limits for the bearing
- Evaluating the performance of a lubricant or grease
- Measuring churning losses at high speeds



These tests are aimed at stressing a single rolling element component, either a ball or a rod. These tests are used for investigating:

- Rolling contact fatigue life of different candidate materials
- Effects of surface finishes and coatings
- Performance of lubricants and greases in idealised contacts

Rolling Test Configurations - Hybrid Component Test Geometries



- Rolling contact fatigue life of candidate materials

PLINT Tribology Products
from Phoenix Tribology Ltd

Hybrid test geometries use the rolling elements and one raceway from standard rolling element bearings, with the other raceway replaced with a simplified and easy to manufacture specimen, of candidate material, in the form of a disc or cone. These hybrid geometries are primarily designed for evaluating:

- Rolling contact fatigue life of candidate materials

Conclusions

- Decide whether a bearing test, a materials test or a lubricant test is required
- Choose appropriate test geometry
- Choose appropriate test conditions, including lubrication regime, to achieve the required failure mechanism
- Include appropriate running in procedures, at contact pressures that will achieve satisfactory elastic shakedown
- Run test at contact pressures consistent with the production of the required failure mechanism

PLINT Tribology Products
from Phoenix Tribology Ltd

Before we start, we need to decide whether we need a bearing test, a materials test or a lubricant test. We then need to choose an appropriate test geometry and test conditions, paying particular attention to the lubrication regime, under which our tests need to be run, hence to achieve the required failure mechanism. Our test must include appropriate running in procedures, at contact pressures that will achieve satisfactory elastic shakedown. Subsequent tests must be run at contact pressures consistent with the production of the required failure mechanism. This will rarely exceed a nominal contact pressure of approximately 2.5 GPa.

It is worth re-stating that for all lubricated wear tests, without exception, knowledge and control of the lubrication regime is an essential requirement of any meaningful test.