The Acoustic Emission from Mixed Elastohydrodynamic Lubrication

Simon Hutt

A thesis submitted in partial fulfilment for
the degree of Doctor of Philosophy

Tribology Group

Cardiff School of Engineering
Cardiff University

June 2018

Summary

This thesis presents an investigation of the Acoustic Emission (AE) from mixed Elastohydrodynamic Lubrication (EHL). In mixed EHL the contact surfaces are partially separated by an oil film, and partially by direct contact between opposing roughness asperities. The proportion of each can vary and excessive amounts of asperity contact can precipitate surface damage. Researchers have recently shown that, in mixed EHL, asperity contacts are a source of AE (a class of mechanical wave) and the aim of this thesis was to characterise the relationship between the AE and the amount / severity of the asperity contacts.

Two experiments were conducted on a disk machine which simulates the lubrication conditions of heavily loaded transmission gears. The rig was instrumented with two AE sensors placed on the test disk, one interfaced using a novel oil-coupling. In the first experiment the thickness of the lubricating film was varied using speed and temperature control. The AE measurements were found to be susceptible to noise but after frequency filtering a clear and precise exponential relationship between the amount of AE and the amount of direct asperity contact was identified and modelled. The oil-coupling was found to be noisy but fundamentally feasible.

In the second experiment two previously unused disks were tested. All operating parameters were kept constant and the test disks were allowed to wear naturally. Surface measurements confirmed a rapid and large initial reduction in asperity height caused by running-in, and then a much slower and more limited reduction caused by further wear, which included some micro-pitting. The wear affected the amount of AE which was found to correlate well with changes in the (RMS roughness) of the contact surfaces. The AE measurements revealed that running-in of asperities occurs immediately, i.e. within the start-up of a new contact.

Acknowledgments

Organisations

I must start by acknowledging the following organisations: The EPSRC, who funded this research (and my lavish lifestyle). The Peter Jost Travel Fund, which supported the presentation of this work at CMMNO2018 (a condition monitoring conference). REM Surface Engineering, who did some excellent pro bono surface finishing for me. The organising committee of the Leeds-Lyon Symposium on Tribology 2017, who had the good sense to award me the Maurice Godet Award (which comes with a hefty cheque) for my paper presented at said symposium. And finally Cardiff University which has been my sanctuary from the real world for the past 8 years.

Individuals

Firstly I would like to thank Dr. A. Clarke, Prof. H.P. Evans and Dr. R. Pullin for their invaluable support. They provided a welcome ‘light-touch’ regulation and yet always made themselves available when I needed assistance.

Next I would like to thank Little Goode, my girlfriend, nay, life-partner and soul mate. She has provided considerable emotional support, never once describing conversations regarding my work as “a natural sedative.” Doubtless she will be overjoyed at the extra-time I will now have to spend with her.

I am also indebted to my parents for the strong work ethic they beat into me when I was younger, and their occasional phone calls to check that I hadn’t expired in the laboratory or at the keyboard.

And last, and least, I would like to acknowledge the existence of all the boys from the office (sadly no girls found their way in). In order of preference: Ben, Will, Dewi, Paul, Kostas, Othman, Sujit, Andy, Salem and Aaron. I have enjoyed the welcome distractions from work that you provided (some more than others), and the excellent Blitz mentality of our office. AE for the boooisss!!!

Notation

(This list is restricted to uses that occur commonly throughout this thesis. Limited uses are defined as needed).

ADC	Analogue to Digital Conversion
AE	Acoustic Emission
BAE_{RMS}	RMS of a broadband AE signal
CV	Contact Voltage
EHL	Elastohydrodynamic Lubrication
$\text{FAE}_{\text{RMS X-X}}$	RMS of a filtered AE signal where X-X refers to the pass band
HL	Hydrodynamic Lubrication
RMS	Root Mean Square
	RMS roughness
SRR	Slide / Roll Ratio
	Specific film thickness

Related publications

As of Jan 2019 aspects of this work have been published / are under review as follows:

Hutt S., Clarke A., Evans H.P., 2018, Generation of acoustic emission from the running-in and subsequent micropitting of a mixed-elastohydrodynamic contact, *Tribology International,* Vol. 119, pp. 270–80.

Hutt S., Clarke A., Pullin R., Evans H.P., 2018, Characterising the acoustic emission from a simulated gear contact in mixed lubrication conditions, Presented at: Condition monitoring of machinery in non-stationary operations, Santander, Spain, 20th – 22nd June 2018. [Peer reviewed, pending publication].

Hutt S., Clarke A., Pullin R., Evans H.P., 2019, The acoustic emission from asperity interactions in mixed lubrication. *Proceedings of the Royal Society, Part A.* [Submitted, pending acceptance and peer review].

Al-Mayali MF., Hutt S., et al., 2018, Experimental and numerical study of micropitting initiation in real rough surfaces in a micro-elastohydrodynamic lubrication regime, *Tribology Letters*, Vol. 66, pp. 150-64.

Contents

1 Introduction

For transmission gears, where the speeds of rotation and loads can be high, the presence of a lubricant is essential. Without this, catastrophic damage to the tooth surfaces would quickly occur. Elastohydrodynamic Lubrication (EHL) is a specific type of lubrication mechanism that occurs in heavily loaded concentrated contacts such as those between gear teeth. The essential feature of EHL is that the motion and geometry of the contacting surfaces entrains (draws) a thin film of highly pressurised oil into a contact area generated by elastic deformation of the surfaces. In normal operation the thickness of this film can change significantly as it depends on many parameters which are often variables, such as speed and temperature.

The surfaces of transmission gears tend to have a roughness, or finish, with asperity (peak) heights that are of a similar scale to the thickness of an EHL oil film. It is possible for the oil film, if thick enough, to completely separate all the asperities of the opposing surfaces. However in many real operating conditions the oil film will be too thin for this, in which case the asperities of the opposing surfaces will impinge on each other. This results in numerous transient instances of direct asperity-to-asperity contact, and lubrication conditions in which this occurs are known as ‘mixed’.

Mixed lubrication is not a single condition but a spectrum representing a variable amount, or intensity, of direct asperity contacts. Operation at the aggressive end of this spectrum, i.e. where there are continuous high intensity asperity contacts, can lead to damaging surface wear. It would thus be useful if the position on this spectrum could be easily measured. Although there are some existing techniques which allow this, they are often impractical to implement outside of the laboratory. AE may offer a new, more flexible, way to measure the position within the mixed lubrication spectrum.

AE are transient high frequency waves that are spontaneously released, and propagate through, dense materials when there is a rapid release of strain energy, such as when a crack grows. Researchers have shown that moving contacts generate AE, the amplitude and frequency of which can vary with the contact parameters. There is agreement that asperity contacts in mixed lubrication gear contacts are a source of AE, but a coherent relationship between the levels of each has yet to be established. Part of the difficulty in trying to uncover one is that many of the contact parameters in meshing gear teeth are always in a state of flux, and this makes correlating the AE to the precise lubrication conditions difficult.

Many of the salient features of real gear contacts can be simulated by the contact formed between two disks, the simplicity of which is more suited for fundamental research. Cardiff University has a highly instrumented disk rig which has been successfully used by previous researchers to investigate various phenomena associated with gear lubrication. The aim of this work was to explore the fundamental relationship between AE and the amount of asperity contact using the simplified contact of this disk rig.

The thesis is structured as follows. In Chapter 2 the topics that have been introduced here are explained in more detail, and the recent literature regarding the AE from mixed lubrication contacts is reviewed. In Chapter 3 the test rig is described. Chapters 4 and 5 present two experiments designed to reveal different aspects of the relationship between AE and the amount of asperity contact; these chapters consist of the procedure, results and discussion specific to each. Finally Chapter 6 concludes with a summary of the important findings, the novelty of this work, and its implications.

2 Theory and literature review

To appreciate the experiments presented in this thesis it is necessary that the reader has some understanding of firstly, the tribology (lubrication and wear) of gear contacts, and secondly, acoustic emission. Both of these are explained in this chapter. This chapter starts by introducing some of the well-established and fundamental theory that explains how ideal, and stationary, gear teeth touch. The complications of lubricated motion, roughness and wear are then gradually incorporated, to build up understanding of how real gears contact. After this the topic of AE is introduced, starting with the fundamentals of what AE are and their broad use. The application of AE measurements for gear contacts is then discussed using examples from recent literature. And finally the resulting objectives for this research are given.

2.1 Non-conformal contacts and their geometry

Consider contact between an unlubricated, and highly polished, gear pair which are loaded but at rest. (This would be an unusual situation but it serves as a useful starting point with which to introduce contact theory.) The geometry of gear teeth is such that the contact between opposing teeth is *non-conformal*. As an illustration of the difference between a conformal and non-conformal contact consider the example, in Figure 2-1, of two *perfectly smooth* and *rigid* cylinders. When touched together end on end, they contact over an area; this is conformal contact. However when touched together in any other orientation they only contact at an infinitesimal point or a line (a line in the pure sense of a length without width), i.e. the contact has an infinitesimal area; this is non-conformal contact.

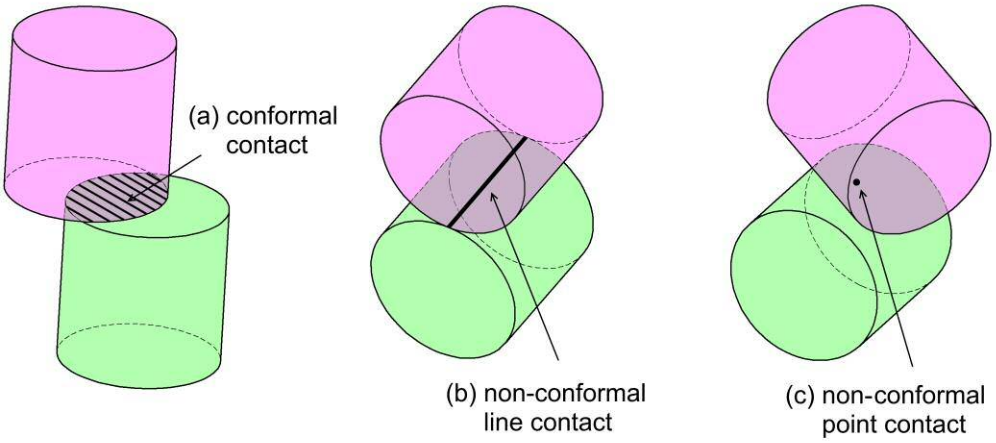


Figure 2-1 Examples of conformal and non-conformal contact between smooth and rigid cylinders.

Essentially the type of contact, conformal or non-conformal depends on the curvatures, or radii, of *both* bodies. Consider the case of Figure 2-1(b): two axially aligned cylinders, (also known as ‘rollers’). Figure 2-2(a) shows their curvatures at the point of contact, and the gap between their surfaces, which is a function of the distance, *x,* from their point of contact. This gap can be modelled as that between a cylinder and a plate where the cylinder has a *relative* curvature which results in an equivalent gap[1]. This relative geometry is often used to simplify analysis of non-conformal contacts, and provided the curvatures are small compared to *x,* i.e. the radii are large, it can be easily calculated.

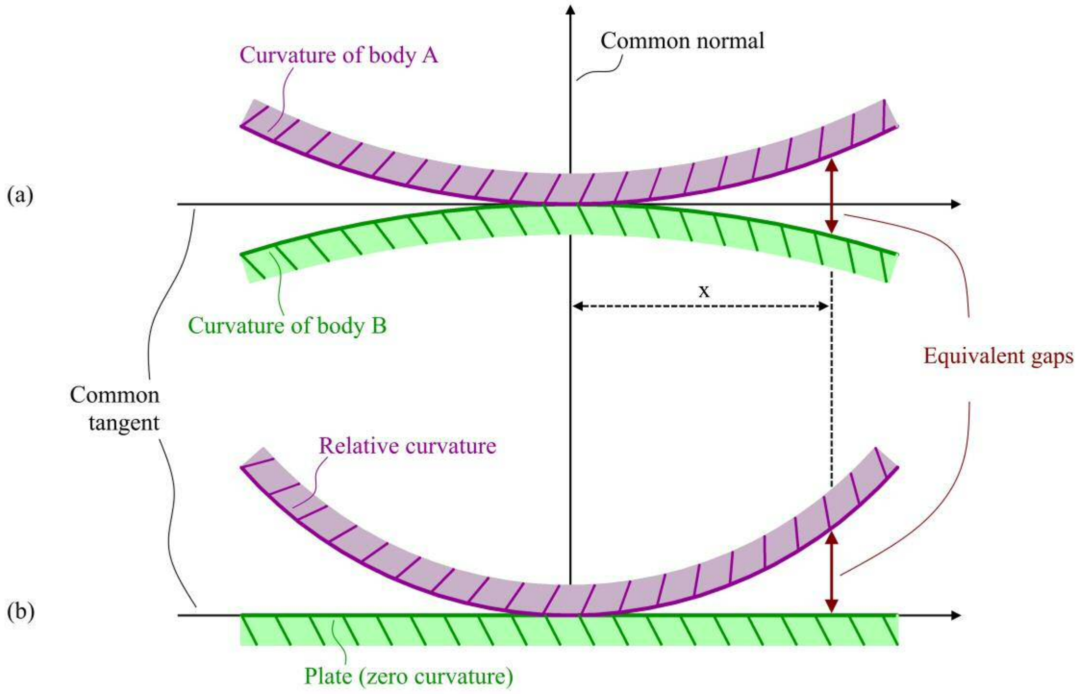


Figure 2-2 The geometry of two rigid cylinders that are touching. (a) their actual curvatures, (b) the relative curvature.

2.2 The Hertzian contact model

In practice materials are elastic, not perfectly rigid, and when the surfaces of non-conformal bodies touch they elastically deform so that they contact over an area. The nature of the contact: its shape, area, pressure distribution and deformation, depends on the geometry, load, and elasticity of the contacting bodies.

In 1882 the German physicist Heinrich Hertz published a seminal analysis of elastic contact [1,2] and its use has endured to this day. His method is based on a number of constraints. The first is that the equivalent rigid surfaces, must be continuous, non-conformal, and have geometry no more complex than quadratic, which Hertz showed would result in an elliptical contact area when they were deformed. There are special cases where the contact area is circular or linear[2], but these can be viewed as special cases of an ellipse where its semi dimensions are, equal for the former case; and for the latter case, one is finite and the other infinite. The second constraint is that the contact deformations are small compared to the dimensions of the bodies.

---

[1] Of course for the case of an actual cylinder on a plate, its curvature is identical to the relative curvature.

[2] A ‘linear’ Hertzian contact is one in which the contact area is a long and narrow rectangle.

And the final constraint is that there is no traction (friction) between the surfaces, and so that the contact load acts in a direction normal to their tangent plane.

Using these constraints Hertz derived expressions that can be used to determine the absolute dimensions of the contact ellipse, as well as the pressure distribution and surface deflection. To calculate these contact parameters all that is required is knowledge of the relative curvature between the two bodies at the contact, their elasticity and the contact load. From first principles the analysis is not straightforward as it requires the manipulation of elliptic integrals. However the solution only varies with the ratio of two orthogonal relative curvatures, and tabular / graphical results for common ratios are available. This makes the calculation of the Hertzian contact parameters a straightforward matter.

Figure 2-3 illustrates an elliptical Hertzian contact. (a) shows the surfaces of two convex solids at the point of touching, where both contact load and area are infinitesimal, and the surfaces are not yet deformed. For a Hertzian analysis the co-ordinate system is orientated as shown, with the origin at the point of touching and the plane coincident with the common tangential plane. The and axes are aligned with the principal axes of relative curvature, meaning, one is aligned to the surface cross-section having the greatest relative curvature and one with the cross-section having the least. (For a quadratic surface, these principal axes are always 90° apart). (b) shows the deflected surface when the load is increased. A common elliptical interface is generated and the major and minor dimensions of the ellipse are aligned with the principal axes of relative curvature. A semi-elliptic pressure distribution is generated over the area of the ellipse with the maximum pressure acting at the origin. (In Figure 2-3 (b) the pressure acting on the x axis is only shown for the lower surface. The pressure acting on the upper surface will be equal and opposite.)

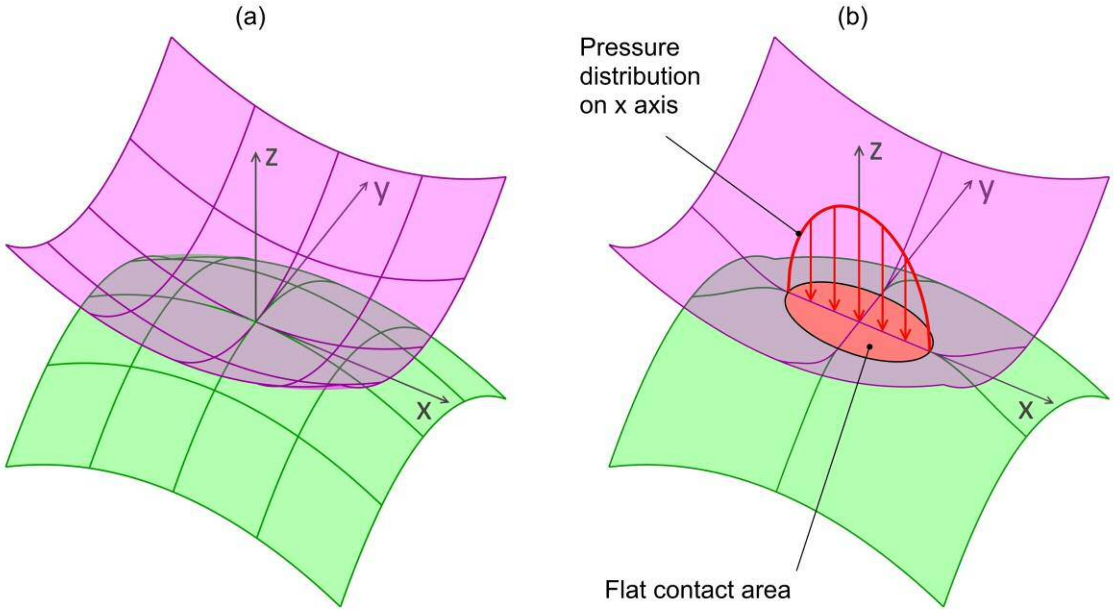


Figure 2-3 Illustration of an elliptical Hertzian contact. (a) the surfaces of two convex bodies at the point of touching. (b) the surfaces deformed under loading.

2.3 Hertzian analysis of gears

The Hertzian model has been found to be useful starting point for analysing many of the real contacts found in machinery, in particular those found in rolling element bearings and between gear teeth. These components have non-conformal geometry and high stiffness which makes the Hertzian contact model suitable. And when applied it reveals that, for typical operating conditions, their contacts have small areas: in the order of millimetres squared; and extremely high pressures: in the order of gigapascals.

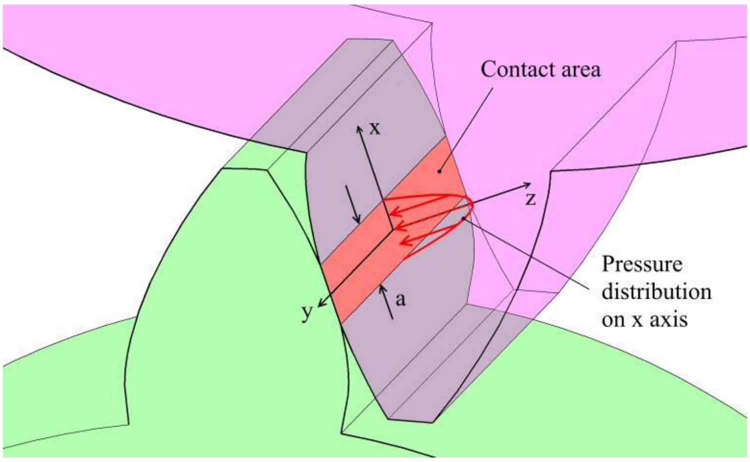


Figure 2-4 Hertzian contact between spur gears.

Returning to the example of a loaded gear pair at rest: Figure 2-4 shows a pair of spur (straight) teeth in the middle of the meshing cycle, i.e. the teeth are in contact about halfway up their flanks. and are the principal axes of relative curvature and z is normal to the tangential plane. Spur tooth geometry results in one of the special cases of Hertzian analysis. Both surfaces have an infinite radius of curvature in the direction thus generating a linear contact, of width . The pressure distribution is semi-elliptic in the direction and uniform in the direction. Most gear teeth have an involute[1] profile, thus do not have a constant radius of curvature. The implication of this is that the Hertzian contact parameters vary over the meshing cycle. And at the edges of the idealised teeth in Figure 2-4, the Hertzian constraint of continuous geometry is not satisfied as so the analysis becomes less accurate as they are approached. Despite this, provided the tooth face is wide enough so that the contact area is much longer than it is wide, satisfactory results may be obtained. And in practice teeth can be axially crowned and radiused at the ends which will alleviate, to some extent, the effects of the geometric discontinuity.

When the teeth of spur gears mesh the contact is generated across the entire face width instantaneously. This can lead to significant noise and vibration when the gears operate at speed. To alleviate this problem helical teeth are often used. As in spur gears the tooth profile is an

---

[1] An involute is a type of spiral. The end of a taut string unwrapped from a cylinder will trace an involute curve.

involute but instead of the profile being ‘extruded’ straight along axis of rotation it is ‘extruded’ along a helical path. The result of this is that when two helical teeth mesh they do so gradually. At first a point contact is generated at the end of the teeth and as the teeth move through the meshing cycling the contact extends along the tooth face becoming a line. This line then retracts and becomes a point at the other end of the teeth as they disengage at the end of the meshing cycle. This complicates the Hertzian analysis but nevertheless it can still produce satisfactory results.

In summary, the Hertzian model is a simple yet accurate starting point for examining the contact between gears. But it does not account for complications caused by rough surfaces, lubricants and motion, and once these are introduced the contact parameters, in-particular the pressure distributions, can vary significantly from the Hertzian ideals.

2.4 Lubricants

Lubricants are ubiquitous and their use in machinery with moving contacts is essential (as anyone, the author included, who has allowed a car engine to run out of oil will no doubt attest). In gears, as generally elsewhere, their primary function is to reduce friction. There are two different lubrication mechanisms which may achieve this: ‘boundary’ and ‘fluid film’ [3], and both or only one may be present in a contact. Boundary lubrication acts by altering, at the atomic scale, how *directly touching* surfaces interact chemically and mechanically. Many lubricating oils are formulated with specific additives that will react with a metal surface to deposit an extremely thin low friction coating which functions as a boundary lubricant. In contrast, fluid film lubrication acts by physically separating the two surfaces by a thin pressurised film which shears to accommodate their relative motion. Fluid film lubrication is generally the more effective of the two types.

2.4.1 Viscosity

The fundamental property of a fluid that enables it to separate two surfaces in fluid film lubrication is its viscosity. Simply put, viscosity is a measure of how difficult it is for the molecules of a fluid to move past one another. It defines the fluid’s resistance to shearing. The equivalent property in a solid is its elastic modulus which is the ratio of the shear stress to the shear strain. The viscosity is more complex, it involves the *rate* of shear strain. Conceptually the difference between shear in solids and fluids is one of elasticity. When a constant shear stress is applied to an elastic solid there is no motion after the initial deformation, or to put it another way, there is a finite amount of strain, this is not the case for a fluid. Consider a thin film of fluid sandwiched between two infinite and parallel plates. If opposite shear (tangential) forces are applied to the two plates points on each plate, initially above each other, will continue to move apart ad infinitum, i.e. there is no final distance between them. There is however a rate at which they move apart. If the two fluid boundaries move with their respective plate, i.e. there is no slip, then the rate at which the points move apart is the shear strain rate, this is also the velocity gradient across the film.

If, for a constant temperature and pressure, the shear stress in a fluid is proportional to the shear strain rate, then the fluid is known as *Newtonian*, and the *dynamic viscosity* is simply the constant of proportionality. Many fluids, often those with simple molecules, such as water, can be considered Newtonian. However there are classes of *non-Newtonian* fluids: where the shear stress is not proportional to the shear strain rate, i.e. the viscosity is not constant, even for a given temperature and pressure. Non-Newtonian fluids can exhibit *shear-thinning* behaviour whereby the viscosity falls with increasing rates of shear, or *shear-thickening* behaviour whereby it increases. Figure 2-5 illustrates the viscosity behaviour for these three categories of fluids:

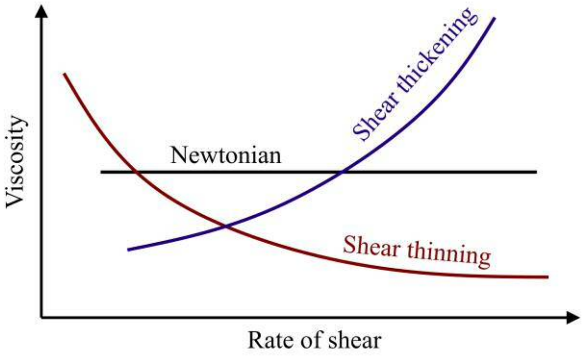


Figure 2-5 Viscosity and shear behaviour for three categories of fluids. Adapted from [4].

Shear-thinning behaviour can be beneficial for lubricating greases, as it reduces their viscosity when in operation, i.e. under motion, thus mitigating viscous friction. Researchers [5,6] have shown that when lubricating oils are subject to contact conditions found in gears they exhibit shear-thinning behaviour and empirical relationships have been developed which can describe this behaviour. The shear-thinning behaviour of lubricating oils may be beneficial or adverse depending on the specific contact conditions.

2.4.2 **Temperature, pressure and viscosity**

For Newtonian and non-Newtonian fluids both the temperature and the pressure can affect the viscosity. The viscosity of most liquids decreases with increasing temperature, however the scale of the decrease, depends on the specific liquid. For example, an increase in the temperature of an oil results in an extreme decrease in its viscosity whereas for water it has a comparatively negligible effect. The temperature-viscosity relationship for lubricating oils is non-linear with the decrease in viscosity typically resembling an exponential-like decay, i.e. a small change in temperature will have a much greater effect on the viscosity of a cold oil compared to a hot one. The temperature-viscosity characteristics of an oil are typically supplied by its manufacturer. Since it is the viscosity of an oil which enables it to separate to surfaces, the implication of its dependence on temperature is that the effectiveness of fluid film lubrication depends on the temperature of the contact, and by extension its friction, since friction generates heat. It is possible to conceptualise a self-reinforcing situation whereby frictional heating reduces the viscosity and thus effectiveness of a lubricant which then further increases the friction and so on.

In low-pressure systems it may be acceptable to assume that the effect of pressure on the viscosity of oil has a negligible effect. However this is not the case for non-conformal, Hertzian type, contacts. At the extremely high pressures generated by these, oil has *piezo-viscous* properties where the high pressure results in a dramatic increase in viscosity. As will be discussed later it is the piezo-viscous nature of oil that allows it to effectively lubricate the non-conformal contacts found in gears.

Barus [7] proposed a simple exponential model to describe the relationship between viscosity and *isothermal* variation in pressure. In this, the viscosity is proportional the exponent of the gauge pressure and a ‘pressure viscosity coefficient.’ This model can be satisfactory for rolling contacts, where there is often little frictional heating, however it is inaccurate in the case of sliding contacts, where frictional heating is significant. Later empirical relationships were developed by Roelands [8] and Houpert [9] which take into account both temperature and pressure variations. These relationships are suitable for calculating the viscosity of oils in realistic gear contacts.

2.5 **Hydrodynamic lubrication (HL)**

In most practical situations the viscosity of a liquid lubricant alone is not sufficient to enable it to separate two loaded surfaces indefinitely. If a heavy plate is placed onto oil that is pooled on a flat surface it will squeeze out the oil until it is resting on the surface. Although the viscous resistance of the oil prevents it from flowing out from under the plate instantaneously it only delays the inevitable contact. For the oil to *continually* separate two loaded surfaces it needs to be pressurised. This pressure can be supplied by an external source, as is the case for hydrostatic bearings, in which a pump provides a continuous supply of pressurised oil that separates the bearing surfaces. However, it is possible for the motion and geometry of contact surfaces to self-pressurise the oil and keep them separated. This type of lubrication is known as *hydrodynamic* and the principles that govern it are explained in the rest of this subsection.

To understand hydrodynamic lubrication two types of fluid flow need to be explained. The first is flow driven by a pressure gradient. An example of this is the flow up a drinking straw, an illustration of which is shown in Figure 2-6. The lungs create a low pressure at the outlet of the straw compared to the high (atmospheric) pressure at the inlet. This creates a negative pressure gradient in the straw as the drink flows from high to low pressure. If there is no slip at the boundaries, i.e. the fluid is stationary where it contacts the edges of the straw, then there must be a non-uniform velocity profile across a cross-section of the straw for the fluid to flow. Provided the flow is not turbulent then the velocity profile will be parabolic and the fastest flow will be at the centre-line of the straw. This type of pressure-induced flow is known as Poiseuille flow.

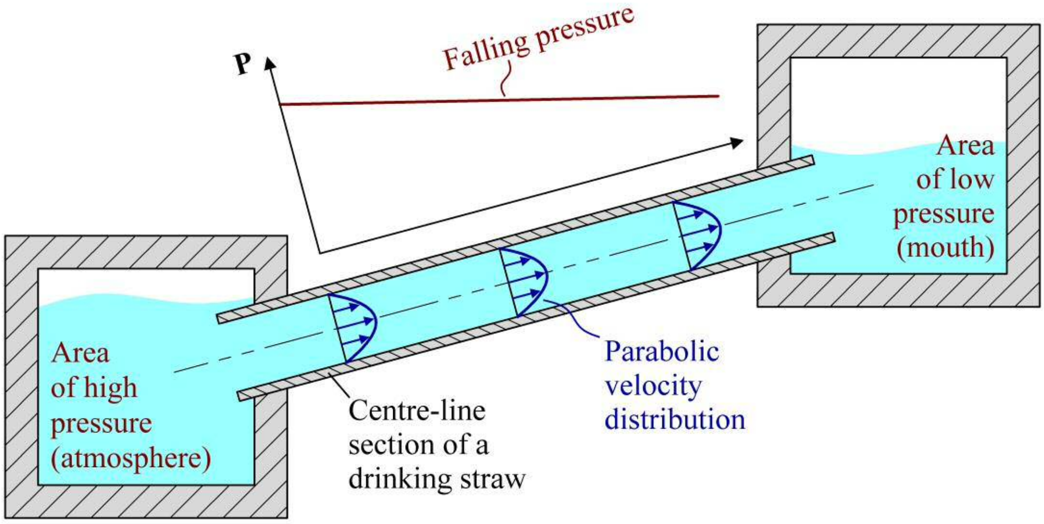


Figure 2-6 Flow through a drinking straw.

The second type of flow is that driven by motion, for example the flow of a fluid film separating two narrowly spaced and parallel surfaces that are sliding relative to one another. Figure 2-7 (a) shows this situation with a stationary top surface and a moving bottom surface. For conditions of zero slip the fluid velocity must be zero at the stationary surface, and must be equal to the velocity of the moving surface at that. If there is no pressure gradient along the direction of motion then the velocity profile will be triangular as shown. This motion-induced flow is known as Couette flow.

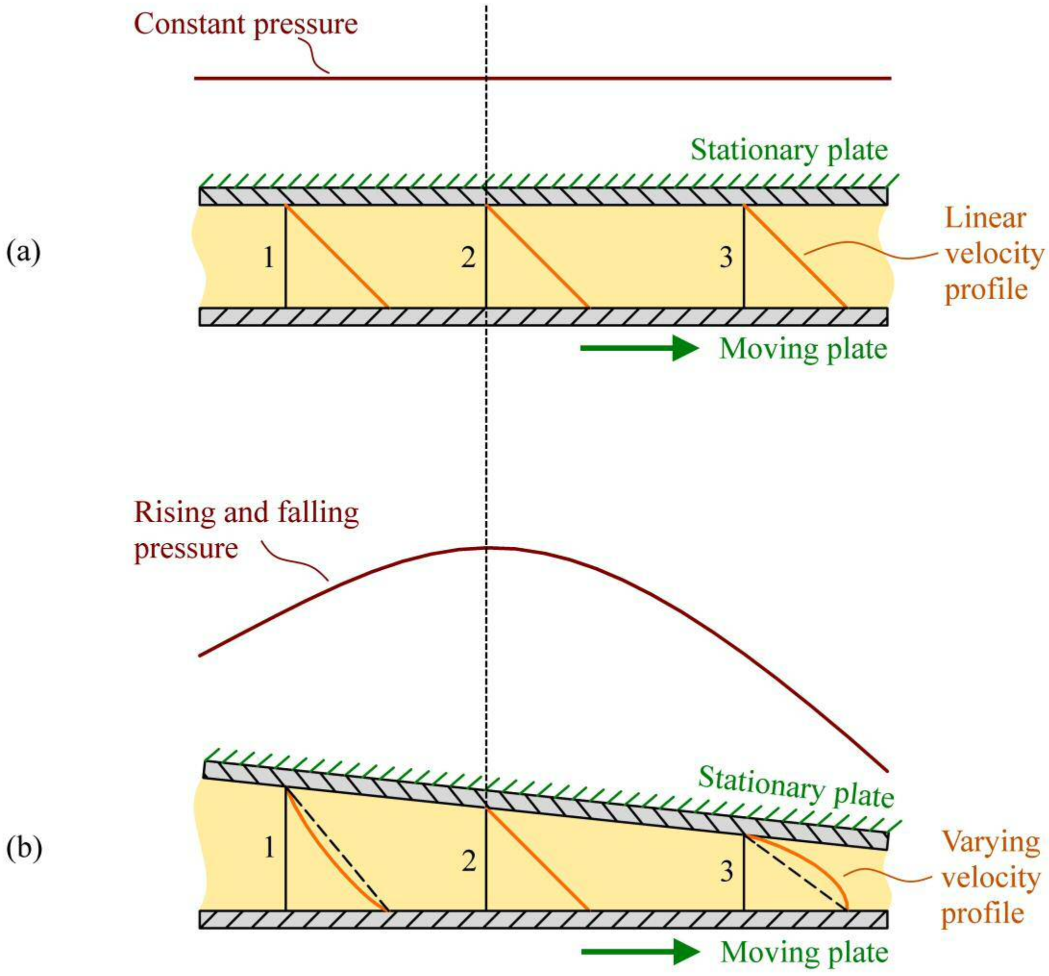


Figure 2-7 (a) Flow between parallel sliding surfaces. (b) Flow between inclined surfaces. Adapted from [10].

Now consider what happens when the stationary top surface is tilted slightly so that it is inclined and the fluid film becomes wedge shaped. This is shown in Figure 2-7 (b). At the pivot point, position 2, nothing changes so the triangular velocity profile remains unchanged. However upstream of this the film thickness increases and downstream it decreases and as a consequence the velocity profile must change in order to maintain a constant flow rate (the area under the velocity profile). Upstream, position 1, the velocity profile bulges backwards and downstream, position 2, it bulges forwards. A forward bulging profile indicates a falling pressure, as shown by the example of Poiseuille flow in Figure 2-6, conversely a backwards bulging profile indicates increasing pressure. Thus the wedge shape must generate a rising then falling pressure distribution as shown and this is capable of supporting load, i.e. separating surfaces in a contact. This, motion-and-wedge generated pressure is the principle behind hydrodynamic lubrication.

Mathematically the relationship between the pressure, velocity and shape of the hydrodynamic wedge is described by a partial differential equation first derived by Reynolds in 1886 [11]. The one dimensional form of the Reynolds equation as it applies to the example in

Figure 2-7 (b) can be easily determined by dimensional analysis [10]. Alternatively by making various simplifying assumptions (for example that the fluid is Newtonian) the three dimensional form of the equation, can be derived by consideration of: the forces acting on a moving element of fluid and the conservation of mass. With the Reynolds equation it is possible for engineers to design effective thrust bearings that provide fluid film lubrication based on the hydrodynamic principle. However it is gear, not thrust bearing, contacts that are of interest here.

The geometry of spur gear teeth in contact can be approximated as two rollers. Applying the hydrodynamic model to these, which necessitates a *rigid* surface assumption; and by selecting suitable inlet and outlet conditions, the pressure and film thickness distributions shown in Figure 2-8 result. However when calculating absolute values of pressure and film thickness for realistic real gear contacts the maximum pressure turns out to be impossibly high and the minimum film thickness impossibly low. In other words, the hydrodynamic model does not explain how gears are lubricated.

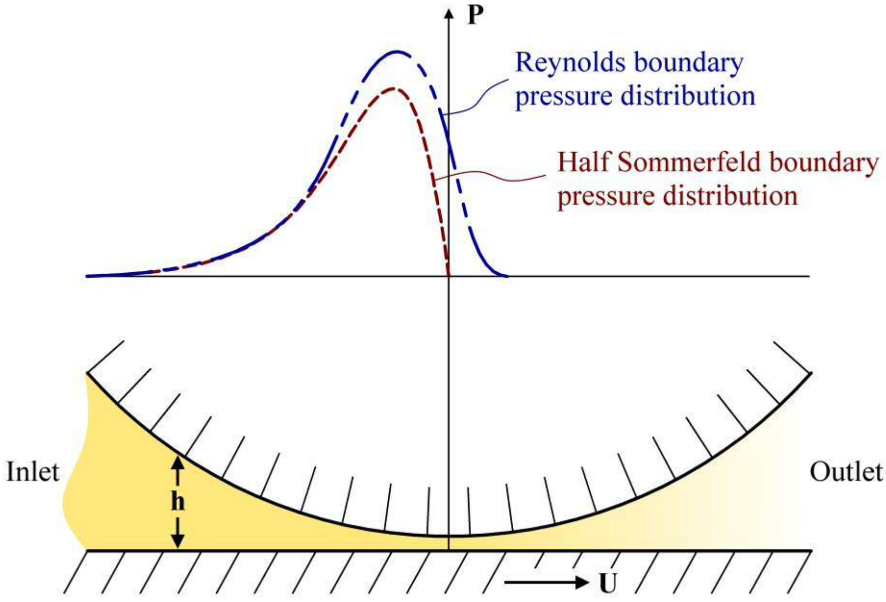


Figure 2-8 Hydrodynamic model of gear contact, where h is the film thickness, or gap, and U is the surface motion. Adapted from [3].

2.6 Elastohydrodynamic lubrication (EHL)

Before EHL theory was developed it was known that there must be some mechanism which provided better lubrication between gear teeth than was predicted by hydrodynamic theory. The latter predicts film thicknesses well below the scale of the machining marks typical on gear teeth, in addition to extremely high maximum pressures, conditions which would induce significant surface damage very quickly, and yet this does not occur. It was recognised by researchers in the early to mid-twentieth century that the hydrodynamic theory, when applied to gears, is inaccurate because of the incorrect assumption of rigid surfaces and an iso-viscous lubricant. This led to the development of EHL theory. The fundamental concept of this is that fluid film lubrication in a non-conformal contact results from the combined effects of a hydrodynamic oil film, elastic deformation of the contact surfaces and the piezo-viscous nature of the oil.

The first realistic EHL model to provide an approximate solution for the film thickness was developed by Ertel and Grubin and published in 1949 [4,12]. In this they made two important assumptions. First, that the elastic deformation of the surfaces is exactly the same as for a dry, Hertzian, contact, but that they are separated by a parallel oil film. Secondly that the hydrodynamic pressure rises sharply at the contact inlet and that within the parallel film section the pressure distribution is similar to Hertzian. The implication of these assumptions is that the parallel film thickness is determined by the geometry of the inlet surfaces. By introducing the piezo-viscous nature of oil into the Reynolds equation Ertel and Grubin were then able to derive an analytical expression for the film thickness for the case of a linear EHL contact. This, unlike the hydrodynamic model, produced plausible values for the film thickness in non-conformal contacts. Figure 2-9 illustrates Ertel and Grubin's EHL model.

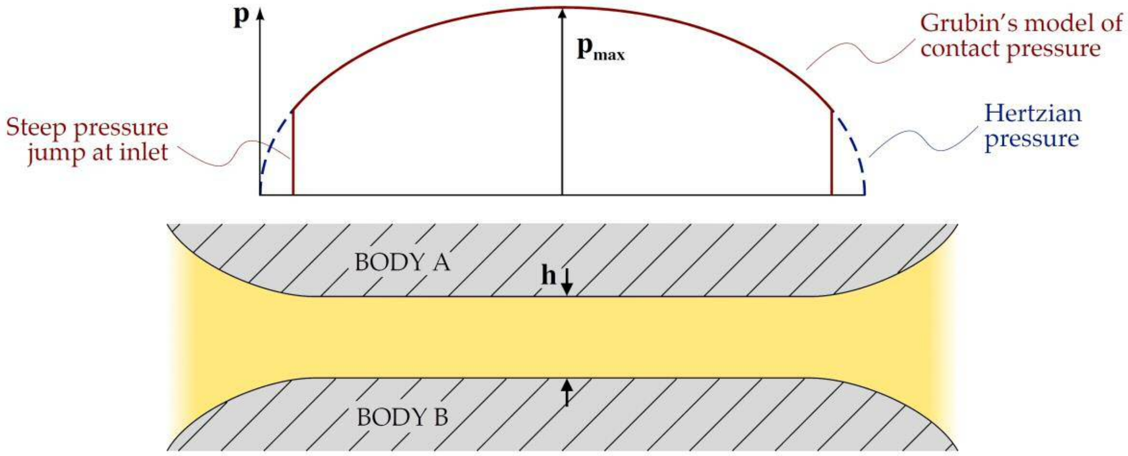


Figure 2-9 Ertel and Grubin's model of the film thickness and pressure distribution in an EHL line-contact. h is the film thickness and is the maximum Hertzian pressure. Adapted from [4].

The difficulty in developing a robust analytical solution for the EHL problem is that it involves complex interplay between fluid dynamics: the Reynolds equation; and solid mechanics: a suitable elasticity equation. Ertel and Grubin avoided this problem by first assuming Hertzian elastic deflection and then applying the Reynolds equation to this. This method essentially decouples the effects of the surface elasticity and the fluid dynamics. However, in practice the elastic deflection in an EHL contact can deviate significantly from Hertzian, particularly at the edges of the contact. As a consequence the analytical solutions do not describe true EHL. To produce more accurate results researchers turned to numerical (iterative) methods to solve the Reynolds and elasticity equations simultaneously.

The work of Dowson and Higginson was particularly important in the development of numerical methods to tackle the EHL problem. In 1959 they published a solution for an isothermal and heavily loaded linear contact [13]. (Other researchers had previously produced solutions for light loading but for typical engineering applications it is heavily loaded contacts that are of most interest). Dowson and Higginson’s 1959 solution, which included various simplifying assumptions such as an incompressible oil, presented the pressure and film thickness profiles for four load cases up to a centre-line pressure of approximately 0.5 GPa (considered a heavy load case at the time). They continued to develop their model and in 1962 presented a summary of their results which included the effects of different materials, speeds and a compressible oil [14]. From results such as these the pressure and film thickness characteristics of a typical linear EHL contact can be described, and they are illustrated in Figure 2-10.

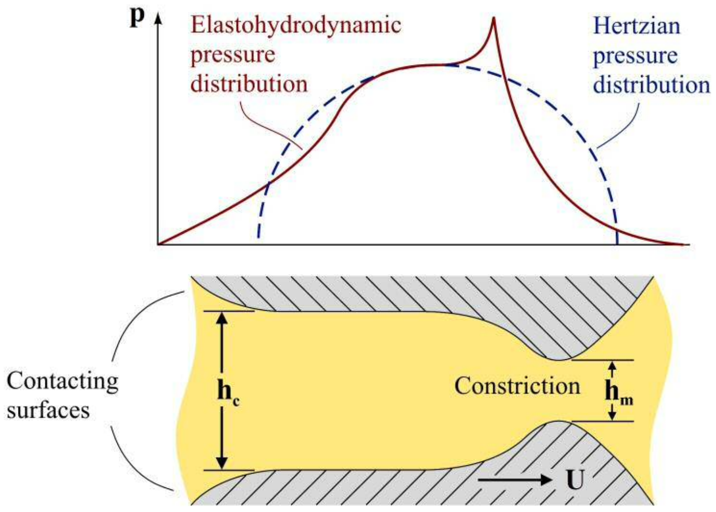


Figure 2-10 Typical pressure and film thickness characteristics of a linear EHL contact. is the minimum film thickness, is the central film thickness, and is the direction of entrainment (lubricant flow). Adapted from [4].

As Ertel and Grubin predicted, the pressure distribution is similar to Hertzian and the film thickness is parallel in the central region of the contact. However at the outlet, where the most significant deviation from Hertzian contact occurs, there is a constriction and immediately upstream of this there is a pressure spike. The existence of a constriction necessitates the use of two values to characterise the film thickness: the central film thickness, , which is the film thickness of the parallel section, and the minimum film thickness, which is the film thickness at the constriction. An additional deviation is that at the inlet the pressure increase is more gradual than in a Hertzian contact and as a consequence is of a slightly lower value. These inlet and outlet characteristics also occur in point EHL contacts in which case the area is elliptical and the constriction extends around the contact's trailing edge and sides to form a horseshoe shape (see Figure 2-11). It should be noted that the prominence of the inlet and outlet effects is greatly dependent on the speed of the contact. At very high speeds the constriction / pressure spike dominates the entire contact, which thus deviates significantly from a Hertzian shape, whereas at very low speeds the deviations become almost negligible.

Of course without some way to validate the accuracy of numerical EHL models they are only conjecture, so at this point it is worth mentioning some of the techniques that were, and still are, used to measure and observe real EHL contacts. The first of these are the electrical techniques. Provided the lubricant is an insulator and the contact surfaces are conductors, which is usually the case in real-world applications, then the resistance and capacitance across the contact will vary with the lubrication conditions. The resistance can indicate the breakdown of the lubricant film. However because the resistance across the contact varies in a somewhat binary manner, either very high or very low, it is difficult to use this to measure the film thickness. The capacitance can be used to estimate film thickness with reasonable accuracy. However it is more difficult to implement as the dielectric properties of oil vary with its temperature and pressure[1], and thus the technique requires significant calibration. Its use was pioneered by Crook who, in 1958, published measurements of lubricated rollers which showed, at high loads, film thicknesses greater than those predicted by simple hydrodynamic theory [15]. As the early numerical models were developed capacitance measurements were used to validate their predictions of the minimum film thickness, an example of which is the work by Dyson at el. [16].

One of the limitations of the electrical techniques is that they do not reveal anything about the film thickness distribution within an EHL contact. To measure this optical interferometry can be used. Its implementation requires that one of the contacting bodies is transparent. Light directed through this from behind will be split at the transparent contact surface with part being directly reflected back and part being reflected back by the opposing contact surface after transmission through the lubricating film. The interference pattern between these split beams, (which often adorn the office walls of tribology professors), produces a contour map of the film thickness and can provide absolute values if the process is calibrated. A crude form of this technique was first used by Kirk in 1962 [17] to study a point contact produced by crossed Perspex

---

[1] Thus the capacitance of the lubricating film will vary with independent changes to any of the following: film thickness, temperature or pressure.

cylinders. However owing to the low contact stiffness of Perspex he was not able to generate true EHL between them. Cameron and Gohar refined the technique, using a glass plate and a steel ball bearing to study a point contact, and in 1966 published a paper [18] using this method to validate a numerical simulation. Figure 2-11 shows a comparison of optical interferometry measurements made by Gohar [19] and the contour maps of a relatively modern numerical simulation of his experiments made by Jalali-Vahid et al. [20]. It can be seen that the numerical predication for the shape of the film is in close agreement to that observed using interferometry. Figure 2-11 also serves to illustrate the typical shape of the film in a point EHL contact. A characteristic horseshoe constriction is evident and it can be seen that at the higher speed the minimum film thickness has shifted outwards towards the sides of the contact.

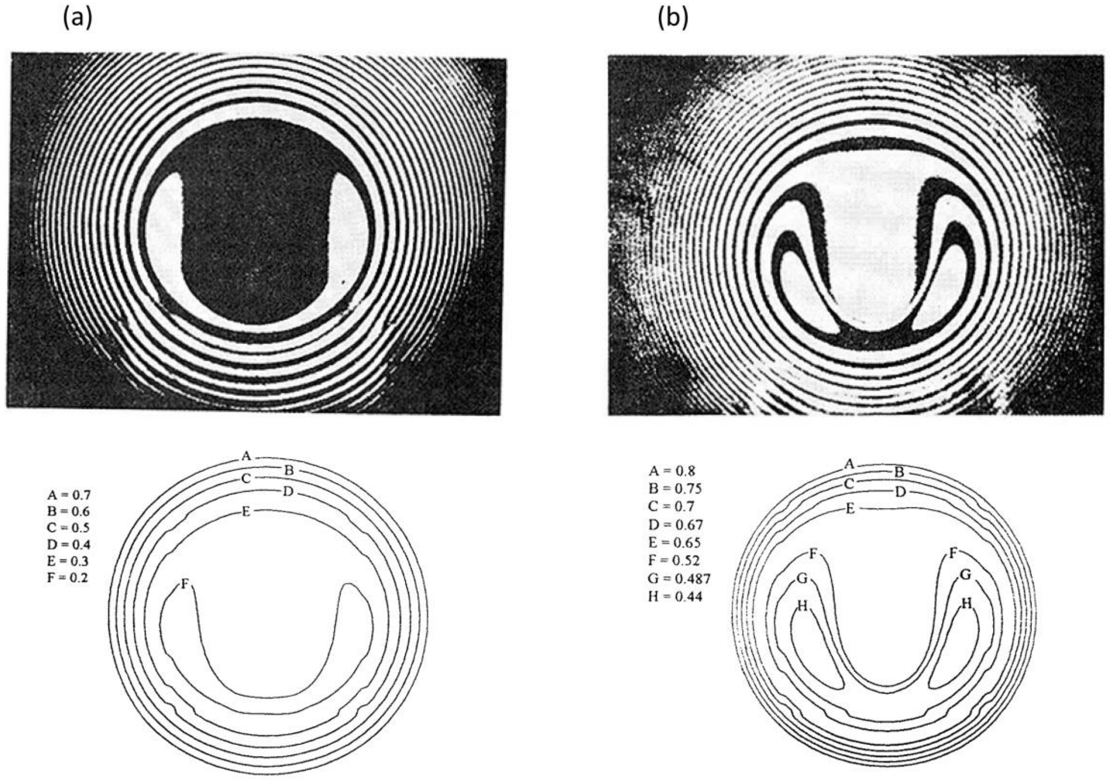


Figure 2-11 Comparison of optical interferometry measurements (top) and numerical simulations (bottom) of a spherical point contact at: (a) lower speed, and (b) higher speed. The results from the numerical simulation show the film thickness contour map. The entrainment direction is from the top to the bottom of the plots.

Both the electrical and optical techniques used to measure film thickness have significant limitations. The former requires isolation of the contacting bodies which is difficult to implement, and the latter requires one transparent, usually glass, body the material properties of which are not representative of those found in applied contacts.

In practical engineering applications the film thickness distribution is rarely required. The minimum or central film thickness value is sufficient to indicate the likelihood of a damaging lubrication regime. So with the development of numerical simulations efficient and robust enough to solve a wide range of control parameters, researchers began to characterise their relationship with the minimum film thickness. This led to the development of formulae that can be used to predict the minimum film thickness based on many practical permutations of the control parameters. The most important of these was developed by Hamrock and Dowson in a series of papers published in the 1970s [21–23]. Their formulae describe the relationship between the dimensionless film thickness and four dimensionless control parameters, defined as follows:

The dimensionless film thickness: — (1)

The dimensionless speed parameter: —— (2)

The dimensionless materials parameter: (3)

The dimensionless load parameter: —— (4)

The ellipticity parameter: – (5)

where:

is the minimum or central film thickness,
is the effective radius of the contact,
is the entrainment velocity (the mean of the velocity of each surface),
is the viscosity of the lubricant at atmospheric pressure,
is the reduced elastic modulus (the effective modulus of both bodies combined),
is the pressure-viscosity coefficient of the lubricant,
is the contact load,
is the semi major dimension of the Hertzian contact ellipse,
is the semi minor dimension of the Hertzian contact ellipse.

Hamrock and Dowson’s formula applies to point EHL contacts with the constraints that they are isothermal, fully flooded[1], and the surface motion is in the direction of the minor axis of the contact ellipse. The formula was derived by studying the effects of variations in all four dimensionless control parameters, one at a time, using a total of 34 simulations. This resulted in correlations between the film thickness and each control parameter which were assembled into a single formula. In 1985 Chittenden at el. [24,25] extended the work of Hamrock and Dowson to cover a contact ellipse in any orientation with respect to the direction of motion. This was based

---

[1] The ‘fully flooded’ condition implies that there is a continuous supply of oil at the inlet and outlet of the contact and thus the possibility of starvation is negated.

on the results of 106 numerical solutions. For the contact studied in this thesis the surface motion is in the direction of the minor axis of the ellipse and for this orientation the following formulae, apply [25]:

— (6)

— (7)

where:

is the dimensionless minimum film thickness,
is the dimensionless central film thickness,
is the major effective radii of curvature of the contact surfaces,
is the minor effective radii of curvature of the contact surfaces,
, and are the non-dimensional groups that represent the influence of velocity, materials and load.

In these formulae, sometimes termed the CDDT[1] formulae, the ellipticity parameter has been reformulated so that it is in terms of the surface geometry rather than the Hertzian contact dimensions; this removes the need to perform Hertzian analysis. The formulae reveal the something of the relative influence of each of the dimensionless control parameters. Of particular note is the small and negative exponent of the load parameter. This shows, perhaps against intuition, that the load has only a very weak effect on the film thickness. Because the geometry and materials in a real contact are almost always fixed the practical implication of the weak load dependence is that, for isothermal conditions, the film thickness is only varies significantly with speed. However, few real contacts are isothermal, and in practice the film thickness varies significantly with both the speed and temperature.

[1] 'CDDT' being an abbreviation of the authors names: Chittenden, Dowson, Dunn and Taylor.

2.7 **Rough surfaces**

At this point it is worth considering why the minimum film thickness in an EHL contact is so important. It is primarily because of surface roughness. For perfectly smooth surfaces any non-zero film thickness will provide full film lubrication; for rough surfaces this is not the case. All real surfaces are rough at some scale, and although there may be situations where this scale is such that its effect on the lubrication conditions can be ignored, this is not the case for gears. In this section the nature and measurement of rough surfaces is introduced.

Surface measurements are typically made using a profilometer, the principle of which is simple. A stylus attached to a pivoting arm is drawn in a straight line across a surface. The pivot allows the stylus to deflect up and down, tracing the height of the surface as it moves. The rate of traverse is kept constant and the height is sampled at uniform time intervals. From this a 2D profile of the surface, its height vs. distance, is generated. By use of a very hard and fine tip, typically a diamond cone with a tip radius of a few microns, the extremely fine detail of the roughness topography can be resolved. Pesudo-3D measurements can be made by incrementing the surface position, taking repeated profiles and then stacking them up. The pseudo-3D measurements usually only have fine resolution in one direction due to the time constraints of taking multiple profiles. Recently optical techniques have been developed that allow true 3D topography measurements to be made [26], however it was not possible to utilised these in this thesis. The machining / finishing processes used in manufacturing often result in a surface with an anisotropic roughness. In this case a 2D profile measurement across the dominant direction of the roughness can be sufficient to characterise it.

Real surfaces tend to have a topography that varies over a wide range of scales. In engineering, surface features are described, in decreasing scale, as *form*, *waviness* or *roughness*. The form, describes the basic shape, e.g. cylindrical, spherical etc.; the waviness describes how much the form deviations from ideal, i.e. the tolerance; and the roughness describes the fine texture, or finish of the surface. From a mathematical perspective the three parameters simply represent surface features of different, and decreasing, wavelengths. The form can be obtained by a low-order polynomial fit and the waviness and roughness by digital filtering, typically a Gaussian filter [27] is used. Figure 2-12 illustrates the form, waviness and roughness of a simulated surface. In reality the scale of features on a real surface, i.e. the spectral content of the profile, tends to vary continuously so there may be no obvious cut-off which can be used to distinguish between waviness and roughness features. In practice there are several standard values of cut-off to choose from with the choice dictated by the purpose of the measurement.

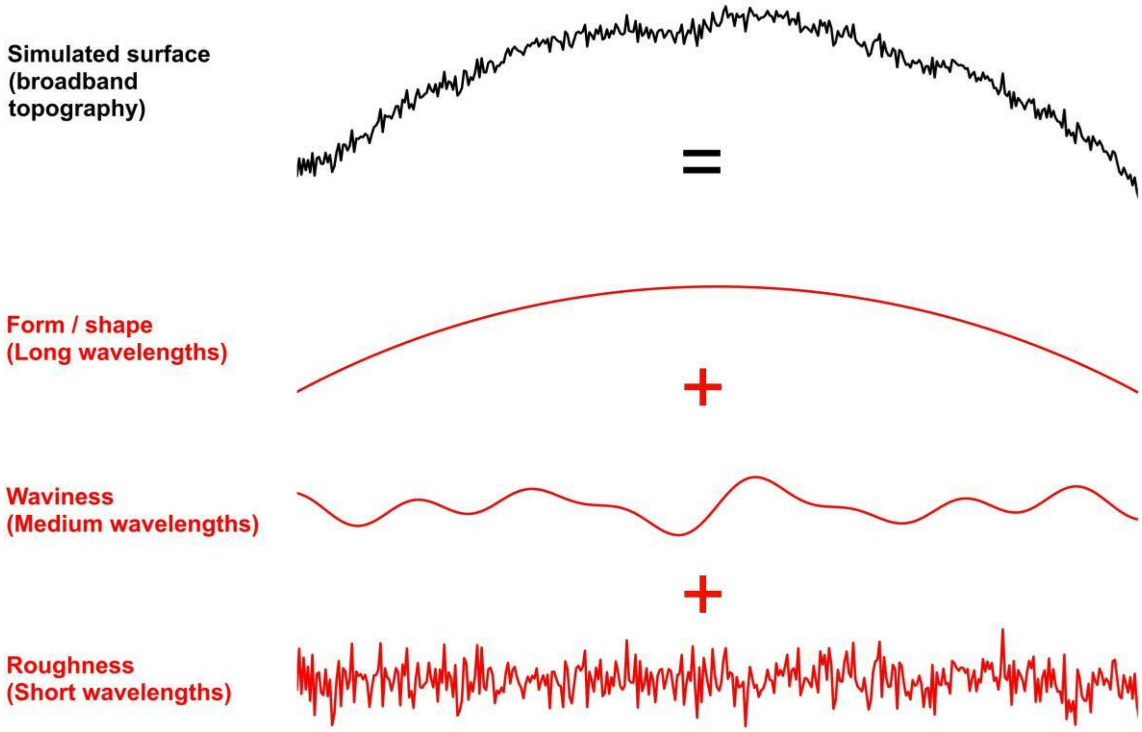


Figure 2-12 Form, waviness and roughness of a simulated surface

Typically when plotting profile measurements different length scales are used for the height and position axes, particularly when considering the roughness. This spatial distortion is necessary to present a reasonable length of the surface but it can give the impression that the roughness features are jagged with steep slopes. This is not the case, typical manufacturing results in roughness features with shallow slopes. Figure 2-13 shows the roughness of a real surface using true and distorted length scales.

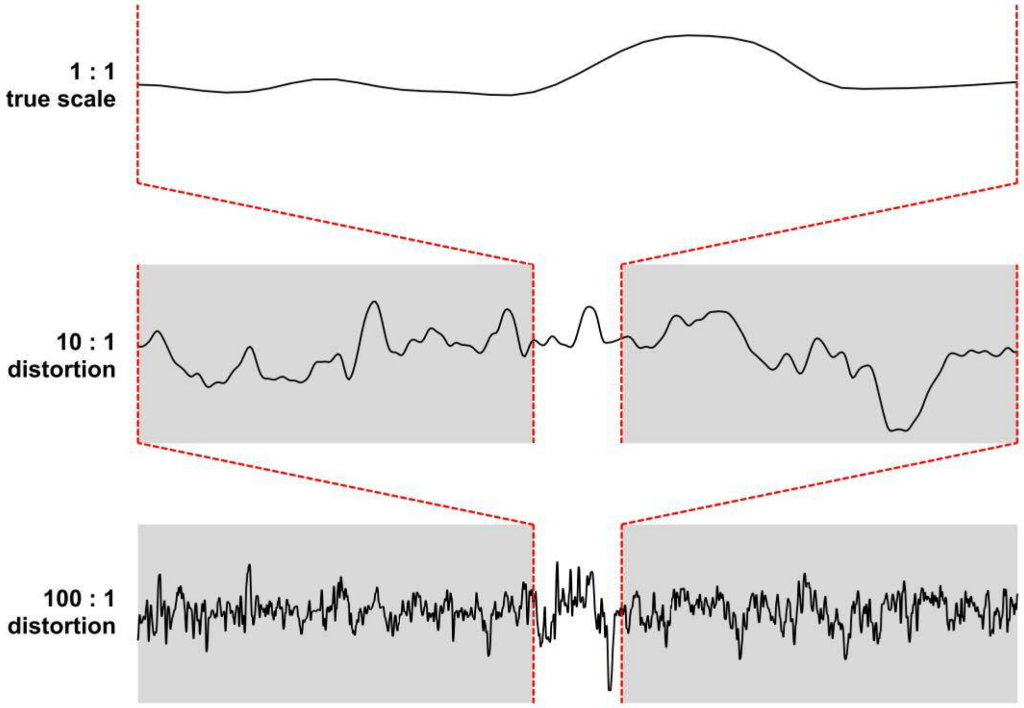


Figure 2-13 The roughness of a real surface plotted using true and distorted length scales.

It is cumbersome to work with roughness profiles directly, so statistical parameters are often used to represent their essential characteristics. These allow engineers to easily specify and compare different types of surface finish. To compare profiles a common datum is desirable however most engineering surfaces do not have an obvious roughness feature to use as such, (in contrast to the Earth for example, where sea level is the obvious choice). For this reason the mean of the measured heights is almost always used as the datum. This can cause uncertainty when comparing measurements made at different surface positions or after surface modification as it cannot be assumed that their mean lines are at the same height. The most commonly used roughness parameters are listed below with the formulae, in discrete form, used to calculate them:

The maximum height

The minimum height

The roughness average:

The root mean square (RMS) roughness:

The skewness:

The kurtosis:

where:

is the number of samples.

is the height.

The and the indicate the scale of the roughness features; the skewness indicates how symmetrical the profile is about the mean line; and the kurtosis represents how outlier prone it is, i.e. the probability of extreme spikes in height or depth.

2.8 Rough surface lubrication regimes

Now that roughness has been introduced its effects on lubrication can be discussed. In Section 2.4 two fundamental types of lubrication were introduced: boundary and fluid film. They cannot coexist in the hypothetical contact between two perfectly smooth surfaces. However in contact between rough surfaces it is possible for both to occur simultaneously, this is known as mixed lubrication. Figure 2-14 illustrates the different lubrication regimes that may occur in rough surface contacts.

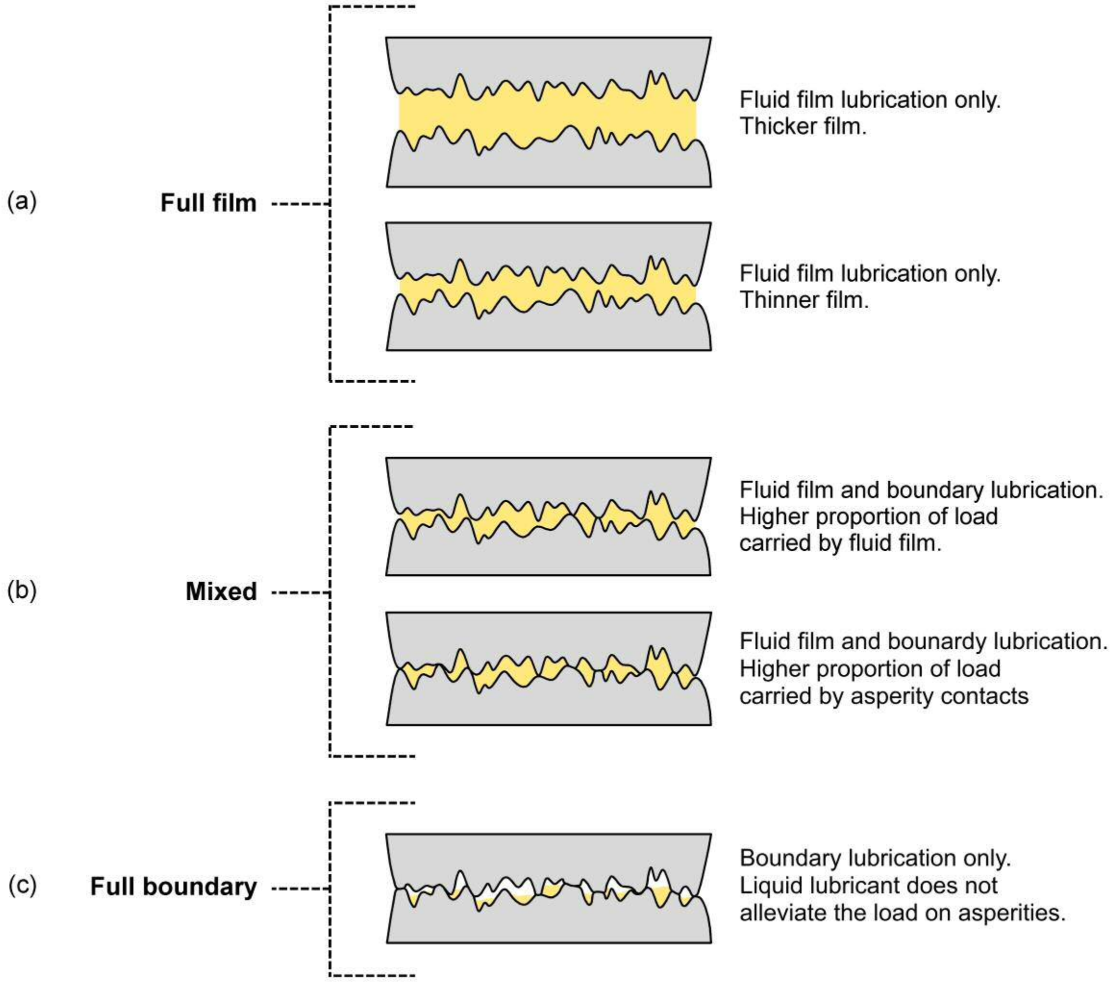


Figure 2-14 Types of rough surface lubrication: (a) full film, (b) mixed and (c) full boundary.

In Figure 2-14 (a) the lubricant film is thick enough to completely separate the surfaces, this is the 'full film' regime. In this the contact load is carried completely by the pressurised lubricant. Now consider what happens when the average film thickness is reduced. At some point the asperities of one surface will start to impinge on those of the other. Where this happens they deform to make a common interface in which boundary lubrication occurs. In this situation the contact load is shared between the pressurised lubricant and the direct asperity contacts. This is 'mixed' lubrication as shown in Figure 2-14 (b). As the average thickness of the film is reduced further more asperities will come into direct contact and there will come a point where the entire contact load is carried by direct asperity contact and the fluid only acts as a boundary lubricant.

This is 'full boundary' lubrication[1], as shown in Figure 2-14 (c). In metal to metal contacts the high material stiffness means that typical loads can be fully supported by relatively few direct asperity contacts; the roughness does not deform to anywhere near the extent that it becomes smooth, significant surface damage would occur before this.

Mixed lubrication does not represent a single state but a continuous spectrum of different proportions of fluid film to boundary lubrication. Knowledge of the specific proportion, is highly desirable as it can give an indication of the likely wear rate, or propensity for damage, of the contact surfaces. This proportion is typically represented by the *specific film thickness,* , which is the ratio of two parameters characterising the film thickness and surface roughness. In 1967 Tallin, when studying the failure modes of rolling contacts [28], proposed the following commonly used definition for the specific film thickness in EHL contacts:

(8)

where:

is the uniform (smooth) film thickness,
is the RMS roughness) of body 'A',
is the RMS roughness) of body 'B'.

The roughness parameter is simply the of both bodies combined, which can easily be measured or estimated from knowledge of the finishing process. The *uniform film thickness* is the film thickness of an equivalent contact, i.e. the same, speed, temperature, geometry etc., but one with perfectly smooth surfaces. It can be calculated using the CDDT formulae (Equations (6) or (7)) for the minimum or central film thickness. Many researchers choose to use the minimum film thickness but in this thesis the central film thickness is used as the pressure spike and constriction (see Figure 2-10) can be much less prominent, or even non-existent in mixed EHL, as will discussed in the next section.

It is important to note that the specific film thickness does not usually give a precise, or absolute, indication of the lubrication conditions. This is because of uncertainty associated with both the roughness and film thickness parameters. Firstly, the only partially characterises the roughness of a surface, for example it does not describe the shape of the roughness features or their orientation (if anisotropic) within the contact. Secondly, it is difficult to ensure accuracy in the calculation of the uniform film thickness. The often used CDDT formulae are based on a number of constraints, such as isothermal contact conditions and Newtonian rheology that may

---

[1] Typically this is referred to as just 'boundary' lubrication but the term 'full boundary' lubrication has been adopted to distinguish it from localised incidents of boundary lubrication that might occur in mixed lubrication.

not be met in practice. And so, unless compared against a direct measurement of the lubrication conditions, it is wise to treat the specific film thickness as an approximate or *relative* indicator of the lubrication conditions.

To suit the purpose of this thesis different lubrication regimes are categorised in terms of the proportion of fluid film to boundary (asperity contact) lubrication. However it is more common to categorise them based on the type of the contact and the friction associated with each. Figure 2-15 illustrates a typical example of this, from Halme and Andersson [29]. Different regimes are categorised as (full) boundary lubrication, EHL and HL, and their position on a Stribeck curve [30] is shown. A Stribeck curve presents the relationship between the coefficient of friction, , and the Hersey number: a dimensionless parameter which is the product of the fluid viscosity, fluid entrainment speed, and the reciprocal of the load. For a surface with stable roughness, the specific film thickness increases with the Hersey number. The transition from (full) boundary lubrication to EHL causes a dramatic decrease in friction, and the transition from EHL to HL a gradual increase. It is important to note that it is possible for an EHL contact (such as that in a ball bearing) to generate full film lubrication conditions as well as mixed. And conversely for a HL contact (such as that in a journal bearing) to generate mixed conditions as well as full film.

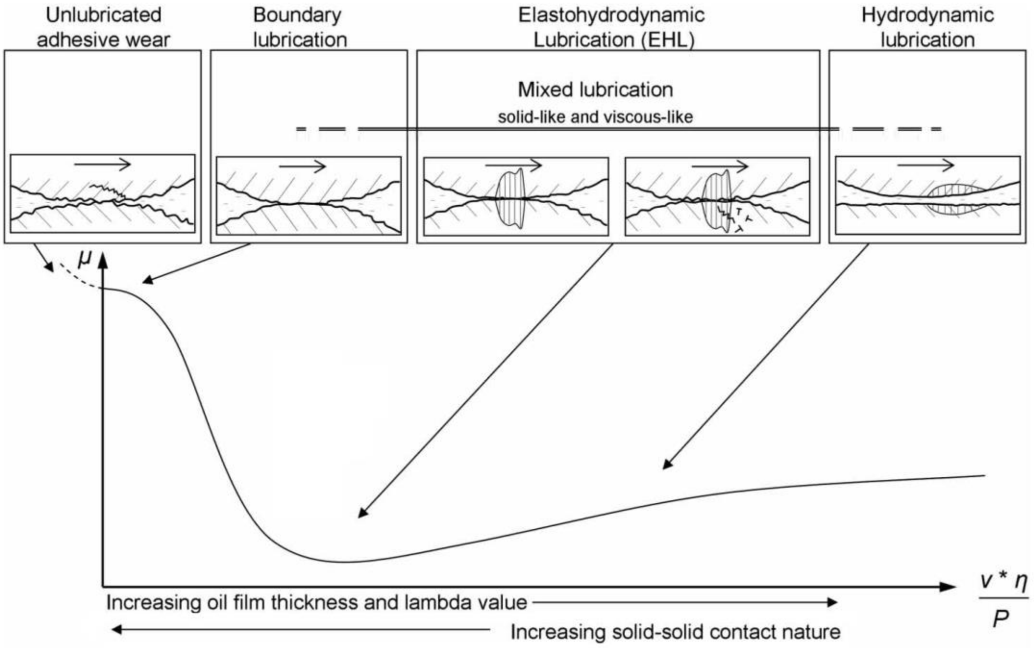


Figure 2-15 Schematic representation of a Stribeck curve and its relationship to different lubrication mechanisms. Adapted from [29].

Some authors have assigned explicit specific film thickness, , ranges to the different lubrication regimes but there is little consensus between them. For example Hamel et al. [31] identify as (full) boundary lubrication, 2 < < 10 as EHL, and as HL, whereas Halme and Andersson [29] identify as (full) boundary lubrication and do not give a value for the change from EHL to HL. The problem with stating an explicit range is that it can give a misleading impression of an easily defined, and precise, boundary between the different lubrication regimes which is not the case. Additionally, due to variations in the methods used to measure / calculate the specific film thickness, the ranges associated with different lubrication regimes may not be transferrable between different experiments / contacts.

In real applications the majority of EHL contacts operate, at least some of the time, in the mixed lubrication regime. From a wear perspective a full film lubrication regime is preferable as it precludes the possibility of wear caused by direct asperity contact. However to achieve full film lubrication under realistic operating constraints typically requires, either or both of: a very low roughness which can only be achieved by expensive finishing techniques, and or, a high viscosity lubricant which increases the contact friction. Thus the mixed EHL regime can represent a good compromise between wear, and efficiency. And for this reason researchers have spent considerable effort investigating it.

2.9 **Modelling rough surface EHL**

The inclusion of rough surfaces into any lubrication theory / simulation significantly increases its complexity. Even for the comparatively simple case of hydrodynamic lubrication, *deterministic*[1] analytical solutions become impractical. This is because the shape of a rough surface cannot be described by a simple deterministic function. However it can be described by a simple *stochastic* (probability density) function. Thus in the 1960s stochastic models for rough surface hydrodynamic lubrication were introduced by Tzeng and Saibel [32], and developed by Christensen [33]. Later Chow and Cheng [34] studied how a stochastic roughness affected the average film thickness of Ertel and Grubin's EHL model for lubricated rollers. Although attractive due to their relative ease of solution, the scope of these stochastic models proved limited. For one reason, there are no *accurate* analytical solutions to the smooth EHL problem with which to base a stochastic model on. As discussed in Section 2.6 numerical methods were required to provide accurate solutions and these were all deterministic. A more important reason is that there are phenomena associated with mixed, and near-mixed lubrication that cannot be easily modelled without consideration of the *exact* shape and position of contacting roughness features. These phenomena include: micro-EHL, a mechanism whereby close passing asperities within a contact are prevented from contacting due to localised EHL between them; and the effects of direct asperity contact on the fluid flow, such as trapped pockets of lubricant.

In contrast deterministic numerical simulations naturally lend themselves to the study of rough surface EHL. Conceptually it is not difficult to modify a numerical EHL simulation from smooth to rough: Roughness profiles can be stored as discrete data sets which can be 'fed' into a pre-computed smooth solution, and the governing equations can be adapted to cope with

---

[1] In a deterministic solution a fixed input leads to a fixed and repeatable output, there is no randomness.

phenomena that don't occur in smooth simulations such as direct asperity contact. However to actually run a rough EHL simulation requires vastly more computations than a smooth. For this reason the development of rough EHL simulations has run in tandem with the development of increasingly powerful computers and more efficient numerical formulations.

The earliest rough EHL simulations were developed in the 1980s[1]. The data processing and storage capability of computers at this time was too limited to run simulations with the high spatial resolution (a high number of 'sample' points) required study real roughness geometry, or the high temporal resolution (multiple 'sample' times) required to study moving roughness. Thus researchers began by studying the effects of simplistic and stationary roughness features. These were mostly a single 2D asperity or a 2D sinusoid, of a larger scale than is typical in real contacts and only in full film conditions. Goglia et al. [35,36] published results for a 2D (linear) EHL contact and soon after Karami et al. [37] and Lubrecht et al. [38] published results for a 3D (point) EHL contact. As might be suspected 3D simulations are more computationally demanding than 2D, but have the advantage that side effects and different roughness orientations can investigated. Real gear contacts tend to have a roughness with ridges and valleys that run transversely to (across) the direction of motion. And although 2D simulations inherently replicate this orientation they cannot investigate the side effects of such an orientation, which can be significant.

Although the early simulations were far from realistic (real roughness is not sinusoidal or singular, and a moving surface with a stationary roughness is an obvious impossibility) they nevertheless revealed important rough-EHL mechanisms such as the flattening of asperities by large pressure ripples in the lubricant film. An example of this is shown in Figure 2-16 using results published by Kweh et al. [39].

In the early 90s there was considerable concurrent research into different aspects of the general rough surface EHL problem. Some researchers developed 2D EHL simulations capable of investigating real, measured, but still stationary, roughness profiles [40,41]. Some continued to use the simplistic roughness features to study the effects of lubricant rheology [42,43]. And some began to produce transient simulations using the simplistic, but now moving, roughness features [44,45]. By the mid 90s researchers were incorporating multiple aspects of the rough-EHL problem to produce more realistic and general simulations. For example: isothermal simulations that were both transient and utilised real roughness profiles, were published by Ai et al. [46] for a linear contact, and Zhu et al. [47] for a point contact. Additionally a thermal simulation of real, but stationary, roughness was published by Xu et al. [48] for a point contact.

---

[1] Actually in 1973 Lee and Cheng [128], working for NASA, developed a transient simulation for a single asperity moving into a linear EHL contact. This was ahead of its time, particular for a transient consideration, however to do this they had to make the limiting assumption of Hertzian distribution away from the position of the asperity.

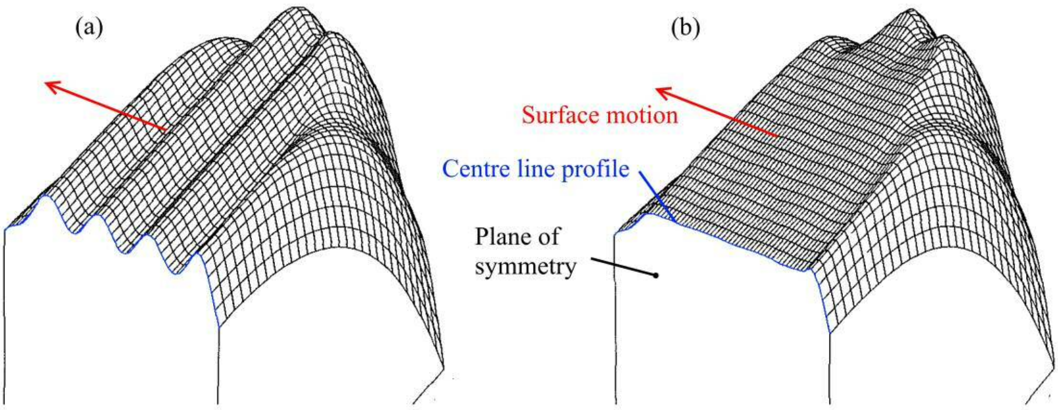


Figure 2-16 Simulation of a sinusoidal toughness in a point EHL contact. (a) Smooth surface solution with superimposed undeformed roughness. (b) Rough surface solution showing flattening of roughness in high pressure region. From Kweh et al. [39].

As an example of the increasing sophistication of rough-EHL simulations a result from Ai et al. [46] is presented in Figure 2-17. This shows a snapshot of an EHL simulation in which a real roughness profile was run, with a high degree of sliding, against an opposing smooth surface. The roughness has caused high frequency pressure fluctuations across the width of the contact and this behaviour is typical in rough EHL.

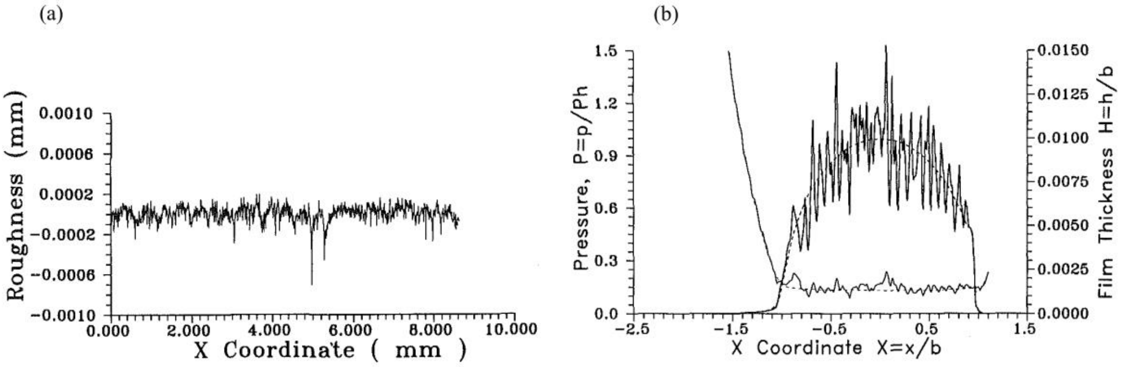


Figure 2-17 Transient 2D simulation of a real roughness profile. (a) Roughness measurement taken from a crankshaft. (b) Single time frame of the roughness in a sliding EHL simulation. From Ai et al. [46].

All the EHL simulations discussed so far have been limited to the study of thick films or fine roughness, where full film lubrication conditions are guaranteed. However it is the mixed lubrication regime that is of most practical interest to engineers. To model this, new rules had to be incorporated into numerical formulations in order to cope with incidents of boundary lubrication. Numerical solutions for mixed EHL were first published at the turn of the century.

In 2001 Zhu and Hu presented the first simulation of mixed lubrication in a point contact [49]. Their simulation modelled realistic 3D roughness, non-Newtonian rheology and the thermal effects of friction. However, its solution depended on the assumption that in very thin films the effect of pressure induced (Poiseuille) flow is negligible compared to that induced by surface motion (Couette flow). The suitability of this assumption has been debated in subsequent papers

[50,51] and the numerical formulation of Zhu and Hu continues to be used, and improved upon today.

At the about same time as Zhu and Hu published their 2001 results, researchers at Cardiff University were developing a different formulation, and this did not require the negation of pressure induced flow at thin films [52,53]. It was soon put to use for investigating the mixed lubrication conditions found in heavily loaded gears. The first results using this were published in 2001 for an isothermal linear contact with real roughness measurements and non-Newtonian rheology [54]. Later work extended the model for point contact geometry, and this was used, with 'extruded' 2D roughness measurements, to study the effects of side leakage in mixed lubrication regime [55,56]. Side leakage occurs when there are transverse roughness valleys. These encourage sideways flow of the lubricant from high to low pressure, i.e. out of the sides of the contact. Consequently the average film thickness can be significantly lower, and thus the probability of direct asperity contact higher, than in the centre. Figure 2-18 shows an example of this from Holmes et al. [56]. The use of an 'extruded' 2D roughness is somewhat unrepresentative of real surface topography. Even surfaces with a largely unidirectional roughness will have some variation in the other directions. To address this shortcoming recent work at Cardiff has utilised 3D roughness measurements [57].

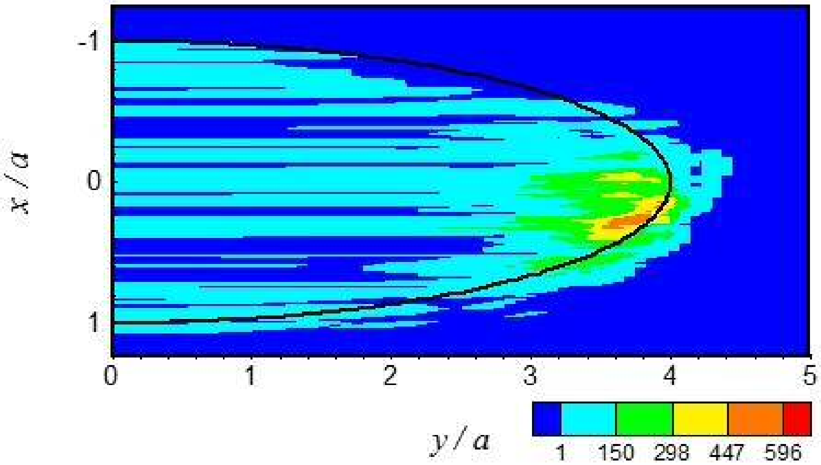


Figure 2-18 The position and frequency of direct asperity contacts in a point EHL simulation. The oil entrainment direction is from top to bottom. The black curve represents the dry Hertzian contact area. The colour scale represents the frequency of direct asperity contacts increasing from blue to red. From Holmes et al. [56].

The sophistication of numerical EHL models, and the computers required to run them, has increased dramatically over the past 40 years to the point where they can now be used to investigate in detail how mixed lubrication causes surface wear. However, it should be noted that simulating the mixed regime is still computationally demanding, even by modern standards. It can take hours or days to simulate just a few seconds or millimetres of mixed lubrication at a useful resolution. Thus investigations are typically limited to a few limited and simplified conditions. For example, attempting to accurately simulate the entire meshing cycle of a single rough gear tooth, (should one want to), would be considered a sizeable task. This would involve complexities such as edge, start, and end effects; continuously varying surface motion, geometry and thermal effects; and the list goes on. Perhaps the biggest limitation of the models that tend to be used for current mixed lubrication research is the assumption of a steady state roughness, i.e. no wear. For example, these treat the surfaces as infinitely elastic, and yet it has been shown that the pressures generated between close-passing / contacting asperities would be sufficient to plastically deform them in reality. Incorporating plastic behaviour into mixed EHL models is thus an active area of research.

2.10 Surface modification

This section gives a brief introduction to two types of surface modification that typically occur for gear teeth operating within the mixed regime. These are running-in and micro-pitting fatigue and they occur at different stages in the life of a contact.

2.10.1 Running-In

When moving contact is *first* initiated between two surfaces there may be a short period of operation before the friction and wear rate stabilise for the given conditions. This period is referred to as ‘running-in[1]’ and it consists of one or more *self-limiting* mechanisms which alter the friction and wear characteristics until they stabilise. An example of running-in can be easily observed when the pads of a disc brake, such as those found on a modern mountain bike, are changed. After fitting new pads braking performance is relatively poor until several aggressive braking procedures have occurred. The period while the braking performance increases is the running-in period.

There are various running-in mechanisms which, arguably, can be split into two categories, those relating to geometrical changes, and those relating to changes in the material properties. Geometrically induced running-in occurs due to non-conformity between the surfaces and it may resolve itself by modification of the impinging regions until the surfaces become more conformal. Changes in the material properties of the contact can occur due to the deposition of extremely thin coatings that form as a result of the initial contact. These can grow due to adhesive transfer of debris from a softer contact surface to a harder one; or due to chemical reactions with components of the lubricant, in which the coating is known as a ‘tribofilm’. If there is a change in the operating conditions of a contact then running-in may be re-initiated for the new conditions, but whether or not this will occur depends on the history of the contact.

This thesis is concerned with mixed EHL contacts between steel surfaces of similar hardness, i.e. gear-like contacts. These will have a running-in period because of direct asperity contacts. Researchers have shown that this involves both: the growth of tribofilms, and flattening

---

[1] The terms ‘running-in’ and ‘run-in’ are both used in this thesis. This is not to indicate a difference in meaning but a difference in tense, e.g. ‘that surface is running-in’ vs. ‘that surface has been run-in’.

of asperities. Considering tribofilms first. Many lubricants contain a small proportion of additives which are designed to form tribofilms that will enhance boundary lubrication of direct asperity contacts, i.e. reduce the friction and wear associated with them [58]. It has been observed that it can take time for some tribofilms to reach full effect as they require initiation, or conditioning, by the contact itself [59], and this will result in a period of running-in. As an example, consider ZDDPs, an important family of additives which form anti-wear tribofilms. Experiments using an Atomic Force Microscope (AFM) have shown that, at the nano-scale, the tribofilm produced by these grows at a rate that increases exponentially with either increasing pressure or temperature [60]. Of course conditions in an (AFM) are not very representative of those in mixed EHL and the practical factors that determine the growth and stability of tribofilms in the latter may be different.

Now consider running-in due to flattening of asperities. Figure 2-19 illustrates this by comparison of a real roughness profile measured before and after running-in. The two measurements have been aligned in space so that the differences can be easily observed. It can be seen that after running-in the asperities have become flattened but the valleys are unchanged. This shifts the height distribution from approximately symmetrical before running-in, to negatively skewed after. It has also caused a reduction in the (RMS roughness). Consequently, all other contact parameters being equal, it results in an increase in the specific film thickness, i.e. a decrease in the likelihood of direct asperity contact in the future.

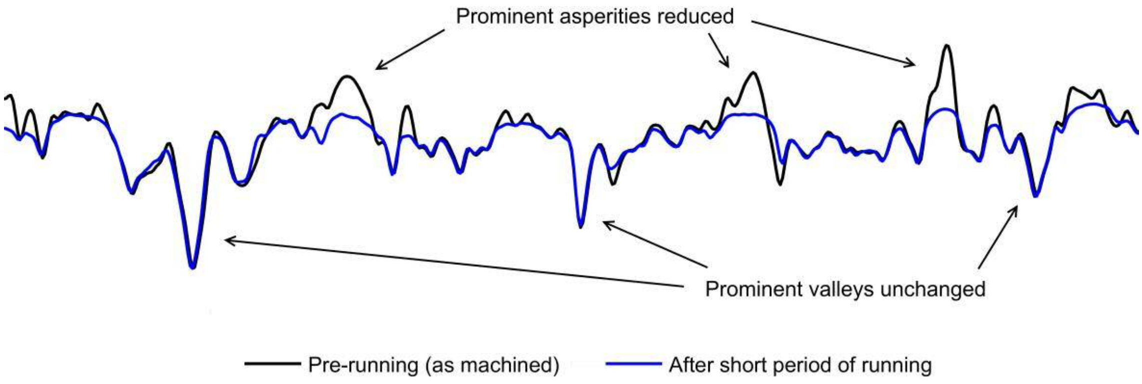


Figure 2-19 Comparison of the roughness of a section of a surface profile before running-in and after.

This geometric running-in has been thoroughly investigated for gear-like contacts using electrical resistance techniques and comparison of roughness profiles in the manner shown in Figure 2-19 [57,61]. Experiments revealed that the flattening procedure occurs and concludes relatively quickly, within less than a minute from the first application of a contact load. It is thought that due to the speed of the change, the process is predominantly one of plastic deformation, rather than material removal. The proposed mechanism is that extremely high compressive stresses induced in close passing or touching asperities causes them to plastically yield until the contact stress is reduced to below the yield point. The decrease in stress occurs because as overloaded asperities yield, the surfaces become more conformal and consequently the load becomes more evenly distributed over the EHL area.

2.10.2 **Micro-pitting fatigue**

If the surfaces of gear teeth that have been operating for long periods are observed it may be noted that patches of the surface have lost the reflectiveness typical of their finish immediately after running-in and instead have become dulled. Figure 2-20 shows an example of this. The difference in appearance indicates changes in the roughness topography, i.e. some form of wear. Under close scrutiny it is revealed that the dulled areas result from numerous ‘micro-pits’, that is pits with dimensions in the order of microns. The scale of these micro-pits is typically similar to the pre-existing topographical features of the surface and so they only initially affect the surface roughness rather than its form.

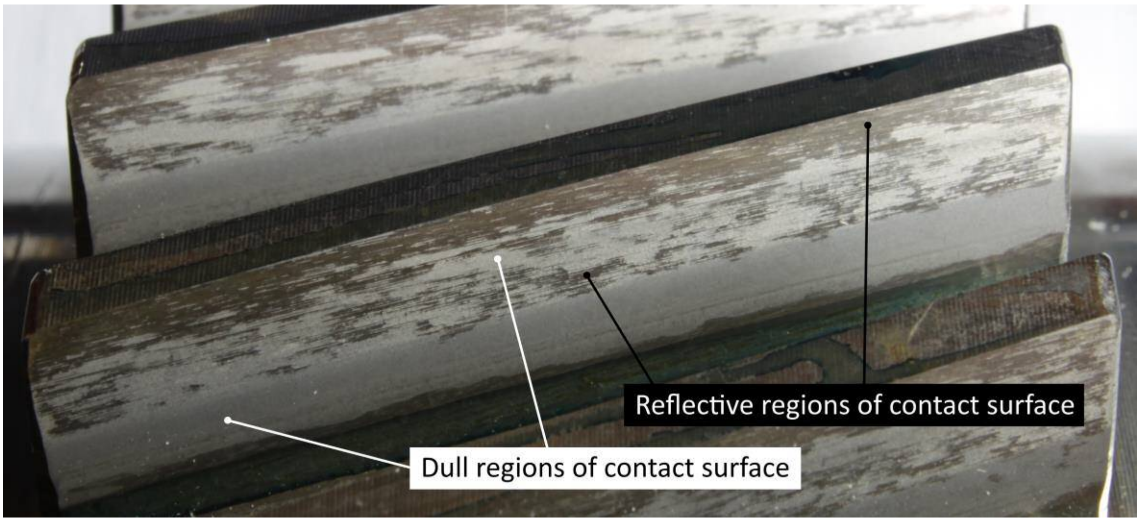


Figure 2-20 Photograph of used gear showing patchy surface finish.

Micro-pitting is thought to be caused by contact fatigue, a progressive process whereby cyclic loads initiate micro-cracking at or near the surface. These cracks propagate at shallow angles through the subsurface material until they cause break-away of the bulk surface material leaving behind micro-pits [62,63]. An example of micro-cracks and the resulting micro-pits can be seen in Figure 2-21. Many current researchers are attempting to develop models that predict when micro-pitting will occur however there are many operating parameters that can either encourage or inhibit it so this is not a simple task.

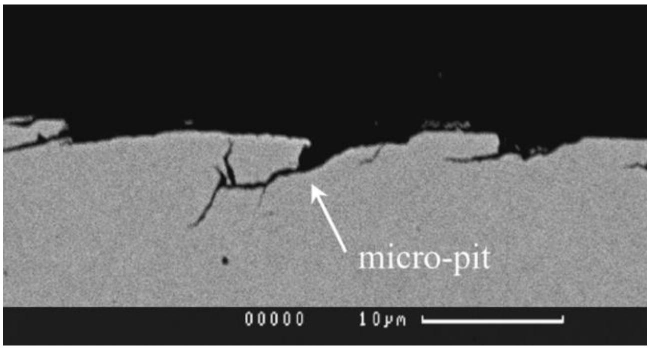


Figure 2-21 Cross-section of a micro-pitted surface. Reproduced from [64].

2.11 Acoustic Emission

This chapter is split into the following sections: Section 2.11.1 provides an introduction into AE and its measurement generally. Section 2.11.2 introduces some of the common signal processing techniques used to analyse the AE from moving contacts. Sections 2.11.3 and 2.11.4, which make up the bulk of this chapter, focus on recent research into the nature of AE from pure and applied moving contacts respectively. Finally, Section 2.11.5 provides a summary of the current understanding of the AE from moving contacts.

2.11.1 Introduction

The term 'Acoustic Emission (AE)' refers to a specific category of mechanical wave that propagates through a non-gaseous elastic medium. At its most broad definition, it might simply refer to any ultrasonic (high-frequency) wave whatever the source. However the definition is usually restricted to a class of waves that are *naturally* released by a material. There are various different mechanisms that might cause this release, but because the early practical applications of AE measurements were for Non-Destructive Testing (NDT), the most commonly described relate to the damage of materials due to overloading. During this, AE is released by various deformation related micro-mechanisms, such as crack growth and changes to the crystal structure in metals, and fibre breakage and dis-bonding in composites [65]. It is easy to see why, writing for the National Bureau of Standards in the 1980s, Wadley described AE as 'nature's ultrasound' [66]. In contrast to other NDT techniques, which typically use a probe to periodically detect if and/or where a structure has been damaged, AE measurement allows continual monitoring of the structure which will 'tell' you itself, by a burst of AE, when, and potentially where, damage occurs. Although typical examples of an AE source involve mechanisms that directly result from loading of solids, there are other mechanisms which are not directly related to this. For example it has been shown that liquid flow phenomenon such as cavitation and turbulence can cause AE, as well as liquefaction, solidification and solid-to-solid phase-transforms [65].

AE is nearly always measured using piezoelectric transducers. These convert the extremely minute motion caused by the propagating waves into an electrical signal, i.e, voltage vs. time. The transducers, referred to subsequently as sensors, typically consist of an encapsulated piezoelectric element which is bonded directly behind a thin faceplate. For the sensor to function the faceplate must be 'acoustically coupled' to the surface of the material being monitored so that the AE waves can propagate across the interface and into the piezoelectric element. This simply means that a good quality contact must be made which can be easily achieved by using a gap-filler such as a grease or an adhesive, and / or a clamping force, both of which help to prevent voids caused by surface roughness.

Sensors vary widely in terms of their sensitivity to different frequencies, and their size. Typically AE sensors operate somewhere within the 10 – 1000 kHz range and may have a broad or narrow-band spectral sensitivity. Across this band the sensitivity is not flat, and all sensors, even those described as 'broadband,' will have a resonant frequency, i.e. a frequency at which they are most sensitive. Because of unavoidable differences in the piezoelectric element of one sensor to the next, even sensors of same type will have slightly different spectral sensitivities. For this reason every off-the-self sensor is typically supplied with calibration data that describes its specific spectral sensitivity based on a standard test method. (Figure 3-12 in Section 3.4.5 shows an example of difference in spectral sensitivity for two sensors of the same type).

The variable spectral sensitivity of the sensors introduces an important, but sometimes overlooked, characteristic of AE measurements: that the *observed* spectral content of a specific AE mechanism is not unique, but varies depending on how it is measured. This is not just due to sensor differences, but is also because AE is not usually measured at the source, but after transmission through material, the geometry and natural frequency of which will affect the spectral content of the signal. It is also worth emphasising that 'AE' sensors are dumb devices, they will transduce any motion of a detectable frequency regardless of its source, i.e. whether it falls within the classic definition of AE or not.

An important characteristic of AE is that, because it has a generally high frequency, it attenuates very rapidly. This can be a positive or negative attribute depending on the practical application. In situations where it is necessary to monitor the AE from a large body, for example when attempting to monitor cracking in concrete structures, it is likely to be negative, as it means many sensors are required to cover an area of interest. However in situations where there is simultaneous AE from multiple sources at different locations, as might be the case in machinery with many moving parts, it can be positive. In this case the high attenuation may allow a source of interest to be preferentially measured if the sensor can be placed close to this but far from others.

An AE sensor provides a signal which needs to be interpreted. There are different methods depending on whether the AE occurs *continuously* or *discretely.* By discretely it is meant that the AE from a specific 'event' can be isolated from that of others. As an analogy consider the ripples in a pond caused by throwing stones in, where the impact of each stone is an 'event' and the resulting ripples are the corresponding AE. If the stones are thrown in sporadically it is easy to attribute a set of the ripples to a specific impact because there is a period of calm in-between. This is the case even if the point of impact is hidden from view as it can still be inferred by looking for the characteristic set of ripples elsewhere on the pond. Provided genuinely discrete AE events can be extracted from the sensor signal, which is typically done using a combination of amplitude thresholds and duration limits, then the AE can easily be described in terms of simple parameter measured for each event. For example, the peak amplitude, which for the pond analogy would be the highest peak in the set of ripples caused by one impact. Figure 2-22, from [67], shows the idealised waveform of a discrete AE event, along with commonly measured parameters. The AE event triggers a 'hit' when the waveform oscillations first cross an amplitude threshold, the hit ends when the oscillations no longer reach the threshold. If it is possible to differentiate between specific events, then advanced measurement techniques can be used to reveal useful information about each, such as its point of origin; which can done by comparing the times at which two or more sensors in differing locations detect a specific event.

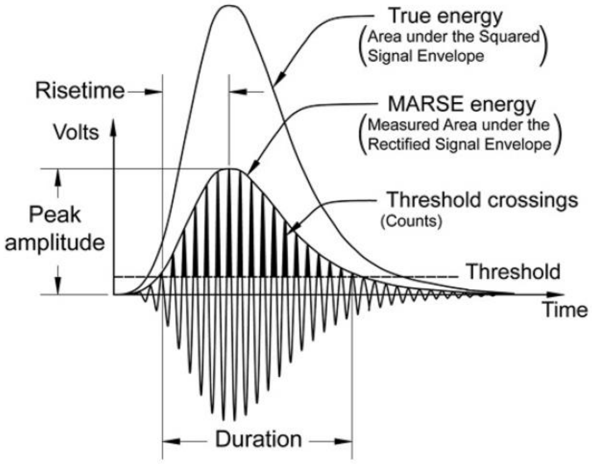


Figure 2-22 The waveform of an AE event along with commonly measured parameters. From [67].

Returning to the pond analogy, now consider what happens if the stones are thrown in in rapid succession, so that new sets of ripples are generated before others have a chance to spread out and die away. In this case it becomes impossible to study a set of ripples associated with one impact as the ripples from multiple impacts interfere with each other and there is no period of calm to separate them. This represents the case of continuous AE, in which case there are no individually identifiable events and the threshold based method of analysis is of less use. In reality the difference between discrete and continuous AE is not clear cut, for example it is possible for a signal to consist of discrete high amplitude AE events superimposed on continuous lower amplitude AE. However as will become clear in Sections 2.11.3 and 2.11.4, it is often beneficial to treat the AE from moving contacts as continuous.

2.11.2 **Signal processing techniques**

This section introduces some signal processing techniques that are commonly used in the analysis of AE from moving contacts. The techniques are only given a brief overview but are sufficiently common that more details can be found in most signal processing textbooks.

Windowing

Analysis of AE measurements often involves quantifying how specific signal parameters change over time, this can be achieved with a moving window. A window, in the mathematical sense, is a function that has a value of zero outside of a finite interval, the 'window'. To use a window function it is multiplied with the signal and this essentially returns only the part of the signal within the window. Any parameter, such as the RMS or peak amplitude, can then be calculated for this limited view of the signal. Incrementing the window along the signal and recalculating the parameter for each windowed position thus reveals how the parameter changes with time. The width of the window controls the extent to which high frequency variations are filtered out, with wider windows increasing the filtering.

If the window is rectangular, i.e. all coefficient of a window function are equal to one within its boundaries, then the window only segments the signal, it does not otherwise modify it. However a window can take any shape, and non-rectangular ones will modify (or distort) the view of the signal. At the edge of a rectangular window the signal is abruptly truncated and this can have undesirable consequences in some applications. Thus most commonly used windows are designed to taper off towards the edges.

Figure 2-23 shows an example of a continuous AE signal (from a measurement made for this thesis) and some of its parameters vs. time as calculated by moving window. For this example a Gaussian window was used. For the analysis of the AE from moving contacts the windowed RMS is particularly useful as will become clear in Sections 2.11.2 and 2.11.3. Typically, useful AE is of very high frequency, and so very high sample rates (typically MHz) are required to measure it. Thus a continuously recorded signal can very quickly become too large to analyse and store effectively. So for many long test procedures (long being greater than seconds in AE terms) the AE is only sampled periodically. This is analogous to rectangular windowing, only with no overlap, and large gaps between consecutive windows.

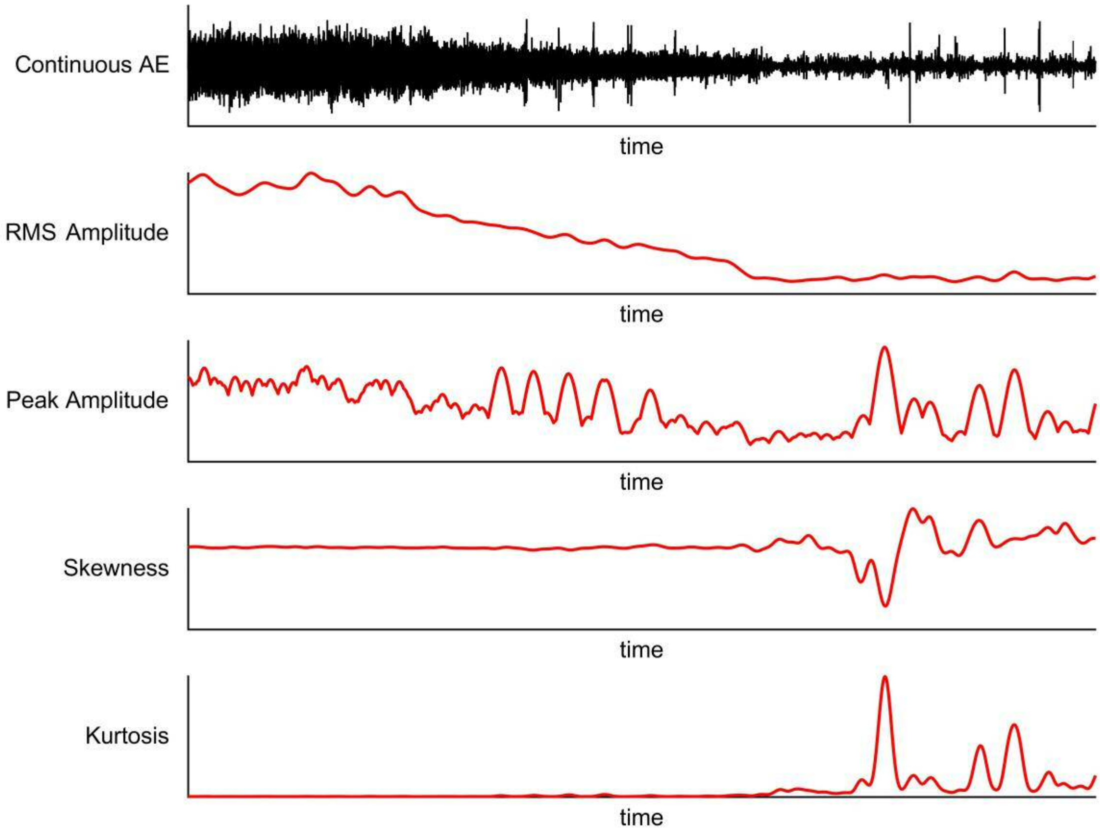


Figure 2-23 Example of a continuous AE signal and some signal parameters calculated over a moving window.

Fourier Transform

Pitch is a fundamental property of music, which allows discrimination between notes based on how 'high' or 'low' they sound, rather than their amplitude or duration. The pitch of a note is the perception of its waveform frequency. Just as in music, the frequency is an important attribute of AE. It is impractical to discern the frequency contact of most signals in the time domain: their amplitude vs. time form, so they are transformed into the frequency domain: their magnitude vs. frequency form.

In the early 1800s Joseph Fourier showed that any periodic waveform can be represented by the sum of an infinite series of sin and cos terms. The Fourier transform is based on this, it decomposes a signal into its constituent sinusoids, i.e. converts it to the frequency domain. There are different Fourier transform algorithms. The Discrete Fourier Transform (DFT) is used for discretely sampled and finite data, i.e. digital samples, and the extremely popular and important Fast Fourier Transform (FFT) is a more efficient (less computations) method of calculating the DFT.

Fourier and Fourier-like transforms are used not just for frequency (spectral) analysis, but also for other tasks such as the solution of differential equations and image compression, and they can often be applied as a standard tool, i.e. without expert knowledge of the mathematics involved. That said, when used for frequency analysis there are some, important characteristics of a DFT that need to be recognised and will affect how a signal is represented in the frequency domain.

Any DFT, returns the amplitude of the frequency components at regularly spaced and discrete points along the frequency spectrum. The amplitude of any components falling in-between these points is distributed to the neighbouring points thereby distorting their true amplitude, this is known as scalloping or the picket-fence effect. The frequency resolution depends on the width of the sample window, with a wider window increasing it.

Two other usually unwanted characteristics of a DFT are spectral leakage and spectral smearing, and these are caused by the finite nature of the sample window. Spectral leakage occurs for all components with a wavelength that isn't divisible by the width of the window. And in the frequency domain this causes the amplitude of its frequency to leak across the entire spectrum in a tapering fashion which may obscure other low amplitude frequencies. Spectral leakage can be mitigated by using a window shape that tapers to zero, or near zero, at its edge but this this comes at a cost of increased spectral smearing, whereby the amplitude of the component becomes wider, or smeared, at its locality.

Now consider a transient burst of high frequency activity (for instance that caused by a discrete AE event) in a long signal. In the time domain this high frequency activity has a position in time, i.e. it might occur midway through the signal. However in the frequency domain it has no position in time, it is not possible to identify when it occurred. This is a problem when analysing frequency content that changes with time. One simple solution to this is to repeatedly calculate the DFT using a narrow moving window. This operation is known as a Short-Time Fourier Transform (STFT) and it generates a three dimensional data set of window position (approximate time) vs. frequency vs. amplitude. The heat map (or surface plot) of the amplitude is known as a spectrogram and using this it is easy to see how the amplitude of specific frequencies changes with approximate time. As the width of the moving window narrows the temporal precision increases but the spectral precision decreases. In other words it is not possible to determine the precise time and frequency. (This is related to Hisenburg's Uncertainy Principle which states that the more precisely the position of a particle is known, the less precisely its momentum is known and visa versa).

Wavelet Transform

The Wavelet Transform is a relatively new type of transform that allows frequency analysis at a higher temporal resolution than a DFT or STFT. Wavelet Transforms have not been utilised in this thesis (partial due time constraints) but a brief description of them is presented here as they are often applied to the analysis of AE from moving contacts (as will be demonstrated in Sections 2.11.3 and 2.11.4). (For those who are interested, there is a very accessible series of presentations on the basics of Wavelet Transforms published by MathWorks [68]).

A wavelet is a transient, or finite, oscillation that rises from and decays to zero and has a zero mean. In a Fourier transform a signal is decomposed into a series of infinitely long sinewaves of different frequency, but being infinitely long: no position. In contrast a wavelet, being finite, has both a positional (time) parameter, it's 'shift', and also a frequency parameter, its 'scale'. The scale defines how stretched out in time (wide) it is and the shift refers to its position on the signal being transformed. In a Wavelet Transform, a particular shape of wavelet (there are many) is chosen and then wavelets of identical shape but different scale (frequency) are shifted along the signal and their forms compared (convolved, in mathematical terms) with that of the local signal. This returns a three dimensional data set of shift (time equivalent), scale (frequency equivalent), and magnitude. The heat map of a Wavelet Transform is known as a scalogram. This advantage of a scalogram over a spectrogram is the higher temporal resolution it affords.

2.11.3 **AE from pure moving contacts**

In the context of this thesis a 'pure' contact is a type of contact designed for ease of investigating fundamental tribological properties, rather than to replicate real-world systems. AE research using pure moving contacts is often focussed on the micro-mechanisms that cause it, and how the AE relates to parameters such as the coefficient of friction and rate of wear. In a pure moving contact typically one of the contact bodies is stationary which allows for easy instrumentation, and contact bodies are usually simple and easily varied so that the effects of different combinations of materials, surface finishes, and possibly geometries can be tested; these parameters tend to be fixed for real-world contacts. This section presents some of the recent research into the AE from pure moving contacts. Many of the published experiments have used a similar methodology, so only three papers are reviewed in detail, others are listed with less detail at the end of the section. The three examples presented in detail were selected based on the clarity with which they highlight AE characteristics which are relevant to this thesis, and / or their novelty.

The first example is from Asamene and Sundaresan [69] who used unusual kinematics to investigate the characteristics of AE from unlubricated steel-on-steel contacts. Their rig consisted of a long bar, fixed at one end, which was clamped partway down between two friction pads. The pads were reciprocated, using a sinusoidal displacement. At the start of each reciprocation, elastic deflection of the bar induced by the contact friction allowed a period of stick (static contact). As the displacement increased the elastic stress increased until it overcome the contact friction resulting in a period of slip (sliding contact). AE sensors were coupled directly to the bar and used to measure discrete events (see Figure 2-22). It was found that AE events predominantly occurred in the slip period as can be seen in Figure 2-24, in which slip periods are indicated by a constant axial load.

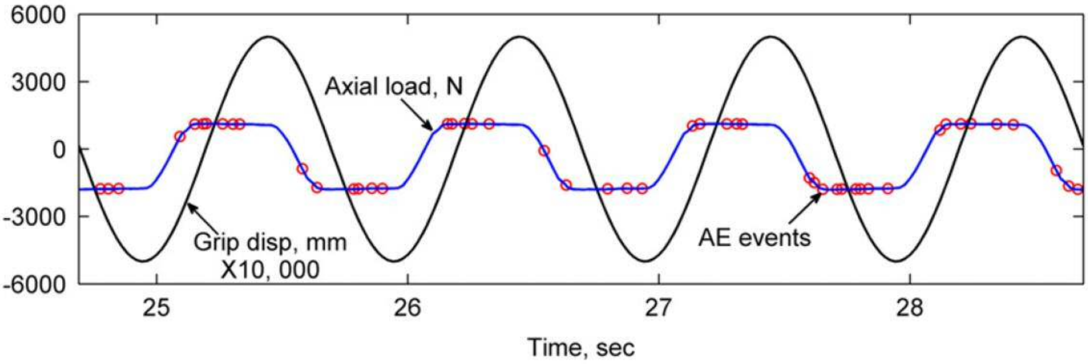


Figure 2-24 Distribution of AE events within a slip-stick cycle, from [69].

The authors assert that there were likely many more AE events in the slip periods than were observed and these were missed because they did not meet the threshold requirements either because they were of too low an amplitude, or because they occurred too soon after a preceding event and were thus merged with that. This serves to highlight the limitations of using a simple event count to measure the amount of AE, i.e. the count may vary depending on how an event is defined. In this case it would have been beneficial to treat the AE as continuous instead of, or as well as, discrete.

The cause of the AE was attributed to ‘collisions’ between opposing asperities on the steel bar which is logical as, given that the contact was unlubricated, these collisions are the raison d’etre of the contact. However the exact nature of the asperity collisions is not considered in great detail. For example, there is no detailed topographical information that might be used to reveal the geometry of the collisions or if they resulted in significant wear (although the and values did remain consistent over the tests).

Tests were conducted at two different cyclic frequencies (speeds) and pressures, for both smooth-on-rough and rough-on-rough surfaces. The spectral content of AE events resulting from these different conditions was compared using wavelet transforms. Subtle differences in the spectral content was noted, for example at the leading edge of the events involving smooth-on-rough surfaces 70% had a significant high frequency component, i.e. that above 500 kHz; whereas only 20% of events from rough-on-rough surfaces had a high frequency component. Figure 2-25 presents a comparison of wavelet diagrams which show this. As the authors describe, there are clearly differences in the leading edges, but these are just two events. It would have been interesting to see the spectral distribution averaged over a number of multiple events. They also found subtle differences when comparing the different pressures (2 and 4 MPa) and cyclic speeds

(0.5 and 1 Hz). These results show that the spectral content of the *observed* AE can change with operating conditions.

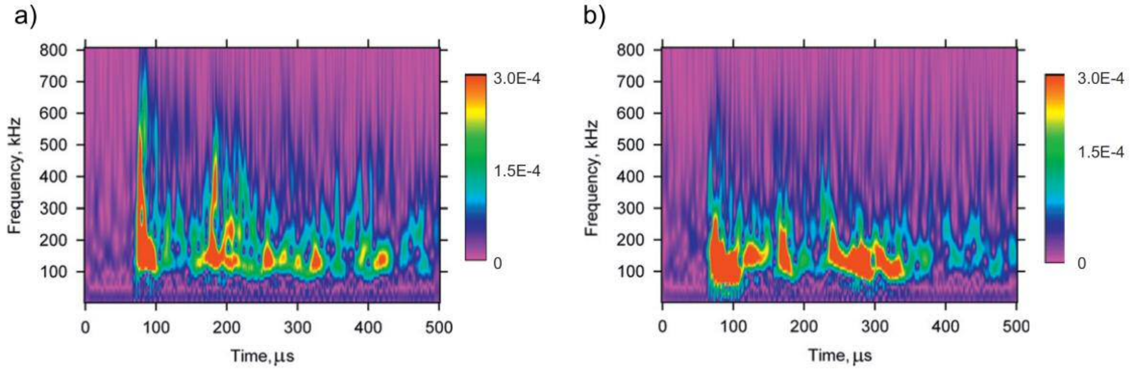


Figure 2-25 Wavelet diagrams (scalograms) of an AE event from a) a smooth-on-rough test and b) a rough-on-rough test [69].

The next example is from Hase et.al [70], who were concerned with distinguishing the AE from two different wear mechanisms, adhesive and abrasive. Their method consisted of measuring the AE during a single, and slow, pass of a lightly loaded and rounded pin over a flat counterface. To induce adhesive wear dry contact was made between pin and block specimens of different metal combinations (iron-iron, copper-iron and silver-iron). To induce abrasive wear lubricated contact was made between highly polished pins (< 50 nm) run against different grades of emery paper (400 and 800 grain size). After testing the worn surfaces were examined, using micrographs and atomic force microscopy, to verify the wear mechanisms were as expected. Prior to testing the authors used a pencil lead break test [71] on the pin tip as a reference for the resonance of the system. This identified significant peaks in the spectral content at frequencies below 200 kHz which were removed in subsequent testing using a 500 kHz high pass filter. They did not explicitly state why this was deemed necessary but presumably it was because a uniform, or flat, system response was considered beneficial when making spectral comparisons. However a potential downside of filtering out the resonance of the system is that any AE with a similar spectral content will be lost.

During the test an amplitude threshold was used to record discrete AE events (termed ‘bursts’ by the authors). For the adhesive contacts these only occurred in tests where the metal combinations resulted in significant wear debris, or transfer particles. The spectral contact of these events was characterised by a significant peak at 1.1 MHz. For the abrasive contacts discrete events also occurred, and with increasing amplitude for increasing grit, or grain, size. The spectral content of these was characterised by numerous distributed peaks in the range from 0.25 to 1 MHz. It is worth noting that below 0.5 MHz the ‘natural’ amplitude of these can be expected to have been both partially supressed by the high-pass filter but also possibly amplified by the system resonance so it is difficult to analyse the low end of the range. Nevertheless what is clear is that the spectral distribution associated with adhesive wear was different to that of abrasive wear. In both wear types the authors reason that the mechanism of AE is material fracture rather than plastic deformation because AE requires the *rapid* release of strain energy (elastically stored energy) and this release is more rapid in fracture than plastic deformation (flow). If this is the case it follows that the spectral differences must be due to differences in the characteristics of the fracture such as scale and duration; Figure 2-26 shows the authors' illustration of this. The discrete, or burst-like, nature of the AE is attributed to discrete instances of *significant* fracturing. This is plausible given the micro-scale of the contact area, and it follows that if the contact is scaled up the AE would become more continuous due to the greater probability of instantaneous asperity contact.

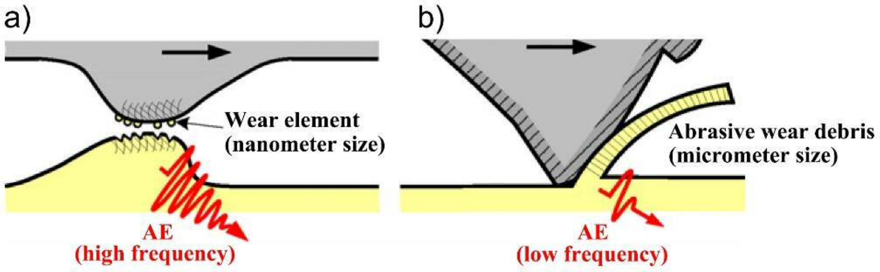


Figure 2-26 Illustration of the AE mechanism from a) adhesive wear and b) abrasive wear [70].

In their paper Hase et.al [70], attempt a comparison of the AE amplitude and frequency characteristics from deformation and fracture processes investigated in other experiments [72–77]. Figure 2-27 shows their resulting 'correlation map'. Hase et.al acknowledge that the magnitude scale is only an approximation because the measured AE amplitude is affected by the sensor characteristics, filtering, amplification and transmission path, and also that 'various factors' affect the measured frequencies. Despite these uncertainties they assert that the map makes it clear that the AE frequencies change depending on the wear mechanism (adhesive or abrasive) and that the AE spectrum from each is not dependent on the materials involved. However it is argued, that until all the parameters that affect the *observed* AE (as opposed to the AE at the source) have been quantified, comparisons of amplitude and frequency characteristics between different types of experiments must be treated with great caution. And if anything, the wide range of frequencies on the correlation map suggests that the frequency of AE from a particular type of contact cannot be easily predicted.

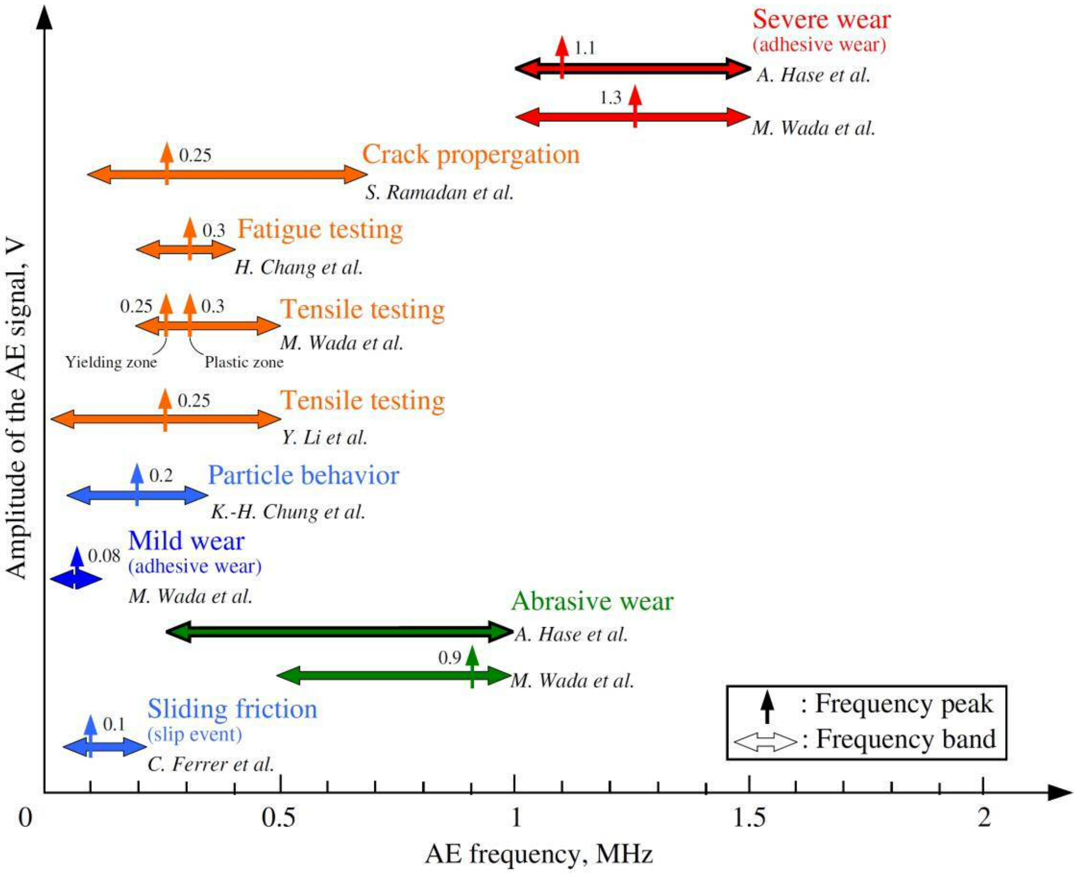


Figure 2-27 Correlation map of the AE frequency spectra for phenomena involving deformation and fracture [70].

The final detailed example involving a pure contact is from Benabdallah and Aguilar [78]. They tested an elliptical sliding contact using a stationary ball on a rotating cylinder. The contact was tested both dry and grease lubricated. Various different speeds were tested under a light load, i.e. with a small contact area. Unlike with the two previous examples the continuous AE was analysed. This was done in terms of the RMS amplitude (voltage) calculated over samples lasting a tenth of a second.

Figure 2-28 shows the relationship between the friction and the RMS AE for the dry tests at a range of different sliding speeds. Low friction coefficients, the six values below 0.6, only occurred at the lowest speed, which the authors attribute to the formation of a tribofilm. They conclude that there is a correlation between friction and AE that is *independent of speed*, however the robustness of the correlation as presented in Figure 2-28 is debatable (goodness of fit statistics were not given). The non-dependence of AE on speed does appear convincing for the top 4/5 of the speed range where there is no obvious systematic change in either the friction or the AE. The implication of these results is that the amplitude of AE isn't solely dependent on the *rate* of asperity interaction as this would lead to a speed dependence which in the case of dry contact is not evident.

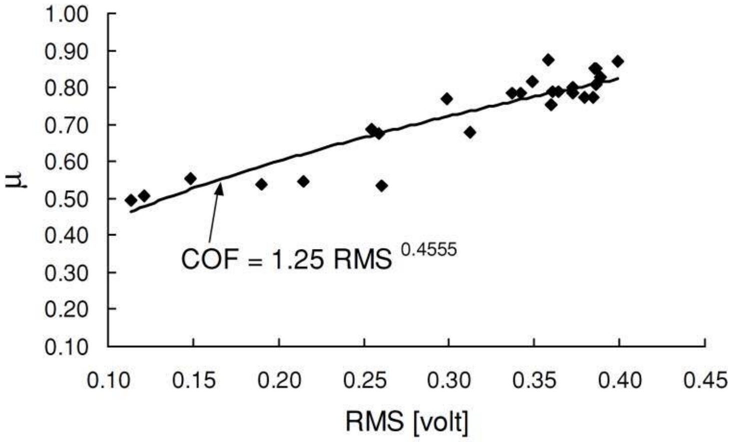


Figure 2-28 The AE RMS amplitude vs the coefficient of friction for dry contact at different speeds [78].

For their results with lubricated contact Benabdallah and Aguilar calculated the minimum film thickness for the different sliding speeds. The authors acknowledge that the calculation uses a simplified approach, it does not include possible starvation effects (the grease was not replenished), but they reason that because of the short test duration of each test, 120 s, and low load, 10 N, the calculation is still an acceptable estimate. There is no discussion related to temperature so presumably the calculation was conducted under the assumptions of isothermal conditions. Again this may be reasonable given the low load and short duration but it is not proved. The relationship between the minimum film thickness, and the 'intRMS', which is the integral of the RMS AE calculated over uniform sliding distances, is presented in Figure 2-29.

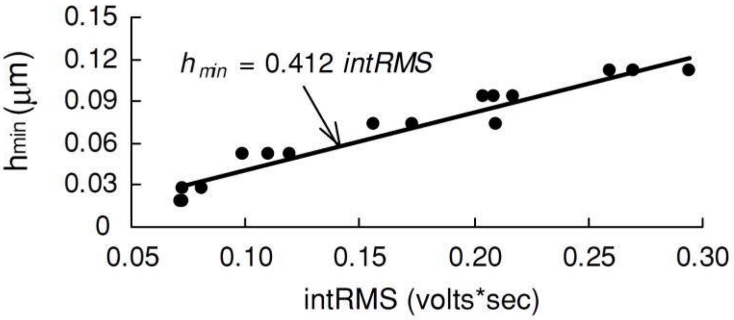


Figure 2-29 The AE intRMS vs the predicted minimum film thickness for grease lubricated contact at different speeds [78].

There is a definite, if somewhat imprecise, relationship between the intRMS and the predicted minimum film thickness. Benabdallah and Aguilar do not explicitly suggest the cause of this relationship. They do show, in a different figure, that the friction decreased with increasing sliding speed in a manner that suggests their contact was operating in the mixed lubrication regime, i.e. the amount of asperity contact decreased with increasing speed. The results of Figure 2-29 imply that the AE increased *despite* decreasing asperity contact.

Some other recent experiments into the AE from pure moving contacts are now listed (in chronological order):

Price et.al [79] investigated scuffing (a type of severe surface damage) and pitting using a four ball tribometer (a pyramid like arrangement, with one ball placed on the centre of a triangle created by three others). They used SEM to identify the wear type, and spectrograms to examine how the AE changed with time. They concluded that the onset of scuffing and pitting could be detected by changes in the AE.

Sun et.al [80] used both AE and the electrostatic charge to investigate the wear mechanisms associated with the unlubricated sliding contact of a bearing steel. They used a pin-on-disk geometry. Three different wear regimes were identified: running-in, delamination and oxidation. The characteristics of the AE RMS were found to be distinct for each, and it correlated well with the friction.

Wang and wood [81] monitored the growth and deterioration of a tribofilm, and eventual scuffing, in a lubricated bearing steel contact. (The growth of the tribofilm was validated using X-ray photoelectron spectroscopy). It was found that the AE could identify the deterioration of the tribofilm and the contact via its amplitude and frequency characteristics.

Wang and wood [82] also investigated the AE from silicon nitride coated steel in lubricated pin-on-disk tests. They found a strong relationship between the coefficient of friction and the AE RMS.

Nagata et.al [83] studied the bearing materials used in marine diesel engines (various tin and aluminium alloys and one with a resin overlay). A sleeve to plate tribometer was used and the AE, friction and temperature were measured during start-stop and seizure tests. The AE RMS was found to be a more sensitive indictor of surface condition than the coefficient of friction and temperature. It was found that the AE RMS could detect damage in the resin overlay before the onset of seizure.

Hase et.al [84] evaluated the relationship between burst-type (discrete) AE and the debris from adhesive wear in a minute contact. (Atomic force microscopy was used to examine it). They found a good correlation between the amplitude and duration of the AE waveform, and the quantity of debris on the worn surface. And the AE hit rate was proportional to the adhesion force between the two materials.

Tian et.al [85] investigated dry sliding between four metals (carbon steel, carbide, aluminium alloy and brass) using a ball on plate configuration. They measured the AE, coefficient of friction and contact resistance for different wear regimes which were identified using SEM. It was found that the AE RMS increased with speed for all metal pairs, but its response to load depended on the materials. It had both positive and negative correlations with the coefficient of friction depending on the speed and materials, this was attributed to the different wear regimes.

Saeidi et.al [86] monitored the AE from a lubricant starved, cast iron on steel, flat on flat, reciprocating contact. They developed an automatic classification algorithm that attempted to detect incipient scuffing. The classification was based on the Wavelet Packet Transform (a specific type of wavelet transform) of the AE signal. Once trained on a data set (i.e. given prior knowledge) the classification algorithm could correctly identify AE signals from just prior to scuffing with 88 % success, and from scuffing with 100 % success.

Moshkovich et.al [87] studied the correlation between friction and AE for pin on disk tests of nickel and silver against steel during the transition from EHL to full boundary lubrication. The different lubrication regimes were identified by SEM and by Stribeck curves. A strong relationship between the coefficient of friction and AE energy was observed. In the EHL regime the frequency content of the observed AE was similar for both material combinations, but in the full boundary regime, the frequency and amplitude characteristics differed between material combinations. The AE signals had characteristics associated with different scales of friction: larger contact ‘spots’ and smaller single interactions.

2.11.4 **AE from applied moving contacts**

In the context of thesis an ‘applied’ contact is one designed primarily to replicate a real-world system. As such they generally involve contact between functional bodies, i.e. actual gears or bearings. However the contact environment is likely to be partially unrepresentative, for example, free from sources of noise and modified for practicality of making AE measurements. Additionally the operating conditions are likely to be simplified, for example only examining state steady conditions and only varying parameters singly. Experiments on applied contacts are often concerned with the practicalities of using AE measurements for condition monitoring. For example, can the sensor still detect damage when placed on a housing rather than directly at the contact? And can AE from surface damage be distinguished from background sources. In this section some of the research into the AE from applied moving contacts is presented. The papers studied in detail have been chosen based on their relevance to this thesis (i.e. they are concerned with gears and mixed lubrication), and / or their novelty. Other examples are presented in less detail at the end of the section.

Considerable insight into the nature of AE from gear contacts has been made by various researchers at Cranfield University [31,88–103]. They have published a number of papers where the AE from spur and helical gear pairs was measured using an AE sensor bonded to the side of pinion (smaller) gear. These ‘Cranfield experiments’ have primarily involved studying the effects of operating conditions on the amplitude of continuous broadband AE, rather than on its spectral content. Some of their observations are now presented.

The first is that the amplitude of the continuous AE signal is prone to oscillation, or pulsing, at the meshing frequency of the gears. An example of this, from Tan and Mba [88,95], is shown in Figure 2-30, and they describe this characteristic waveform as the superimposition of ‘continuous’ and ‘burst’ type emissions.

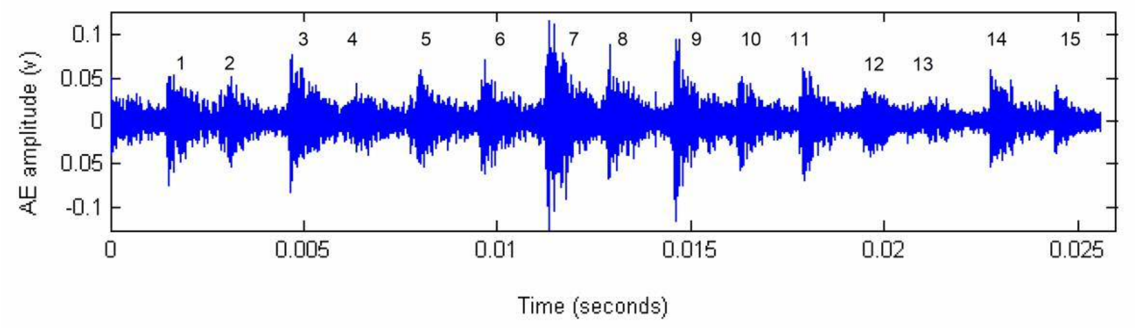


Figure 2-30 Example of the continuous AE during the meshing of fifteen spur tooth pairs [95].

They attribute the transient bursts to *rolling* asperity contact, which occurs at the pitch (mid) point of the meshing cycle, and the continuous emission to *sliding* asperity contact which occurs away from the pitch point. This explanation is plausible, but it has not been proved beyond doubt, and there remain unanswered questions. The most pressing of these is, why should the amplitude of AE amplitude of the rolling asperity contacts be greater than the sliding contacts? There are at least two possible explanations. Firstly it may be because the sliding to rolling ratio (SRR) *directly* affects the amplitude of AE from *any* instance of asperity contact. I.e. for asperity interactions of equivalent 'magnitude' and duration a rolling motion results in AE of higher amplitude than a sliding one. However there is no evidence that this is the case. Secondly it may also be because the variations in entrainment speed over the meshing cycle simply induce the thinnest film at the pitch point, and therefore the greatest probability of asperity contact. However the mechanisms that affect film thickness over a meshing cycle are intricate and it is not immediately obvious that there is greater probability of asperity contact at the pitch point than elsewhere. The variations in the contact parameters over a meshing cycle have different and competing effects on the film thickness. Near the pitch point the lower sliding velocity encourages both a reduction in film thickness: due to reduced entrainment, but also an increase: due to reduced shear heating. In short, the complexity of gear meshing makes it difficult to conclusively identify the fundamental mechanism which results in the pulse-like behaviour.

To try and confirm that the AE pulses were an effect of the variable lubrication conditions over the meshing cycle Tan and Mba [88] ran a comparison test where the gear pair was allowed to run dry. During this the waveform did not show the pulses characteristic of lubricated contact but a problem with this methodology is that the dry running caused significant surface damage which raises the question of whether the change was caused *directly* by the lack of lubricant, or by the wear it precipitated.

A more robust comparison method was employed by Raja Hamzah et al. [90] using temperature control to force changes in the lubrication conditions. In tests conducted on spur gears liquid nitrogen was used to temporarily cool the contact from its natural operating temperature to near 0 °C. The AE during the cooling and re-heating were observed and the waveform characteristics compared. It was found that at the coldest temperature the AE was almost free of pulses, whereas nearer the natural operating conditions, pulses were in evidence. It can be reasoned that this was because near 0 °C the extremely high oil viscosity ensured full film lubrication at all times, despite variations in the film thickness caused by the meshing cycle. Whereas at the more natural operating temperatures, and consequently thinner films, the variations caused cyclic transitions to and from the mixed regime, resulting in cyclic fluctuations in the amount of asperity contact and thus the AE. This test does suggest that the pulse-like nature of AE waveform is due to variable lubrication conditions in the meshing cycle but as the changes in asperity contact were not explicitly measured it is not certain that this was the specific cause.

The Cranfield researchers have also compared the waveforms of spur and helical gears and found that those from helical gears seem to be much less prone to pulse-like behaviour. Figure 2-31 shows a comparison of the two types of waveform from tests at identical speeds and loads and similar temperatures, from Raja Hamzah and Mba [91]. The authors attribute the more continuous nature of the waveform from helical gears to the more continuous rolling motion that occurs over the meshing cycle of a helical tooth pair. However this explanation relies on accepting rolling contact asperity interactions as the primary cause of AE, and as discussed this has not been explicitly proved. There may be alternative explanations relating to other aspects of the meshing cycle that are smoother / more continuous than in spur gears, such as tooth engagement and disengagement. And it is worth noting that the amplitude scale of the waveform for the helical gears is an order of magnitude greater than that for the spur gears, but there is little discussion on why this should be the case.

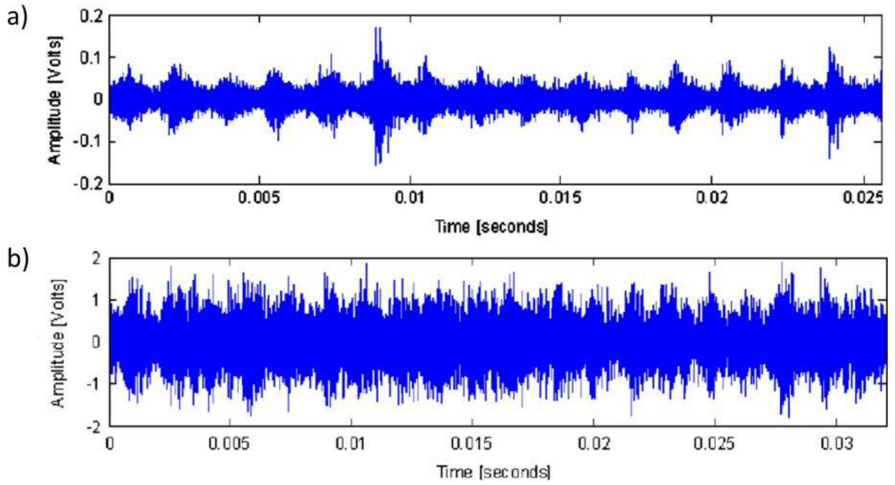


Figure 2-31 Comparison of the waveform from a) a spur gear pair and b) a helical gear pair for tests at identical speeds and loads and similar temperatures [91].

The Cranfield investigations into meshing cycle induced variations is interesting but limited and there is much more that could be done. For example it would be useful to do experiments where the waveform is synchronised with meshing cycle position. This would conclusively show where in the cycle the bursts occur. Additionally there is significant variation in the peak amplitude of each pulse and there has been no analysis of the cause of this. For example can it be explained by variations in tooth surface conditions? In which case it should be periodic. Finally, a spectral analysis of the waveforms might be useful to help quantify the extent of the pulse-like behaviour under varying conditions.

As well as investigating AE variations on timescales associated with the meshing cycle, the Cranfield researchers also considered variations over longer scales, using averages to filter out the meshing frequencies. In the tests by Raja Hamzah et al. [90], where liquid nitrogen was used to force a temperature change in operating spur gears, the specific film thickness was calculated and compared against the RMS AE over the warming period. Figure 2-32 shows this comparison for three loads tested at two speeds. For each constant load constant speed curve, the results show the highest RMS AE at the lowest specific film thickness, i.e. when there was most probability of asperity contact. But beyond this some of the tests show that the RMS AE did not increase significantly until the specific film thickness decreased below a threshold. This behaviour is consistent with the change in asperity contact for a lubrication regime transitioning from full film to mixed lubrication. To explain: For surfaces having asperities with a Gaussian-like height distribution, (which is broadly the case, even after running-in), the initial decrease in film thickness can be expected to have a negligible effect on the number of asperity contacts due to the small number of outliers involved. However as the decrease continues, its effects can be expected to become more and more pronounced due to the greater numbers of asperities becoming involved. Thus an exponential-like behaviour is convincing evidence that the RMS AE is directly dependent on the probability of asperity contact as there are no other obvious candidate mechanisms that would satisfy it.

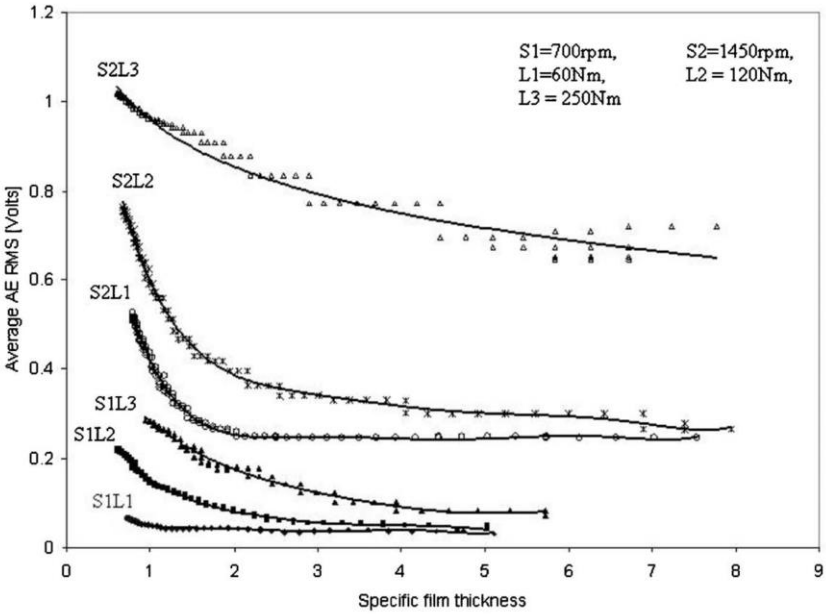


Figure 2-32 RMS AE vs. the specific film thickness during warming of a spur gear pair for three different loads at two different speeds [90].

The tests in Figure 2-32 also show that the RMS AE was *directly* affected by load and speed changes, i.e. outside of their influence on the specific film thickness. It can be seen that for a constant specific film thickness, an increase in either load or speed increased the amplitude of the RMS AE, with the speed having the greatest affect. The following hypotheses may explain both of these effects. Regarding the load: it is well established in EHL theory, that the load primarily affects the contact area rather than the film thickness. Thus, providing the lubrication is in the mixed regime, increasing the load should increase the probability of asperity contact through increased contact area rather than by decreased film thickness. Regarding the speed: as it increases, the *rate* at which asperity contact occurs increases. So for a fixed sample length, as was the case for the work under discussion, an increase in speed results in an increased number asperity contacts per sample, which, when using integral (summing) statistical parameters such as the RMS, could explain the increase in AE. If this is the case then the increase is simply a symptom of measurement inconsistency, i.e. a sample length not synchronised with the system speed. However as the rate of asperity contact increases so does the rate of deformation (plastic or elastic) and consequently the rate of release of strain energy. It has been posited that the amplitude of AE is dependent on this rate and if so then, even with speed synchronised measurements, the AE would be expected to increase.

Although the results of Raja Hamzah et al. [90] reveal the strong dependence of AE on the probability of asperity contact the specific relationship between them is not precise. There are inconstancies in the form of the AE increase for the different speeds and loads, and although these are fairly minor for the spur gears (Figure 2-32), they are major for the helical gears (not shown). These inconstancies may be due to inaccuracy in the specific film thickness calculations which, given the complexity of gear meshing, are likely to contain significant uncertainties. For example, and as the authors acknowledge implicitly, the Dowson and Higginson equation used to predict the oil film thickness is based upon many assumptions that are not met in a real gear contact, e.g. no edge effects and isothermal conditions. Additionally the authors assume constant surface roughness, i.e. negligible wear, which may be possible depending on the test duration but needs to be proved either way. Finally there may be inconsistencies relating to AE noise. For example during a static calibration test the authors identified AE associated thermal strain caused by the speed of the temperature change.

In different experiments the Cranfield researchers, Raja Hamzah and Mba [91], again compared the relationship between the RMS AE and the specific film thickness, but calculated as an average during steady state conditions for four different loads at three different speeds and all at similar temperatures. For these tests the potential for AE from thermal strain was reduced but the uncertainty in the specific film thickness calculation largely remains. Figure 2-33 shows the results for tests on a spur gear pair. The direct speed effect is again in evidence, with the RMS AE at constant load increasing *despite* an increase in the specific film thickness. However it is difficult to draw conclusions on any direct effect of load; unlike the results using variable temperature (Figure 2-32), there is little data that allows comparison of different loads at a constant specific film thickness.

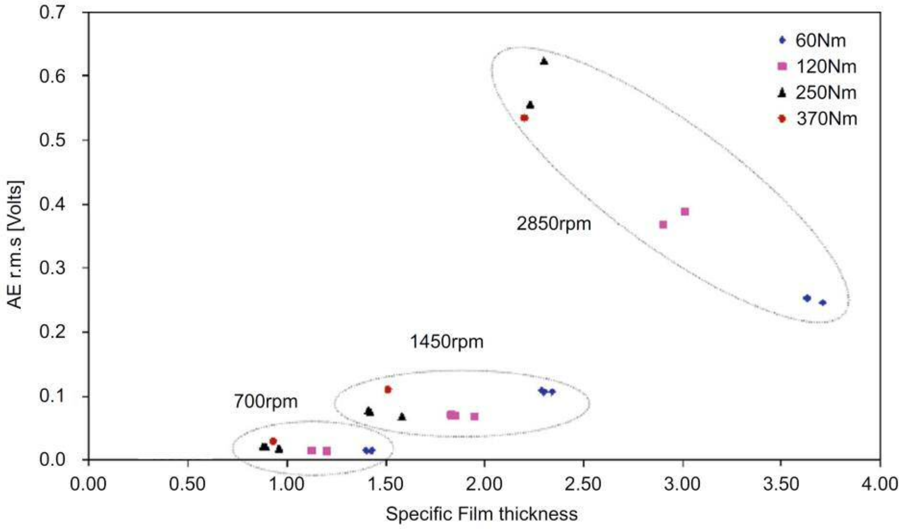


Figure 2-33 RMS AE vs the specific film thickness for tests on a spur gear pair at different speeds and loads and similar temperatures [91].

Raja Hamzah and Mba [91] proposed a mathematical fit for the relationship between load, speed and the AE from their spur gear tests, this can be seen in Figure 2-34. As the authors acknowledge, it relies on a number of limiting conditions which include constant surface roughness and a constant temperature. Unfortunately given the very limited number measurement points, and the significant residual error at some of them, the fit is not very convincing. Nevertheless it is significant because it represents one of the few attempts to mathematically express the relationship between AE and operating conditions for applied contacts.

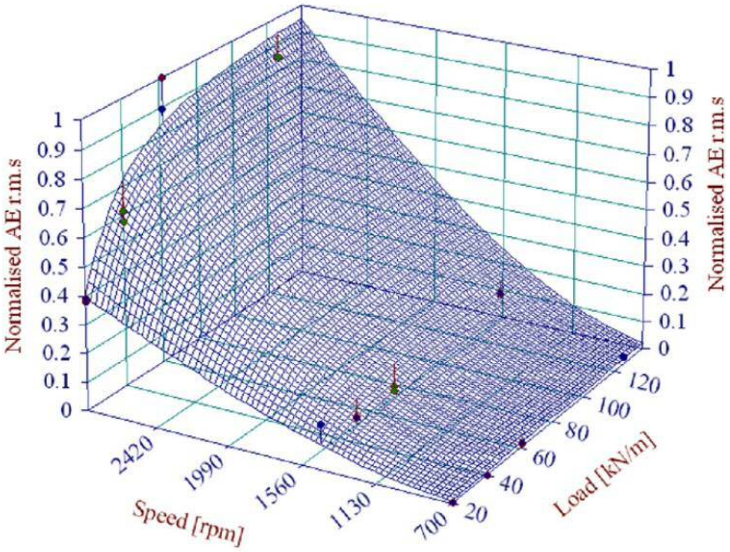


Figure 2-34 Fit for the relationship between load, speed and AE in a spur gear pair [91].

The remaining examples presented in this section were not part of the Cranfield experiments. The first is from Douglas et al. [104] who investigated the AE from the piston ring / cylinder liner interface of a diesel engine. Using sensors bonded to the cylinder block and with the cylinder head off (so no compression effects), and using a motor to drive the crank, they found that the amplitude of the AE varied in accordance with the speed of the piston head. Figure 2-35 shows this for four strokes of the piston head. At the top and bottom of the stroke when the velocity was at a minimum the AE was also, and at the midpoint of the stroke when the velocity was a maximum so was the AE. They also conducted tests on operational engines, i.e. with the cylinder head on and the piston driven by the expansion / expansion cycle. Similar oscillatory behaviour was found but with additional spikes which were attributed to valve activity.

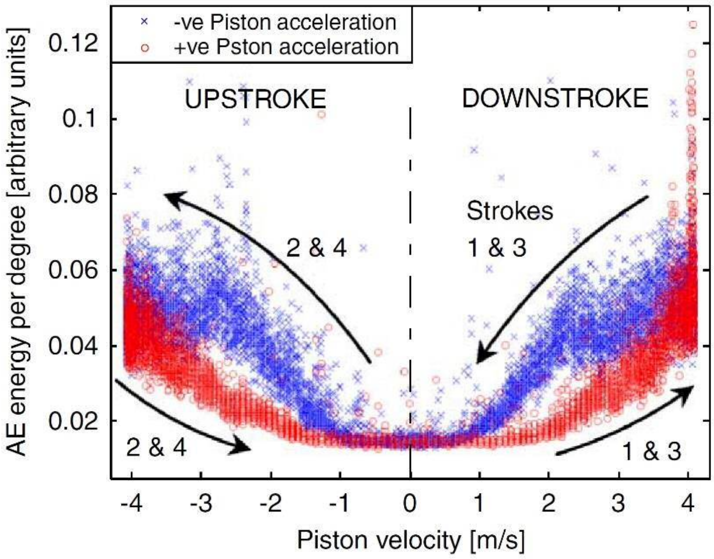


Figure 2-35 The relationship between AE energy and piston velocity for four strokes when driven by the crankshaft [104].

Douglas et al. discussed possible causes of the AE, but because of the complexity of the system they could not establish a definitive one. Piston heads make contact with the cylinder liner using several different rings. They ruled out asperity interactions with the compression rings because these contact under EHL and thus increasing speed should have reduced asperity contact and thus AE. Instead they attributed the AE to asperity interactions on the oil-control ring because this contacts by boundary lubrication and additionally its friction behaviour was similar to the AE. This test is a good illustration of the difficulty of isolating an AE mechanism in a complex system, i.e. one with multiple and different types of contact.

The next example, from Mirhadizadeh et al. [100,103], concerns the AE from a hydrodynamic journal bearing carrying a radial load. This study is interesting because these typically run with a much higher minimum specific film thickness than is generated in an EHL contact. Thus it is possible, in theory at least, to identify any AE mechanisms that are associated with full film lubrication. Mirhadizadeh et al. investigated three grades (inherent thicknesses) of oil at five speeds and three loads. The temperature was allowed to fluctuate based on the operating conditions and the RMS AE was measured at steady-state conditions. The load was applied to the journal (shaft) through rolling element bearings located either side of the test bearing. A novel method was used to help isolate any AE from the loading bearings and the test bearing, this involved placing a baffle, in the form of a Nylon sleeve, between the journal and the inner raceway of the loading bearings.

The results showed that the AE increased with speed *irrespective* of decreasing or increasing film thickness; this can be seen in Figure 2-32 which presents the results for the largest load tested. For each oil there was a maximum film thickness which was due to the competing influences of increasing speed and temperature. The authors assert that because the film thickness, even at its thinnest, was an order of magnitude greater than bearing roughness there was no asperity contact under any of the test conditions. Whilst this is indeed likely, no evidence to support it is provided. The roughness of the bearing surface is simply given as ‘approximately 3 μm’ without reference to how it was measured and the roughness of the shaft is not given at all. Although it is possible that, when the effects of combined roughness and outlier asperities are considered, there may have been some limited asperity contact at the thinnest films, it is doubtful that this significantly influenced the experiment given the consistency of measurements between relatively thin and relatively thick films.

After some analysis the authors found the AE had the most satisfactory correlation with power loss in the bearing. But here, the most important thing to note is the existence of an AE source not related to asperity contact, but rather fluid flow phenomena, primarily shearing, the authors conclude. This raises an interesting question regarding the AE source in mixed lubrication, namely: is it asperity contact, the fluid flow, or a combination of both?

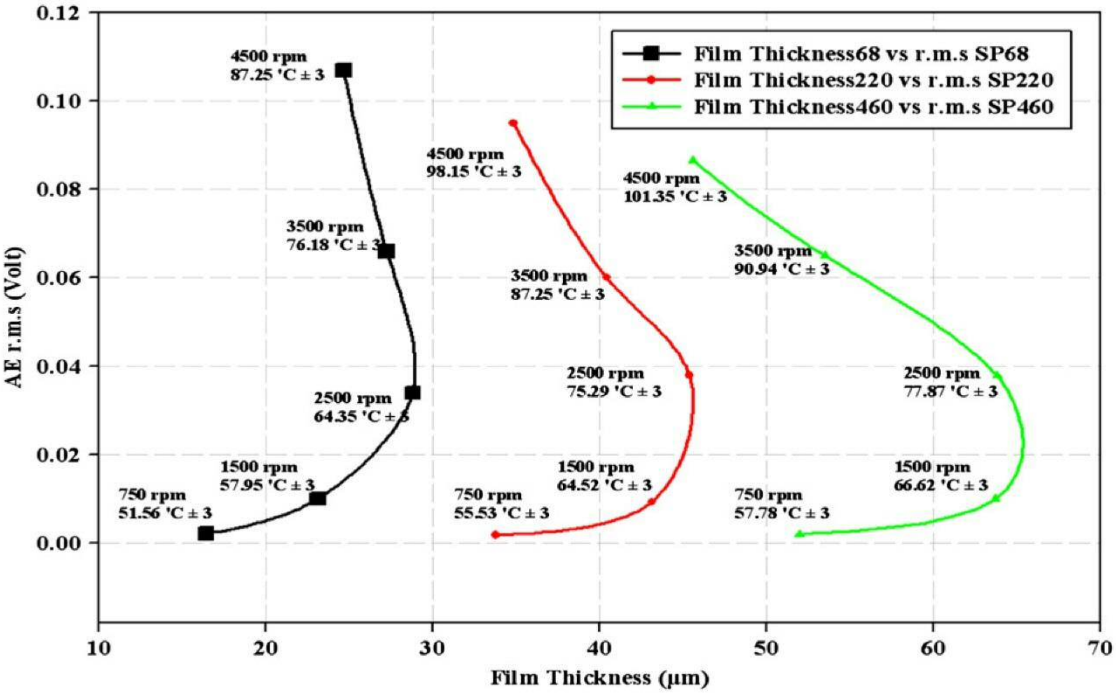


Figure 2-36 The relationship between RMS AE, film thickness and speed at the maximum load, for three different oils [100].

The final detailed example, from Loutas et al. [105], involves spur gear contact and is included because of the novel way in which the AE sensor was interfaced with the gears. For all the Cranfield experiments the sensor was bonded to the side of a gear, consequently it rotated necessitating the use of a slip ring to interface the signal with the acquisition hardware. Loutas et al. avoided the need for a slip ring by interfacing a sensor to the side of a gear using a sliding contact, i.e. a static sensor. The sensor was bonded behind a bronze wear plate which was spring loaded against the side of the gear. It might be predicted that the sliding sensor / gear interface itself would generate AE thus completely obscuring any signal from the test contact, but the results show that the sensor was, at least partially, able to detect changes in operating conditions.

Figure 2-37 shows how the RMS AE varied during a constant speed test where the rig was started cold and under maximum load, then allowed to heat to a steady state temperature, whereupon the two lower loads were tested. During the heating it can be seen that the RMS AE increases which may be explained by increasing asperity interaction at the test contact. A correlation between the AE and gear temperature would help confirm this but only the oil temperature is provided and it cannot be assumed that they are the same. (The gears were lubricated by partial submersion in sump oil but the contact was out of this). Additionally the sensor / gear interface is likely to have been lubricated by oil splash and so the changes in AE during heating may have been from this rather than the test contact. The effects of changing load offer better evidence that the sensor was detecting AE from the contact: When, after heating, the load was removed completely (load 0) the AE dropped significantly. Since the sensor / gear interface should not have been affected by this it follows that the drop was due to a change at the test contact. There are however some inconsistencies in the load related changes, for example why does the increase from load 1 to load 2 (a doubling) only elicit a temporary increase in AE, and why does the AE not return to the previous highs? There are significant unanswered questions regarding this work, and unfortunately there is insufficient data to assess the efficacy of the novel sensor interface.

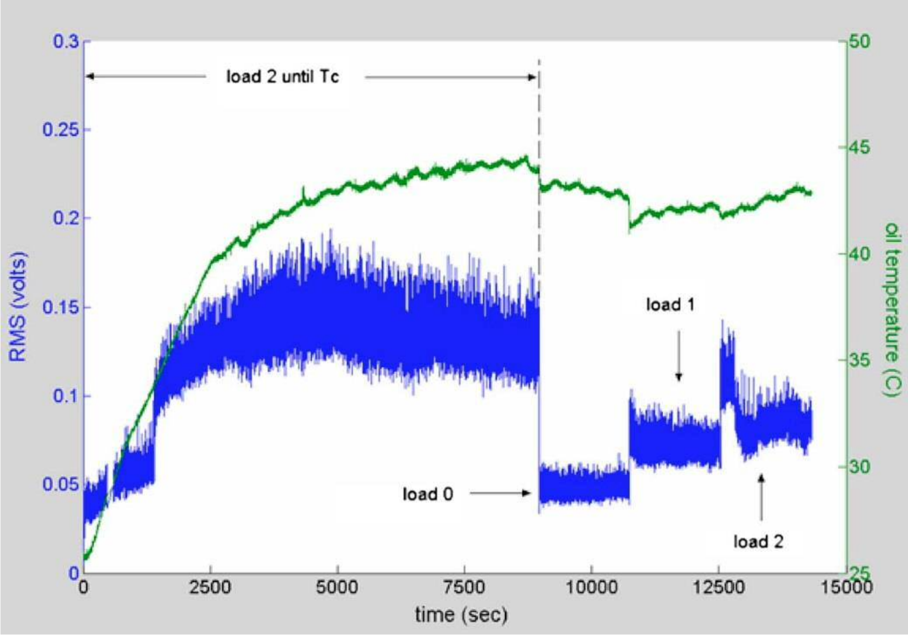


Figure 2-37 RMS AE and oil temperature vs. time, for a spur gear pair at constant speed and several loads [105].

Some of the other recent experiments into the AE from applied moving contacts are now listed:

Singh et.al [106] measured the AE from crack growth in a static fatigue (cyclic bending) test on a gear tooth, the AE sensor was placed directly on the gear tooth. (Although a static test, this is relevant as it concerns a possible source of AE in operational gears). They also measured the attenuation of AE signals through mechanical interfaces (bearings and splines) under stationary torque using one AE sensor to transmit a simulated pencil lead break test and another to receive the signal.

In a similar fatigue test on a static gear tooth, Pullin et al.[107,108] used Digital Image Correction to identify the onset of cracking. The crack growth was compared against the AE, as measured by a tooth mounted sensor, to demonstrate that AE has the potential to detect fatigue fracture in gear teeth.

Scheer et.al [109] measured the AE from a gearbox using a variety of sensors at different positions on the housing and on the ends of two rotating shafts. The end of a running-in procedure was identified a decrease in initially high AE activity to steady state levels. And a reduction in slack (excessive clearance) in the bearings, caused by thermal expansion, was also detectable by an associated decrease in the AE. They also measured the AE from pitted and cracked gear teeth using wavelet transforms and found AE signatures associated with the two types of damage. Finally, using the high temporal resolution of the wavelet transforms and the different sensor positions they were able to approximate the location of specific AE events.

Al-Dossary et.al [92] investigated the AE from seeded defects (pits) in the inner and outer raceways of large, radially loaded, split roller (Cooper) bearings. Their analysis was based on examination of the cyclic bursts of AE associated with the passing of the rollers over the defects. They found a correlation between the size of the defects in the outer raceway and the parameters of the AE bursts. There was no correlation with the size of defects on the inner raceway, which was attributed to greater signal attenuation.

Elforjani and Mba [110] measured the AE during an accelerated natural (not seeded) wear test of a slow speed thrust bearing raceway. The test was run until damage, cracking, was visually observed. Using an assortment of signal processing and statistical techniques they were able to show that the onset and propagation of cracking could be identified by the AE measurements.

Nikam et,al [111] evaluated the effectiveness of different AE parameters to distinguish between dry and lubricated low speed spherical roller bearings. Amplitude thresholding was used to identify and record the waveforms of discrete events. It was found that the standard deviation, variance, peak amplitude, mean amplitude could all be used to identify when the bearings were running dry.

Vicuña [112] investigated the effects of common operating variables: lubricant temperature, speed and load, on the AE from a healthy planetary gearbox, when operated in a lab and in a bucket wheel excavator. For lab tests at constant speed and load but variable temperature he observed both an increase in AE RMS, attributed to an increase in asperity contact caused by reducing film thickness, and a decrease in AE RMS attributed to a decrease in asperity contact caused by running-in. He also determined that rotating speed has a significant effect on the AE RMS and load only becomes important at low speeds.

Qu et.al [113] compared the capability of AE and vibration measurements to detect seeded partial tooth cuts (truncation) in a spur gearbox, when the same low sampling rate was used (low by AE standards). Using a bespoke frequency reduction technique developed by the authors (involving heterodyne demodulation), the AE measurements were able to identify more subtle tooth damage than the vibration measurements.

Wirtz et.al [114] tested a spur tooth gearbox using AE sensors mounted on the housing. They found that the meshing frequency of the gears was easily identifiable by transient bursts of AE in the time domain. The Wavelet Transforms of AE from gears with micro-pitting and pitting damage were compared against those from healthy gears and it was found that the AE from damaged gears could be distinguished by its spectral signature.

In their paper, Novoa and Vicuña [115] discussed various hypotheses for the source of AE in gears. Other than asperity contacts, they proposed that pressure perturbations caused by mixed lubrication, tooth deformation and tooth pair engagement impact could all be sources. They also tested the effects of load, speed and lubricant viscosity on a planetary gearbox using both healthy and damaged teeth. They analysed the envelope (of the AE signals and found that both load and speed affected the AE amplitude and that a thinner oil generated more AE. Tooth defects were also detectable in the spectrum of the enveloped AE.

As well as using AE measurements to monitor the condition of gears and bearings there has also been significant research into using them to monitor machining and finishing processes. As an example of this Strömbergsson et.al [116] investigated the AE from a 'machanochemical' finishing process which involved the deposition of a tribofilm during the burnishing (polishing) of a shaft. They found that the AE could identify the end of a reduction in the coefficient of friction caused by the tribofilm deposition reaching a maximum. Tool wear could also be detected by changes in the AE signals, but not at an early stage.

Lastly some researchers have focussed on making it more practical to measure the AE from rotating and inaccessible components. An example of this is the work by Tian et.al [85] who developed a self-powered (energy harvesting) wireless AE sensor in order to detect tooth cracks in gears. Their sensor relied on the use of an on-board, and low-power, diagnosis algorithm that compared the spectral signatures of damaged and healthy gears. The sensor showed good performance in laboratory tests and the diagnosis of damage was effective at all speeds tested.

2.11.5 **Summary**

This section presents a summary assessment regarding the current understanding of AE from moving contacts:

a) There are a sufficient number of different experiments reported in the literature to suggest that most moving contacts, whatever the lubrication regime, will cause some AE which is liable to vary with contact parameters such as load, speed, materials, wear and friction.

b) There is often a good (positive) correlation between AE and the coefficient of friction in simple contacts but it is not guaranteed.

c) Contact between asperities can cause AE, and will do if there is any associated wear. But it is unknown if wear is a precursor for an asperity contact to generate AE. In other words it is unknown if wear free asperity interactions will cause AE.

d) In idealised small-scale contacts the spectral contact of the observed AE can vary depending on the materials, lubricating conditions and type of wear. The extent to which it varies in large-scale, applied, contacts is less understood.

e) In mixed-lubrication contacts the amplitude of the AE increases with increasing probability of asperity contacts, i.e. decreasing specific film thickness, but the mechanisms at work have not been conclusively established. There may be a combination of effects from wear-free asperity contact, wearing asperity contact and fluid flow phenomenon.

f) In a lubricated contact an increase in speed will *directly* encourage AE of greater amplitude, i.e. independently of any affects relating to a change in film thickness. An increase in load may also directly affect the AE but there is less evidence of a consistent systematic relationship.

g) In gears and bearings AE measurements can be used to detect various types of surface damage. This detection usually relies on detecting the effects of the damage rather than its formation mechanism (cracking and fracture). The detection usually relies on prior knowledge, i.e. comparisons with the AE signature from a healthy component.

h) In gears there are often transient bursts of AE at the meshing frequency. These are usually attributed to asperity contacts but it is not easy to demonstrate that this is definitely the case. There are other potential sources of AE in gears and some of these have yet to be ruled out.

i) Experiments involving applied contacts typically make many assumptions, rather than measurements, when attempting to identify / predict the precise contact conditions; this creates considerable uncertainty any relationship with AE. (For example, the influence of the specific surface topography, and changes to it, have not yet been considered).

2.12 Conclusion

The literature review has shown that for gear contacts there is a lack of experimentation which couples AE measurements *and* detailed knowledge of the lubrication conditions. The complexities of real gear contacts makes determination of the precise lubrication conditions difficult, thus there remains some considerable uncertainty regarding the fundamental relationship between AE and the severity of the lubrication conditions. The aim of this thesis was to shed new light on this relationship by conducting experiments on a ‘twin disk’ rig. This rig, which is introduced in the next chapter, was designed to simulate gear-like lubrication conditions as an alternative to testing directly on gears.

3 Method: the twin disk rig

3.1 Introduction

This chapter describes the principles and operating and measurement procedures of the twin disk rig. This rig is used to study a lubricated contact made between two axially parallel, and geometrically similar, disks. Disk contacts such as this have few practical applications, instead they are used to simulate the contact between gear teeth. Direct study of gears is complex as, even when operating in steady state conditions, many of the contact parameters are variables. Over the meshing cycle the contact kinematics, and sometimes the geometry, are in continual flux; and the load will change abruptly with the number of teeth in contact. Additionally there are start and end effects to consider, such as vibration caused by teeth engaging and disengaging. In contrast, a disk contact is simple. In steady state conditions all the contact parameters, with the exception of the surface topography, are constant, and thus much easier to measure and analyse. A disk contact can be designed to replicate the conditions found at a single point on the meshing cycle of gear teeth, i.e. it provides a ‘snapshot’ of contact conditions at a specific periodic time. By altering the parameters of the disk contact, snapshots of different points in the meshing cycle can be examined and compared. A more detailed explanation of the principles by which disks can be used to simulate gear contacts can be found in the work of Merritt [117].

The twin disk rig used in this research was originally designed by Alanou [118] to study micro-pitting fatigue in aerospace transmission gearing. It has now been in use for over a decade to study the effects of mixed lubrication on both running-in and micro-pitting [61,119]. It has also been used to provide empirical data with which to validate mixed-EHL numerical simulations [120]. Over the life of the rig successive researchers have upgraded its instrumentation so that a considerable number of different contact parameters can now be measured. The most novel measurements are: in-situ and repeat position surface measurements of the test disks, the contact voltage, and new for this thesis, the contact acoustic emission. Although all these is techniques have been used to by themselves to investigate mixed lubrication, it is rare to find them implemented together. As AE measurement was new for this thesis the instrumentation considerations, such as sensor choice, placement and fixture are discussed in some detail.

3.2 Twin disk rig schematic

Figure 3-1 shows a plan view schematic of the twin disk rig and highlights the features that will be discussed in detail throughout this chapter.

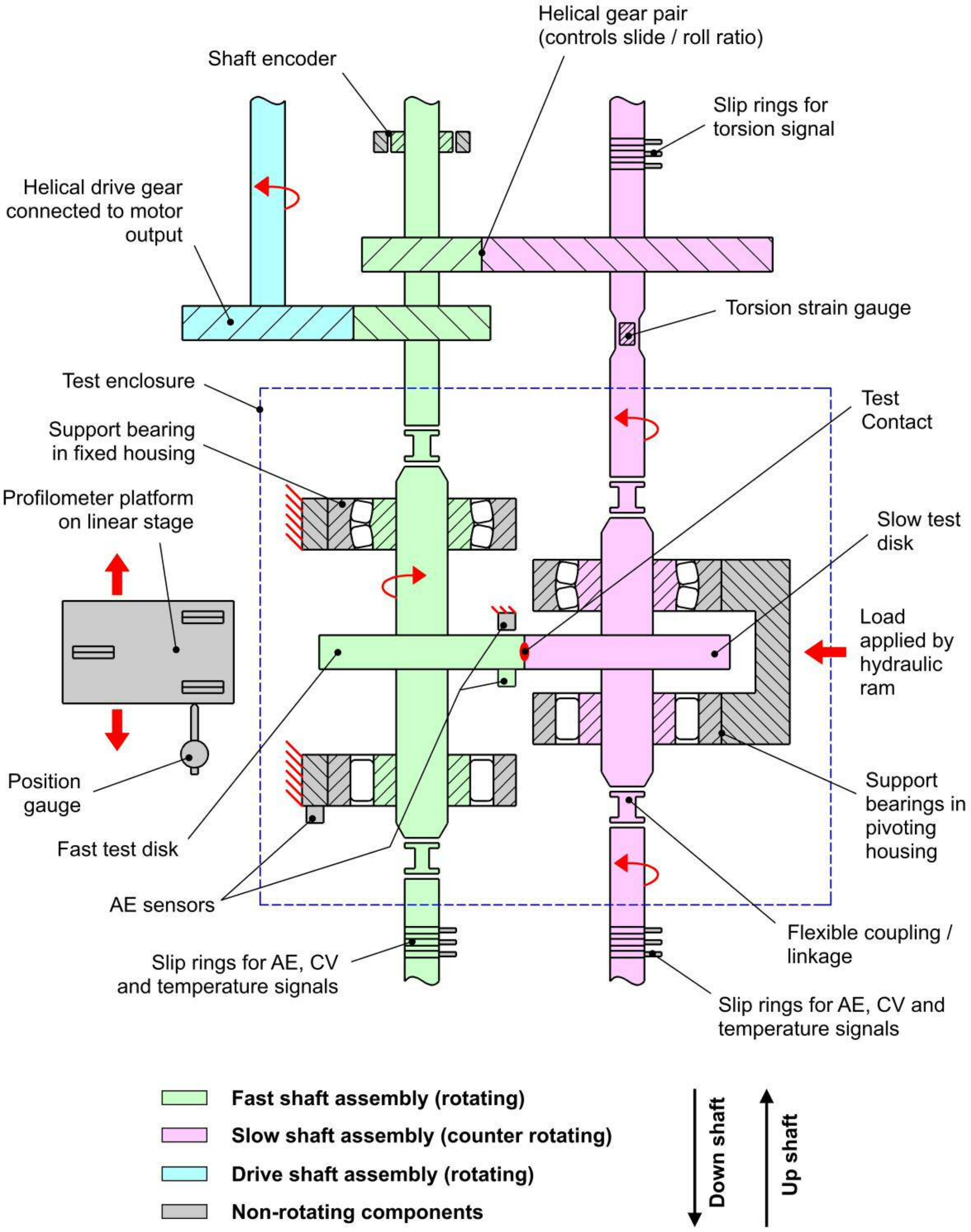


Figure 3-1 Twin disk rig schematic

3.3 Contact characteristics

This section discusses the details of the test contact. These are determined in part by the rig design and operating conditions, and in part by the test disks which are consumable and changeable.

3.3.1 Rig controlled contact parameters

Kinematics

The rig consists of two parallel shaft assemblies, one for each test disk. They are connected by a helical gear pair and the test contact and rotate in opposite directions. If the gear ratio (pitch diameter ratio) of the helical gears and the test disks is equal then there will be no conflict in their desired motion, and both disk surfaces will move through the contact[1] with identical velocity. This is the condition of *pure rolling*. If the gear ratio of the helical gears is not equal to that of the disks (as is the case in the schematic) then there will be a conflict in their desired motion. Consequently the disks will slip, as the helical gears cannot. In this case the disk surfaces move through the contact with different velocities but still in the same direction. There is thus a combination of rolling and sliding which is defined by the *slide / roll ratio (SRR)* calculated thus:

$$\frac{\quad}{\quad} \qquad \frac{\quad}{\quad} \tag{9}$$

where is the gear ratio of the disks (the speed of the slower disk divided by the speed of the faster), and where and are the surface velocities of fast and slow disks respectively.

The helical gear pair can be switched to give different SRRs but not whilst the rig is in motion, i.e. mid test. Although disassembly of the test disks is not required to change the SRR, their angular orientation relative to one another will be lost during in the switching procedure. This has no implications for smooth disks but does for rough, and this is discussed later in this chapter. The helical gear pair are always configured so that the left-hand shaft is the faster of the two, so this is termed the fast shaft and the other the slow.

The rig is driven by a variable speed electric motor which is connected, by a fixed helical gear pair, to the fast shaft. The motor / gear configuration gives a range of fast shaft speeds from 200 to 2000 rpm (the slow shaft speed depends on the specific SRR). Helical gears are used on the twin disk rig in preference to spur because of their smoother transmission characteristics which help to reduce oscillations at the test contact associated with gear meshing.

---

[1] Motion 'through the contact' means motion through a reference frame fixed at the position of the contact.

Lubricant and temperature control

The test contact is lubricated by an oil meeting Defence Standard 91-74 [121]. This oil specification, termed 'OEP-80' is for gear applications with high tooth loads, in particular, marine gearboxes and steam turbines. Oils meeting this specification contain Extreme Pressure (EP) additives. Unfortunately the exact oil formulation was unknown, OEP-80 was used for convenience as it was the oil used in the rig when this work was begun.

To ensure that the contact always operates in the fully flooded condition the oil is sprayed directly onto its inlet and (less critically) its outlet. The oil is pumped from a temperature controlled bath and passed through a 1 μm mesh filter to remove debris that might otherwise cause significant third particle wear at the contact. Oil from this bath is also used to lubricate the bearings and gears of the rig.

The temperature of the test enclosure is primarily dictated by the temperature of the oil bath. Such is the flow of oil through the enclosure that significant heat exchange takes place and given sufficient time the temperature of the test enclosure will approach that of the oil bath to within a few degrees. Depending on the operating parameters the test contact will generate differing amounts of frictional heat. This can have local affects, i.e. on the temperature of the disks, but it is a negligible heat source compared to that of the oil and so does not significantly affect the enclosure temperature.

Loading

The design considerations required for loading the test disks are not as straight forward as might initially be supposed. Firstly, as discussed in the section on Hertzian contact when the disks are in contract their surfaces will elastically deform by an amount dependent on the load. So in order to load the disks their centres distance (the distance between their axes) must be variable. Secondly, EHL is ineffective at slow speeds and thus it is undesirable to load the disks until they are at speed (unless start-up effects are specifically being investigated). So the loading and unloading of the disks must be possible whilst they are in motion. These respective requirements are met by allowing the slow shaft / disk to pivot slightly and by using hydraulic actuation.

The fast disk is supported by fixed bearings whereas the slow disk is supported by bearings in a pivoting housing. With reference to the rig schematic (Figure 3-1) the axis of the pivot is parallel to that of the disks but beneath them, i.e. under the plane of the schematic. With no test force applied to the pivoting housing it swings (due to its centre of gravity) so that the disks are fully apart. To load the disks a force is applied to the pivoting housing which moves the slow disk into contact with the fast. Flexible couplings either side of the pivoting section of slow shaft ensure that it is not restrained by bending forces.

The loading force is applied to the pivoting housing by a hydraulic ram. Hydraulic actuation allows for quick application and removal of the contact load. The hydraulic circuit runs continuously and the fluid pressure to the ram is applied and dumped extremely quickly by use of an off / on bypass valve. This design feature was inherited from a previous rig used to study scuffing, where it can be necessary to quickly stop a test. It has also proven useful on the twin disk rig, but to study running-in. This has previously been found to occur extremely quickly, within tens of seconds [61], and so short tests and correspondingly shorter starting and stopping ramps are required.

To start a test with a short loading ramp the rig can simply be run up to speed and then the ram bypass switched off. There is then a short delay as the ram extends to make contact with the pivoting housing but once the disks are in contact they typically ramp from zero to full and steady-state load within a few seconds. The desired load must be pre-set before testing. This requires that the ram be switched on so that its load can be measured by a load cell located in-between the ram and the pivoting housing. A pressure control valve can then be adjusted to set the desired load. To prevent loading the disks during this procedure blocks are placed in-between the two shafts to prevent them pivoting sufficiently to bring the disks into contact; this necessitates that the load is set whilst the shafts are stationary. During the experiments conducted for this thesis it was observed that the load generated by the ram varies with the temperature of its oil supply. Consequently, the set load only applies for a specific oil temperature. As both the ram and lubricating hydraulics take oil from the same bath the test temperature cannot be varied mid-test without affecting the load[1].

3.3.2 Disk controlled contact parameters

With the exception of the lubricant, the contact parameters discussed thus far: the kinematics, temperature and loading, are those which are, at least partially, determined by the rig. The remainder: the material, geometry and roughness are those which are determined by the specific test disks. Although, it is possible in theory to test a wide variety of disk types, in practice there is a standard design which has been used for all experiments on the twin disk rig. This allows for efficient manufacture (batch production) and comparable testing. The only design variation is the surface finish which is discussed shortly.

Material

The disks are made of a case hardened steel, of a type commonly used for aerospace gearing (at least at the time when the twin disk rig was commissioned). Hardening of gear surfaces is typical as they would otherwise wear quickly in mixed-EHL conditions. The aim of this thesis was to characterise the AE from a *gear-like* contact rather than from a *specific* gear design / component. Thus the specific material properties, such as the steel composition and the heat treatment, are not of primary interest, but can be found in [118] if desired.

---

[1] This limitation was not considered during the design of the rig as it was intended for testing with a constant oil temperature.

Geometry

A schematic of the test disks, in their contact orientation, is presented in Figure 3-2. The dimensions that determine the contact geometry are shown. The contact faces are not cylindrical, as they appear by eye, but are gently radiused (‘crowned’) in the axial direction. Geometrically, each contact surface approximates the shape of a truncated ellipsoid[1]. Under load these generate an elliptical point contact, with an ellipticity (the ratio of the major to minor dimension) of approximately 0.25. The ellipse is orientated with the minor axis pointing in the entrainment direction. The study of a point, rather than linear (roller), contact is preferred for two reasons. Firstly, the contact boundary at its ends is more continuous and predictable than that of a linear contact. Provided the disks are not overloaded, the point contact will not approach the edge of the disks and will be free of the complications associated with them. In contrast the ends of a linear contact would be defined by the edges of the disks, near which, many of the assumptions required for Hertzian or EHL analysis would not be met. Secondly a point contact is more tolerant of disk misalignment than a linear one. An angular misalignment on the common plane of the disk axes would dramatically alter the nature of a linear contact which would shift to the disk edges. However for a point contact it simply alters the position of the contact on the disk face. Thus, a long-thin ellipse, as studied here, represents a good compromise between the need to replicate the long-thin nature of typical gear contacts, and the desire for experimental simplicity.

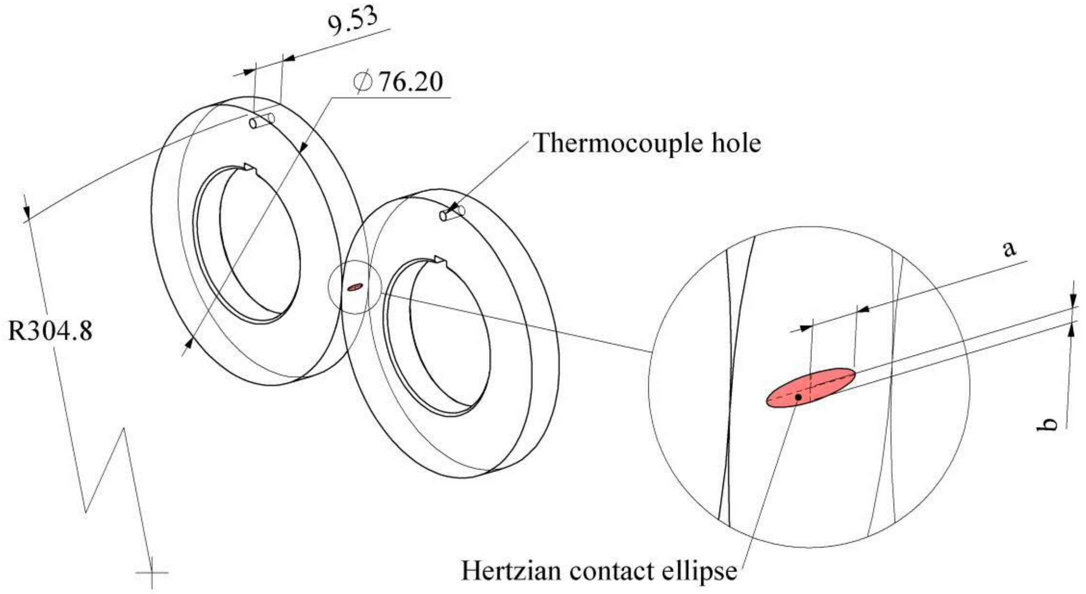


Figure 3-2 Geometry of the test disks and the resulting test contact. All dimensions in mm.

---

[1] An ellipsoid is an ellipse of revolution, with the axis being the minor or major dimension.

Disk Roughness Topography

The roughness topography of production gearing is typically unidirectional, i.e. it has long and aligned, valleys and ridges. These are oriented so that they lie transversely across the contact with respect to the direction of surface motion. To reproduce this on the test disks the contact surfaces are machined using a special surface grinding technique. In one operation this both, forms the crown, and imparts gear-like roughness topography. The technique involves grinding the, as yet, cylindrical disk face against the internal surface of a cone. Specific details of the technique can be found in [118].

A pair of disks in their ‘as ground’ state are used to investigate mixed-lubrication conditions between two ‘rough’ surfaces. However, sometimes, for simplicity or comparison, it is necessary to investigate the case of a rough-on-smooth surface or a smooth-on-smooth. This requires disks with a ‘smooth’ contact surface, and this is usually achieved by isotopically superfinishing [122] the as-ground disks. This largely removes the unidirectional roughness topography, leaving a mirror-like finish, without altering the surface form.

Figure 3-3 shows the typical roughness topography of unused, as-ground and superfinished test disks. The as-ground roughness topography consists of uniform and aligned, ridges and valleys that sweep, in a slight curve, across the width of the disk. There is a small amount of ‘noise’ where isolated peaks break up the ridges and valleys but ostensibly the profile of the roughness is uniform across the width of the disk, (although its position changes due to the slight curve). In comparison the superfinished roughness topography is dramatically different. The well-defined ridges and valleys are gone, leaving a considerably more isotropic finish. There are still some vaguely defined, and gently undulating, ridges and valleys running across the width of the disk, but they are subtle and do not dominate the roughness topography. The height of the features is also considerable smaller, as evidenced by the more limited tonal range of the superfinished surface. The circumferential (RMS roughness) of the surfaces in Figure 3-3 are 0.56 μm for the as-ground roughness topography, and 0.052 μm for the superfinished. An of tenths of a micron is typical for untested as-ground disks, as is hundredths of a micron is for untested superfinished disks. The difference between them, an order of magnitude, is sufficient that in comparison the as-ground disks can be considered ‘rough’ and the superfinished ‘smooth’.

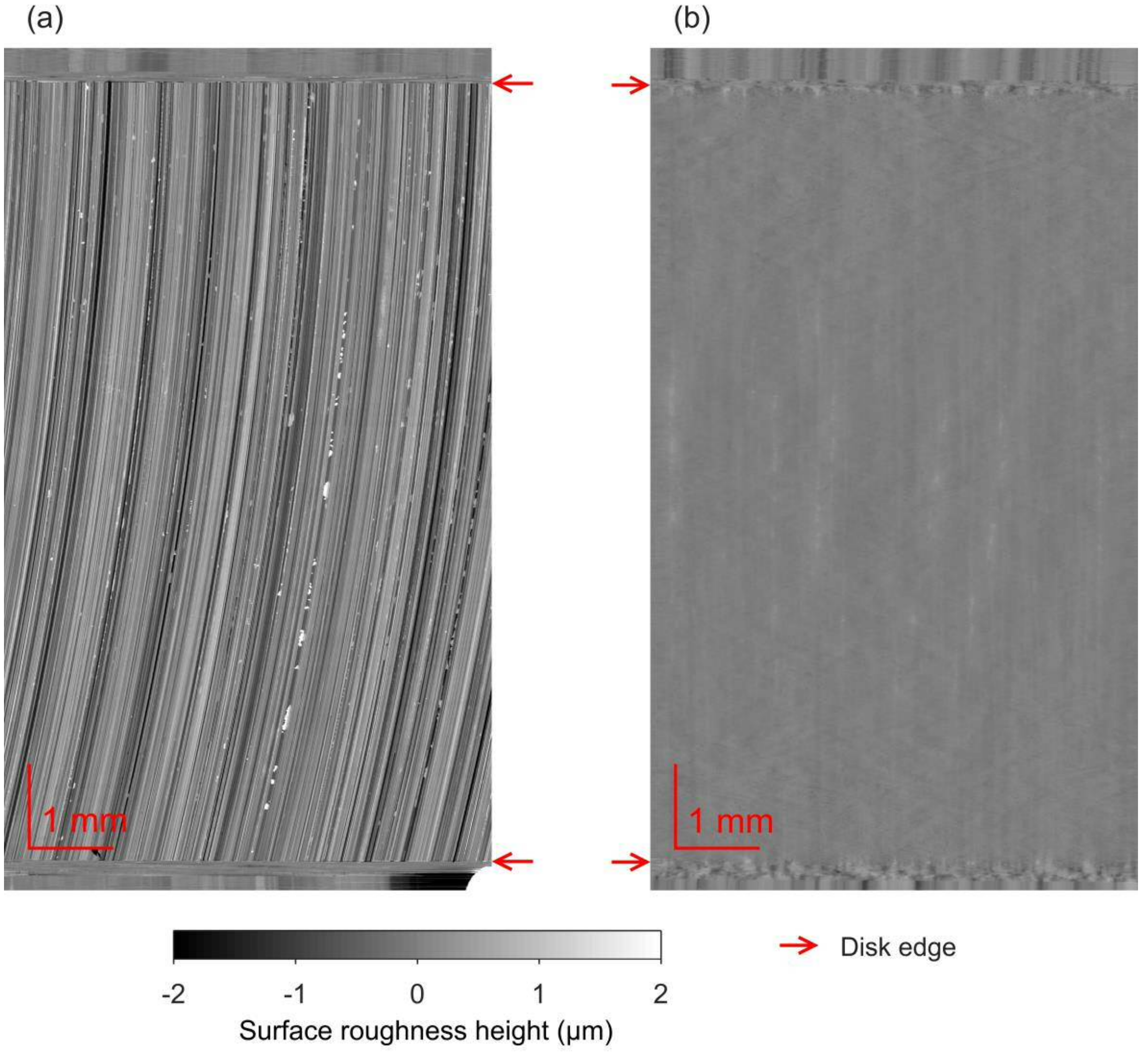


Figure 3-3 Comparison of the finish of (a) an as-ground disk, and (b) a superfinished disk using a six millimetre length of each contact surface measured by 3D profilometry. A common roughness (tone) scale has been used for both disks.

Contact topography

When investigating mixed lubrication it is not sufficient to consider the roughness topography of each disk in isolation, it is how the roughness topographies of both disk *combine* in the contact that is important. This is termed here the *contact topography* and it can be represented by overlaying, i.e. summing, the two topographies, as they would be positioned in the contact.

As discussed in the previous paragraph there is a slight curve to the roughness topography of an as-ground disk. When two disks are assembled in the twin disk rig they are always orientated so that the direction of curvature on one disk is opposed to that on the other. Thus the ridges and valleys of one disk cross those of the other disk at a shallow angle, resulting in a cross-hatched contact topography[1], an example of which is shown in Figure 3-4.

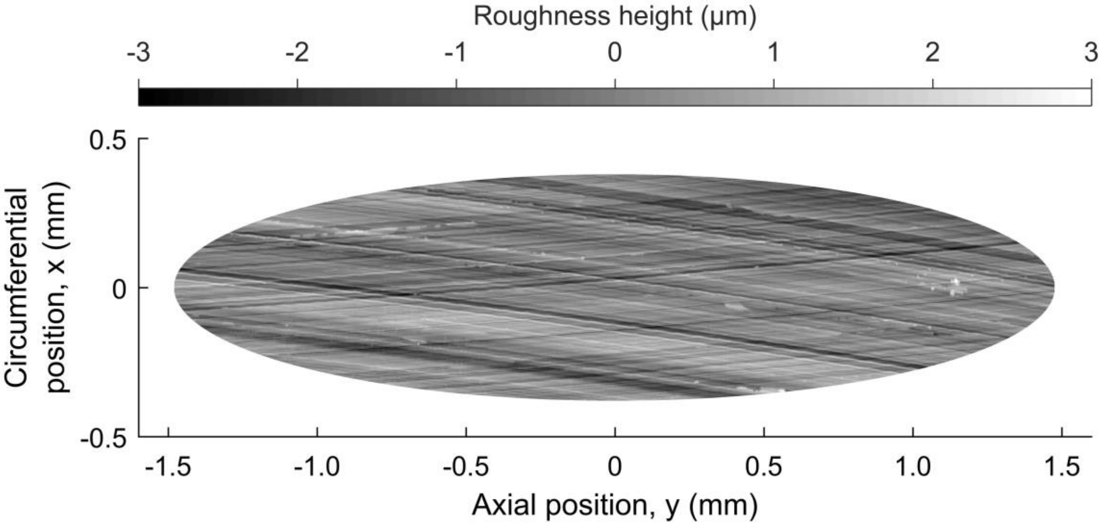


Figure 3-4 An example of the contact topography of as-ground, or 'rough' disk pair. Shown for a Hertzian contact pressure of 1.2 GPa.

It is worth considering the implications of a cross-hatched compared to single-angle contact topography on mixed lubrication. It is reasoned that cross-hatching will 'dampen' the frequency and amplitude of fluctuations in the instantaneous amount of asperity contact. To explain: consider the nature of direct contact between opposing ridges. If they met at the same angle, they would contact over a large area: their entire length, however at an angle they would only be contact over a small area, thus the 'amplitude' of their contact would be limited. However they would be unlikely to meet and would only remain in contact briefly, at a single position. In contrast, at different angles they would only be in contact over a small area, but would remain in contact for longer as the position of contact shifts along the ridges. So it is logical that the fluctuations in the total amount of instantaneous asperity contact would be of lower frequency and amplitude for a cross-hatched contact topography compared to a single-hatched one. It thus follows that the degree of cross-hatching may affect the frequency and amplitude of acoustic emission from the contact. The cross-hatched contact topography of a rough disk pair replicates that of helical gears which are finished with an axial roughness lay. For these the ridges also cross at an angle.

---

[1] There is no practical reason that prevents one of the disks being flipped on its axis when assembled in which case the direction of roughness curvature of both disks would be similar at the contact. This would give a very different contact topography, but as yet this has not been investigated.

Specific contact topography

At this point it is necessary to distinguish between the effects of the *general* and *specific* contact topography. For a given disk pair the general contact topography does not vary, i.e. it will always consist of cross-hatched ridges and valleys, however the *specific* contact topography may do. The specific contact topography is defined by which ridges and valleys of one disk meet those of the other. It will vary with changes to the assembly orientation and SRR of the disks. Variations to the specific contact topography can affect how the disk surfaces wear.

Consider the case of pure rolling (a SRR of zero), for this the specific contact topography has a period of one disk rotation. An arbitrary circumferential position ‘ ’ on the ‘fast’ disk, will always meet with one arbitrary circumferential position ‘ ’ on the ‘slow’ disk. Thus the ridges and valleys found at will only ever meet with those found at and they will thus wear for this specific combination. Consider running-in: Where a ridge at crosses ridges of it will run-in, however where it crosses valleys of it will not. If the assembly orientation of the disks is then changed then a new specific contact topography results. Position will now meet with a new circumferential position on the ‘slow’ disk and this may initiate new running-in as sections of the ridge at which had previously crossed valleys of now cross ridges of The possibility for new running-in can be expected to diminish as the number of different assembly orientations increases and more and more sections of the ridge at have the chance to run in.

When sliding motion (a non-zero SRR) is introduced the specific contact topography becomes more complicated, It will have a period that is dependent on the SRR and will be *at least* two fast disk rotations in length. Thus even for a single assembly orientation, the position on the fast disk will meet with *at least* two different positions on the slow disk. Figure 3-5 shows an example of how the disks will meet for a SRR of 0.5, a gear ratio of 3/5. For any angular position of the fast disk there are five possible positions of the slow disk, (in Figure 3-5 these are shown as A to E for when the fast disk is at 270°). Thus the specific contact topography has a period of five fast disk, or three slow disk, rotations. In this case a ridge found at will wear for five different positional combinations simultaneously (, , , and). Additionally, an increase in sliding motion may increases the length of each individual combination, i.e. the number of opposing ridges and valleys that a ridge at will meet at each position. It should thus be apparent that surface wear is not only dependent on the general contact topography but also the specific contact topography, and this needs to be considered in the design of experimental procedures.

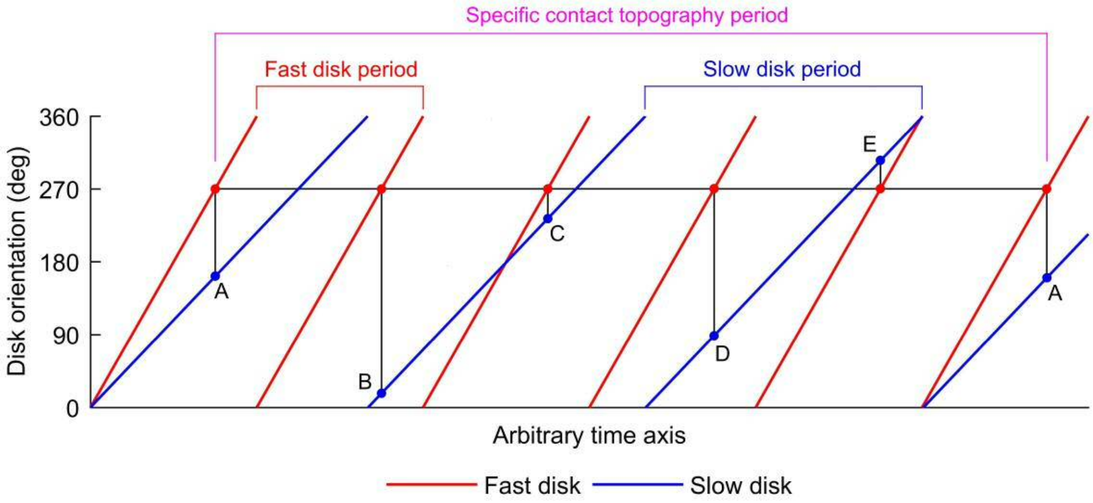


Figure 3-5 Comparison of the orientation of the fast and slow for a SRR of 0.5, a gear ratio of 3/5.

3.4 Contact measurements

This section describes the measurements that can be made on the twin disk rig. The capability for measurement of the AE was developed as part of this thesis, all other measurement capabilities were pre-existing.

3.4.1 Basic measurements

The basic rig measurements include speed, load and temperature. The rotational speed of the fast shaft is measured by an encoder. This, along with the gear ratio between shafts and the disk diameters can be used to calculate the surface velocities of the contact. The shaft encoder also outputs a single pulse per fast disk rotation which is used as a count and to synchronise all measurements.

The contact load is measured by a compressive load cell affixed to the end of the ram. When the ram is on, all the force is transferred to the pivoting housing through the cell. The ram is positioned so that the line of action of its force is coincident with that of the test contact force. Thus the cell directly reads the contact load without the need to correct for any mechanical advantage (leverage). The load cell error can be measured by the method used in [57] which found it to be low enough to disregarded.

The temperature of each disk is measured using a single thermocouple. The thermocouple is located in a small hole drilled 3 mm below the disk's contact surface. (The hole can be seen in Figure 3-2). The thermocouple is typically potted using an epoxy resin. To reduce its susceptibility to electrical noise it is important that it is isolated from the disk. For the experiments presented here this was achieved by pre-coating the thermocouple tip with cured epoxy prior to potting it. The thermocouple of each disk is wired through a bore in its respective shaft to a slip ring which is connected to the rig computer.

Ideally the temperature of the disks would be measured at the contact surface however this is not practical. The thermocouple hole has to be sufficiently distant from the contact surface to prevent affecting its elasticity. Due to the nature of the heat transfer at the disk there will almost always be a small difference between the surface and subsurface (the measured) disk temperature when the rig is operational. The exact difference is not fixed, it depends on the ratios of: frictional heat generated by the test contact; convection to / from the disk surfaces by the lubricating oil / air; and conduction through the disk and shaft. Although in theory it should be possible to use a thermodynamic model of the disks to calculate and correct for the difference it would be a time consuming enterprise and in practice it is easier to simply accept that there is some uncertainty in the surface temperature of the disks.

For this thesis a new thermocouple was added to measure the temperature of the oil supplied to the contact. The thermocouple is located on the nozzle that directs oil into the inlet of the contact. This temperature measurement is ancillary; it is not essential for determining the lubrication conditions at the contact but can be used to reveal something of the thermodynamics of the contact. For example at steady-state conditions the oil inlet temperature is approximately equal to the ambient test enclosure temperature and thus the difference between the disk temperatures and oil inlet temperature reveals the amount of frictional heating at the contact.

3.4.2 **Contact Friction**

The friction at the test contact can be determined by measuring the torsion, and therefore the torque, in the slow shaft. This is done by a strain gauge located in-between the driven gear and the pivoting housing (see Figure 3-1). The strain gauge is self-temperature correcting and is wired through a slip-ring to its conditioning circuitry. The error when stationary has been measured in [57] and found to be reasonably low, less than five percent, for the torque range during typical testing.

The strain-gauge measures the torque induced down-shaft of the driven gear, which is caused by friction from the test contact, bearings and slip rings. Due to the flexible couplings only the test disk and two support bearings are heavily loaded and so the frictional torque caused by other ancillary bearings (not shown on the schematic) and the slip rings is assumed to be negligible. The contact friction can be thus calculated by subtracting the support bearings torque from the measured value and dividing the result by the radius of the disks.

The support bearings torque can be measured by disconnecting the fast shaft at the coupling down-shaft of the motor driven gear. This results in the down-shaft section, which includes the fast disk and its two support bearings, being driven solely by the slow disk which results in the contact operating under pure rolling conditions. In this set-up the strain gauge measurement gives the torque caused by all four support bearings and there is no frictional loss at the contact. It is assumed that each support bearing generates an equal torque and so dividing the measured value by two gives the value for the slow shaft bearings only.

Unfortunately the bearing torque can be expected to vary with load, speed and temperature (lubricant viscosity) and so it needs to be measured for all combinations of these tested. This was done by using a set of torque calibration procedures. These replicated the actual test procedures of this thesis exactly, with the exception that the fast shaft was disconnected, as described above, and a sacrificial set of disks was used. The contact friction during the actual test procedures was thus calculated by looking-up the bearing torque at the equivalent point in the calibration procedure.

3.4.3 Contact Voltage

The contact voltage is used here to provide a *measured* reference against which the specific film thickness calculation can be compared. The techniques used to measure the contact voltage on the twin disk rig were developed by previous researchers and are discussed in detail in [57,119]. A brief description of them is presented here.

To make contact voltage measurements it is necessary that, except via the contact, the test disks are electrically isolated from one-another. This is done by isolating the slow disk from the rest of the rig using non-conductive linkages on the slow shaft and the pivoting housing. A potential difference between the disks is then set up and measured using the divider circuit shown in Figure 3-6. The input voltage and two resistors and are fixed (at least nominally) so that the contact voltage is proportional to the resistance across the test contact . With sufficient direct asperity contact becomes zero (i.e. it shorts) and thus the contact voltage becomes zero. Conversely with no direct asperity contact becomes infinite and the contact voltage reaches a maximum which is dependent on the specific values of , and It is possible for the resistance across the contact to have a transition value (i.e. one between zero and infinity) and thus the same is true of the contact voltage.

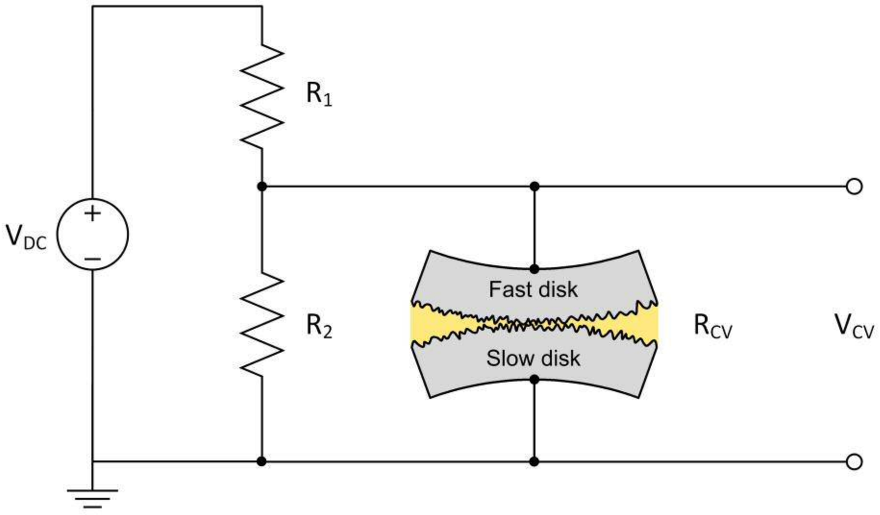


Figure 3-6 Circuit diagram of the contact voltage measurement

For this thesis , and were selected to give a maximum contact voltage of 45 mV. All contact voltage measurements are presented normalised (divided) by this value so that zero represents the minimum, and unity the maximum. The contact voltage signal is sampled at high frequency, 1.25 MHz, and in synchronisation with the fast disk position. The high frequency samples reveal the instantaneous contact voltage, which is used to investigate the rapid fluctuations in the amount of direct asperity contact typical in the mixed lubrication regime. The high frequency samples are also averaged once per fast disk rotation to provide a running ‘mean’ contact voltage. The mean contact voltage indicates the position on the mixed lubrication spectrum.

Previous work by Weeks [57] showed that the contact voltage measurements contained considerable noise in the 100 - 400 kHz frequency range, which was attributed to the rig’s power inverters. To improve the clarity of the instantaneous contact voltage measurements Weeks removed this noise using a low pass filter, specifically a 50 kHz 3rd order Butterworth filter. Despite some improvements to the electrical shielding, similar noise was found in the contact voltage measurements made for this thesis so the instantaneous contact voltage measurements were again filtered. The same 3rd order Butterworth design was used, but this time with a pass band of 30 kHz and a stop band of 40 kHz. The noise and the actual contact voltage had overlapping frequencies so there were no perfect values for the cut-offs. They were selected here so that the results produced clear surface plots. A 30 kHz cut-off corresponds to a surface length cut-off of 27 μm at the minimum rig speed of 200 rpm and 270 μm at the maximum rig speed of 2000 rpm (or 0.04 and 0.4 degrees of rotation). The noise was found to have a negligible effect on the mean contact voltage (indicating the noise had a mean of zero) so this was not filtered.

An example of some contact voltage measurements taken in the mixed lubrication regime is shown in Figure 3-7. The decreasing mean contact voltage indicates a shift in the mixed lubrication spectrum towards more severe lubrication conditions, i.e. an increase in direct asperity interactions. The instantaneous contact voltage, shown for an arbitrary 45° arc of the fast disk, reveals the rapidly fluctuating nature of the contact conditions. Red bands indicate a propensity for significant direct asperity contact, i.e. a point in the specific contact topography which is dominated by one or more aggressive asperities. Conversely blue bands indicate a propensity for minimal direct asperity contact, i.e. a point in the specific contact topography which lacks significant asperities.

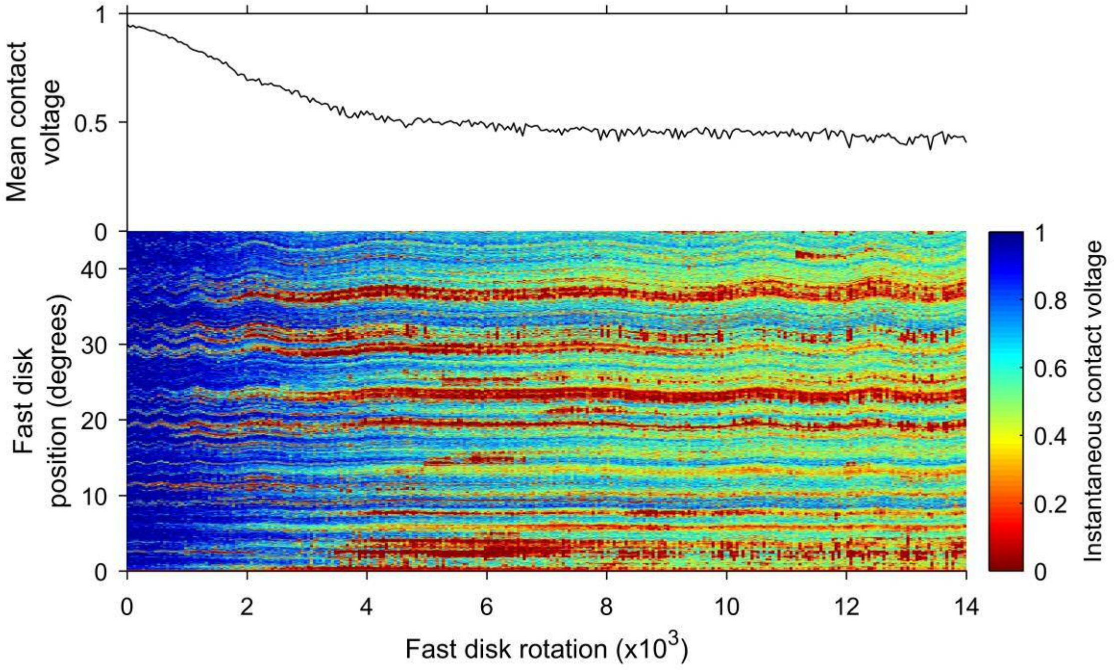


Figure 3-7 Example contact voltage measurement. (A 30 kHz low pass filter was used for the instantaneous contact voltage).

3.4.4 In-situ surface measurements

In this section measurement of the disk surfaces is discussed. As mentioned in the introduction to this chapter this can be done *in-situ*, i.e. without disassembly of the disks from the rig, and at *repeatable positions*, i.e. the same point on the surface can be re-located and measured at any time. In-situ measurements are necessary because, as discussed earlier dis and re-assembly of the disks would alter their specific contact which is often undesirable mid-experiment. Another more mundane reason is that dis and re-assembling the disks takes a long time and so in-situ measurements are more efficient. Repeat position measurements are very useful as they allow the wear history of specific topographic features to be catalogued. This aids understanding of why the general surface wears as it does.

The in-situ measurements are made using a portable 2D profilometer, 3D measurements cannot currently be made in-situ and so their use is restricted to pre and post-experiment. Measurements cannot be made immediately after testing, the disks must first be allowed to cool to the ambient lab temperature to avoid thermal strain error. The lubricant must also be removed from the disk surfaces, this is done using a degreasing solvent. The profilometer cannot be left in place while the rig is running. When measurements are to be made it is placed on a platform, shown on the rig schematic (Figure 3-1). As discussed previously, and as shown in Figure 3-3, the ridges and valleys of the surface topography run in an axial direction so it is the circumferential roughness which is primarily of interest, and which is measured. Grooves in the platform align the profilometer so the measurement will be in a circumferential direction. The platform is affixed to a manually actuated linear stage which is used to control the axial position of the measurement. The axial position is indicated by a dial gauge with a resolution of 10 μm. A vertical stage on the profilometer is used to lower and raise the stylus tip into and out of contact with disk surface. When taking a measurement the profilometer moves the stylus horizontally. Given the vertical range of the stylus: 1 mm, and the curvature of the disks, the theoretical maximum for the chord (horizontal) length that can be measured is ~17 mm. In practice a shorter length is more convenient, and a length of 12 mm was used for this thesis. The stylus used for all the measurements in this thesis had a conical tip with a radius of 2 µm and it is important to remember that the shape of the tip affects the profile measurement. The radius will smooth out sharp drop-offs or steps and the tip will not penetrate into very narrow cracks or overhangs. Thus stylus profilometry cannot detect the entrances of the fine and slanted cracks evident in micro-pitting (see Figure 2-20).

To make repeat position measurements two reference points are needed: an axial and a circumferential one. The latter is given by one or more engraved scores on the down-shaft side of the disk. To use these the disks are simply rotated by hand until the score mark is located at the stylus tip when it is at a known position in the measurement length (e.g. the beginning, middle, or end). For this thesis measurements were made at up to three equispaced scores on both disks. The measurements were positioned with their midpoints at the score lines. Figure 3-8 illustrates the circumferential positioning. It is not especially precise, but it does not need to be as cross-correlation can be used to align different circumferentially measurements precisely after they are made, this technique is discussed later in this section.

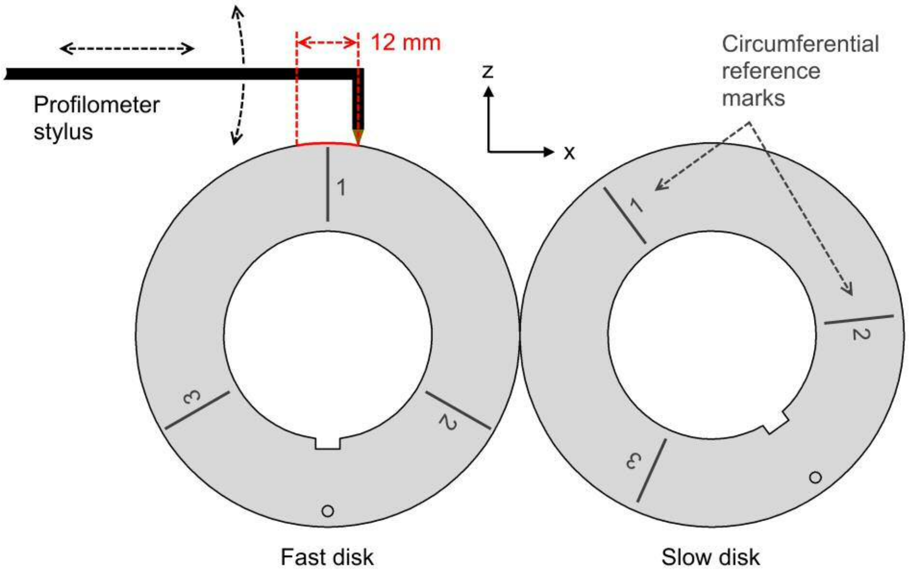


Figure 3-8 Circumferential referencing for profile measurement

Axial referencing is more problematic than circumferential, it needs to be as precise as possible because profiles cannot be axially re-aligned after measurement. The down-shaft disk edge is used as a datum from which all other axial positions are measured. To use the disk edge as a datum its position relative to the stylus must be found. Figure 3-9 illustrates the procedure required to do this using an arbitrary axial profile measurement of an as-ground test disk: The axial position of the profilometer platform is adjusted by eye until the stylus tip is on the disk centre line, position A. It is then slowly moved outwards until the profilometer reads a drop in surface height of 60 µm. This drop is used to indicate the disk edge, position B. The position indicator can then be used to set any axial measurement position. For example, to make a measurement on the disk centre-line the tip is moved back inwards 4.8 mm (half the disk width), to position C in Figure 3-9. When the profilometer is removed from the platform the reference between the stylus and the disk edge is lost and it has to be re-established each time a new set of measurements is made. The uncertainty in axial positioning is at least equal to the resolution of the dial gauge, which is 10 µm. This level of uncertainty may seem small but, as will be shown later, it causes problems when measuring micro-pitted surfaces as the width of these pits are of a similar scale, or less.

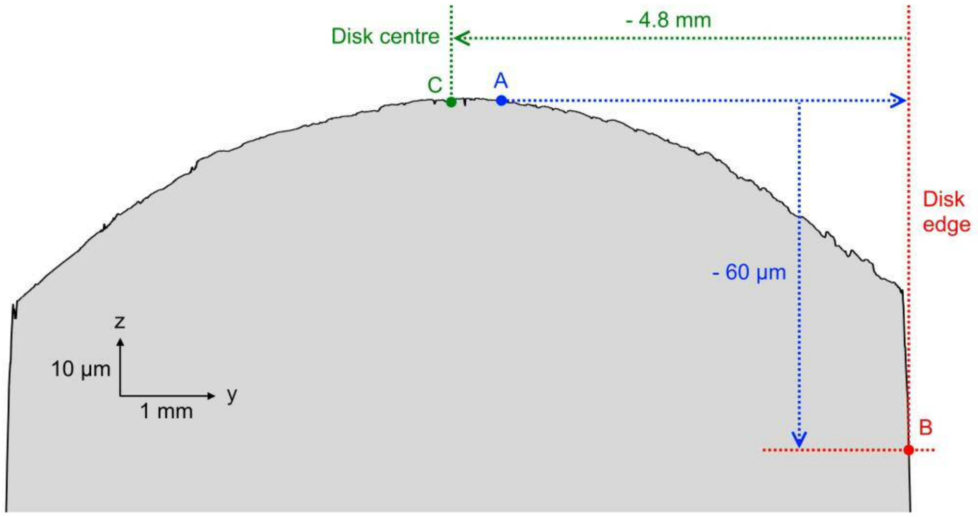


Figure 3-9. Axial referencing for profile measurement

Repeat position profiles are made so that the wear history of specific surface features can be examined. To do this it helps if the profiles can be aligned on a common set of axes. Their axial) position is fixed by the initial measurement position however their circumferential) and vertical) positions can be aligned. This alignment is done on the roughness profiles, i.e. after the form and waviness of the raw measurements has been removed. Their circumferential positions need to be aligned first and this is done using cross-correlation[1]. When analysing the measurements made for this thesis it was discovered that there was some inconsistency in the sample spacing. The profilometer samples the surface heights in horizontal increments that are *nominal* rather than measured. Thus any error between the actual and nominal increment results in a 'stretch' error. This error was found to be inconsistent both within one measurement and

---

[1] This is a standard mathematical technique that can be used to find the lag (horizontal offset) between two similar signals.

between different measurements. The error was systematic (does not have a mean of zero) and so was built up over the length of the measurement. It should be emphasised that despite this the stretch error was very small; it only manifested over long profile measurements. It was subsequently found to be due to a fault with the profilometer which was then serviced. As some of the analysis used later in this thesis requires extremely precise horizontal alignment and all the profiles in this thesis were measured before servicing there was a need to correct the stretch error. A 'windowed' cross-correlation technique was devised to do this and it is described with the aid of Figure 3-10[1].

Figure 3-10 (a) shows two overlapping sections of the same profile. One has been artificially stretched to simulate the stretch error, the scale of the error has been exaggerated for this example. These two profiles can be coarsely aligned using a single cross-correlation of their entire lengths, and (b) shows the result of this. The non-overlapping sections have been removed but because of the stretch error, the profiles are not aligned over their entire length, in (b) they are aligned only near their middle. To remove the stretch, it is necessary to determine the variation in 'lag' (horizontal error) over the length of the profiles. This is done by windowing (sectioning) one of the profiles and cross-correlating each window against the full length of the other 'reference' profile, this returns a lag for each window. (c) shows the lag vs. window position when five windows are used. To correct the stretch error, the lag needs to be determined at each sample point rather than across a window. To do this the lag vs. window position is interpolated using a smooth function, as shown by the solid line in (c). The variation in lag can be complex and so a piecewise function, in this case a cubic spline, has been found to work well. This function is then used to query the lag at each sample point and negate it, i.e. change (to (, which re-stretches the windowed profile so that it is aligned with the reference profile. As a result of this operation the two profiles no-longer have sample points which are coincident, and the re-stretched profile has non-uniform sample spacing. This is undesirable so the final step in the process is to interpolate the surface height of the re-stretched profile at the sample points of the reference profile. (d) shows the alignment of the profiles after the process is complete. Because in this case it has worked so well, it is difficult to tell that there are two profiles overlaid, although minor errors can be observed under magnification.

It should be mentioned that technically the stretch error is not corrected, it is simply made uniform between different profiles. It is not known which has the more accurate horizontal spacing but this is not important as it is the comparison between heights which is of interest. The technique has been described here using only two profiles however, as will be shown later it can be used to align a large set. Its effectiveness depends primarily on the window width: too short and the cross-correlation will not be accurate, too long and the interpolation of lag against position

---

[1] There are other more sophisticated methods that might be used to remove the stretch error such as Dynamic Time Warping but these were not investigated.

will not be accurate. Trial and error was used to determine the best window width for the profiles aligned in this thesis, this was done on a case by case basis.

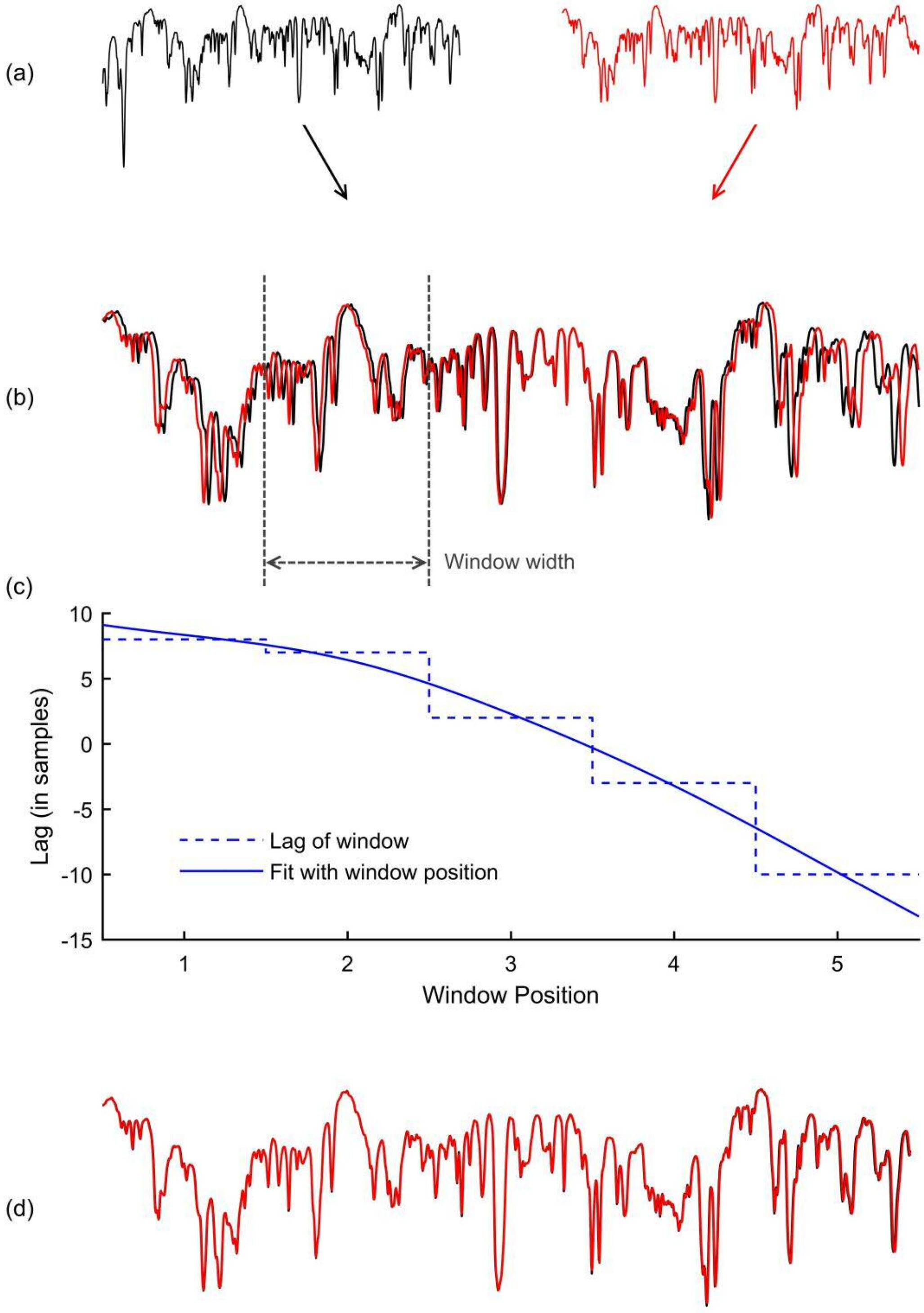


Figure 3-10 Circumferential profile alignment using windowed cross-correlation

Once the profiles have been circumferentially aligned they can then be vertically aligned. Their mean line may change depending on the wear so this cannot be used as a datum. Instead only the deep valleys are used as these are assumed to remain constant for a typical amount of wear. Two profiles are aligned using a uniform vertical offset which is calculated by taking the mean of the differences in their deep valley positions. Micro-pitting can cause problems with this process as it can results in the formation of new deep valleys which need to be excluded from the mean difference calculation. Since the vertical difference between a new deep valley and the prior un-pitted surface is much greater than the vertical misalignment of pre-existing valleys, differences that are too large are simply excluded from the mean difference calculation. The results of the vertical alignment can vary depending on how a 'deep valley' is defined and a how a difference that is 'too large' is defined. However there is no easy way to determine which values give the most truthful vertical alignment. Thus after some calibration experiments it was decided to simply keep the definitions consistent for all alignments performed for this thesis.

3.4.5 **Acoustic Emission**

This section describes how the contact AE was measured. This had not been attempted previously on the twin disk rig so the instrumentation considerations were new for this research. The Cardiff University research group which supported this thesis predominantly uses AE equipment, both hardware and software, manufactured by the Physical Acoustics Corporation (PAC). As it was desirable to use as much pre-existing equipment as possible for this thesis the only hardware considerations actually required were which type/s of AE sensor to use and where it/they should be placed.

It was decided to place two sensors as close to the contact as possible, i.e. on one of the test disks. There were two linked reasons for this: Firstly it was anticipated that the amplitude of the AE from the contact would be at the low end of the detectable range. And secondly it was suspected there might be noise from remote AE sources, the most likely of these being the support bearings. Placing the sensors on a test disk minimises the test contact to sensor attenuation whilst maximising the remote noise-to-sensor attenuation, i.e. it uses proximity to help preferentially detect the contact AE. For pragmatic reasons it was decided to place sensors on the fast disk rather than the slow; the slow shaft support bearings are closer to the disk so space restrictions are greater, and the need to maintain electrical isolation of the slow disk for the contact voltage measurements would have complicated placing sensors on it. Part way through this research a third sensor was added to the support bearing housing down-shaft of fast disk. The purpose of this sensor was to assess the level of noise from the support bearings and to determine if any AE from the contact was being transmitted through the bearing contacts. Unfortunately there was not time to examine the measurements from this sensor in depth.

The two sensors on the fast disk used very different coupling methods in order to compare their effectiveness. The methods are illustrated in Figure 3-11 which shows a cross section through both shaft axes and the contact. The components have been simplified but their outer dimensions, i.e. their space envelopes, are to scale. One sensor was held static by an arm that positions it just off the up-shaft disk face, and next to the contact. The disk and sensor faces were thus in relative sliding motion. They were not supposed to touch, the signal being transmitted from disk to sensor face by an oil film which bridges the gap. The other sensor was clamped to the down-shaft face of the test disk and rotated with it. This sensor was wired through a bore in the shaft to the slip ring. The two methods, and sensor choice, are now discussed in detail.

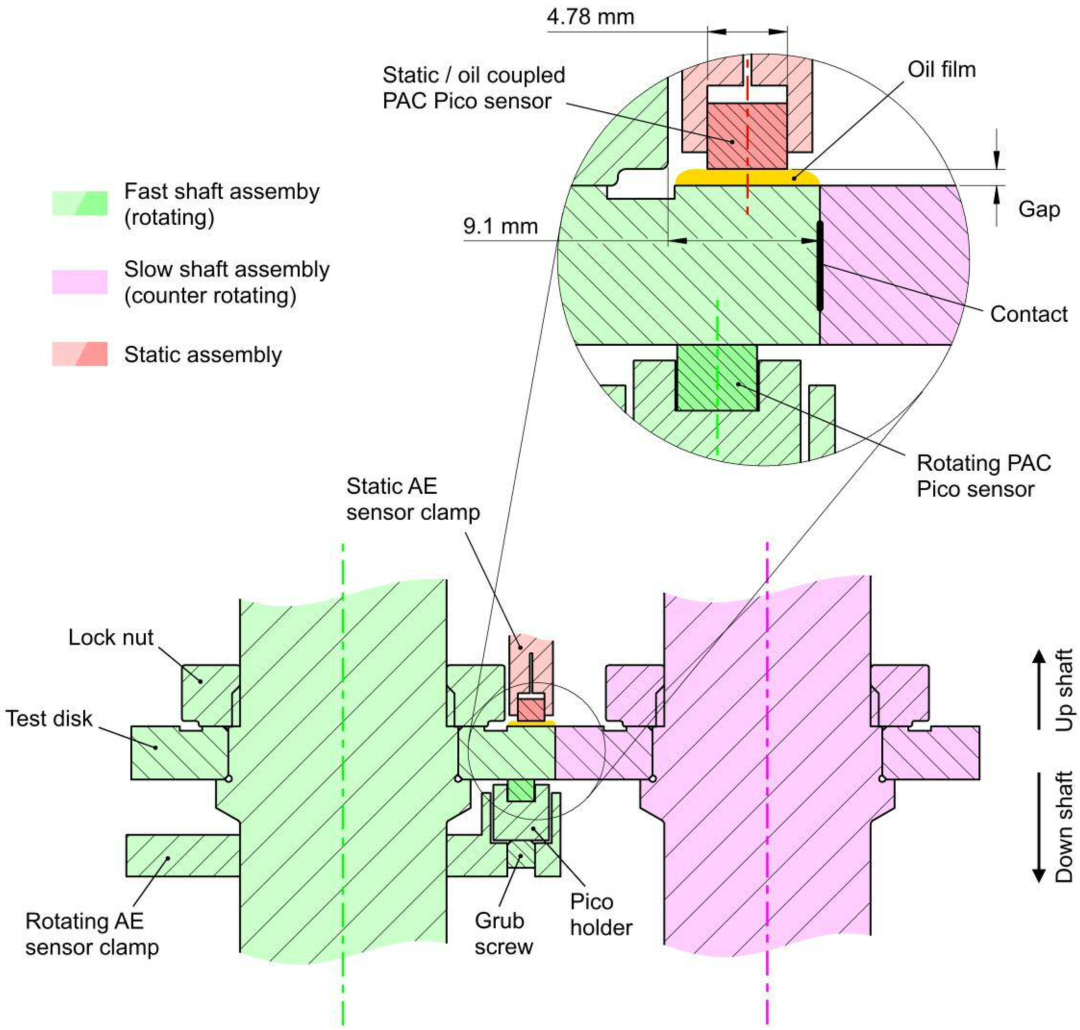


Figure 3-11 AE sensor placement

Static / oil coupled sensor placement

The space within which the static sensor can be placed is extremely limited. Each test disk is axially restrained between a shoulder on its respective shaft and a lock nut. The lock nut and a groove in the disk face restricts the width of the annulus where the static sensor can be placed to approximately 9 mm. To allow for the practicalities of holding the sensor whilst maintaining a suitable clearance from the lock nut and slow test disk, the diameter of the sensor face must be considerably less than 9 mm. As the spectral content of the contact AE was initially unknown it was desirable to use a sensor with a broadband sensitivity in order to be sure of detecting it. However because of the space limitation described, the static sensor choice was dictated by suitability of size rather than frequency response. The PAC ‘Pico’ sensor type was judged to be of the most practical size. These sensors are cylindrical and have a face diameter of just under 5 mm.

An arm was specifically made to hold a Pico sensor static at the position shown in Figure 3-11. The arm was attached to the test enclosure wall and consists of several beams connected by slots that allow limited 3D translation of the sensor to set its position. It was not possible to easily measure the disk-sensor gap (which is heavily exaggerated in Figure 3-11). Instead shims were used to set a nominal gap during assembly. To do this the bolts that fix the axial position of the sensor were loosened and shims of the appropriate thickness placed in the gap. The sensor was then butted against them by hand, the bolts fastened and the shims removed. This method was rather rudimentary as due to both bending of the arm, and elasticity of the shims, the final gap width depended on the force used to butt the sensor. Thus the gap was only expected to remain consistent in-between assembly procedures. An oil jet was directed into the gap to bridge it, the jet was positioned so that the oil would be drawn into the gap by the surface motion of the disk (by Couette flow). Basic calibration tests revealed that as might be expected the signal attenuation between the disk and sensor face increased with the gap width. From an attenuation perspective it would have been ideal to set the smallest possible clearance between the disk and sensor. Unfortunately because of likely fluctuations in the gap width during operation, caused by thermal expansion and vibration, it was not possible to determine a suitable minimum clearance. For this thesis a nominal gap of 0.2 mm was set. This was chosen based on results from preliminary calibration tests however it was not certain that this gap was sufficient to always maintain complete disk-sensor separation.

Rotating sensor placement

The rotating sensor was fixed on the side of the fast disk using a specifically designed holder which attached to the fast shaft. The sensor was located in a recess and was clamped against the disk side by tightening a grub screw. During clamping a silicone sealant was placed in-between the sensor and disk faces. The majority of this was squeezed out under the clamping force, it was simply used as a gap filler to improve the interface quality. The space restrictions were less severe on the down-shaft disk face so the recess was designed to accommodate cylindrical sensors with a diameter up to 11 mm. Smaller sensors could have been used by placing an insert in the recess to reduce its diameter and length. Despite the capability for a larger sensor it was decided to use a Pico sensor so that the two placement methods could be compared using a consistent sensor type.

Discussion of placement method

Of the two placement methods the rotating sensor had by far the greatest precedent. This type of arrangement has been often used in the laboratory when it is desirable to have a sensor as near to the contact as possible. The main benefit of using this method is that a consistent and high quality interface can easily be made between the sensor and the rotating component it is measuring from. There are two obvious drawbacks of using this method. Firstly, that the transmission of the signal from the sensor to the external equipment is complicated. Slip rings can be used, as in the case of this work, but these are costly and difficult to implement outside the laboratory. In theory wireless transmission could be used but this comes with its own set of problems such as how to supply power and ensure a reliable transmission. Secondly, the distance and therefore the transmission path from the contact to sensor is not constant. In Figure 3-11 the rotating sensor is shown next to the contact, but in half a fast disk rotation it will be opposite it, and the signal will have to travel around the disk to reach it. This is unlikely to be a problem for small components, but for large components the variation in distance may be such that the sensor only recives a good signal when it passes near the contact, i.e. periodically. In essence the problem is, that a signal from a rotating AE sensor is likely to become somewhat modulated by its own rotational frequency.

This thesis represents the first known use of a static and explicitly oil-coupled sensor to measure the AE from a rotating component. It has been attempted here because it has almost converse benefits and drawbacks to a rotating sensor. The transmission of the signal from static sensor to conditioning circuitry is trivial, they can be directly connected; and the distance between the sensor and contact is constant. However, the interface between the sensor and the rotating component it is measuring from is comparatively complicated and much less likely to be consistent. The interface is a sliding contact in-itself and unless carefully designed may generate its own AE thus obfuscating the AE from the contact of interest. It was not required that the sliding interface used in this thesis, which is of a rudimentary design, would perform as well as that of a static one, rather that it would prove the feasibility of the concept which could then be developed in the future.

Hardware specifics

This section discusses the specific hardware used to measure the AE and condition the signal starting with the Pico sensors. These are available in two types: one with better lower frequency sensitivity and one with better high frequency sensitivity. The lower frequency type was used for this work as it has a similar frequency response to other sensors being used to investigate the AE from rotating machinery at Cardiff University. Pico sensors of this type have a nominal operating frequency range of 200 – 750 kHz with resonance at 250 and 550 kHz depending on the excitation method. Figure 3-12 shows the calibration curves for the specific sensors used to measure the test contact AE.

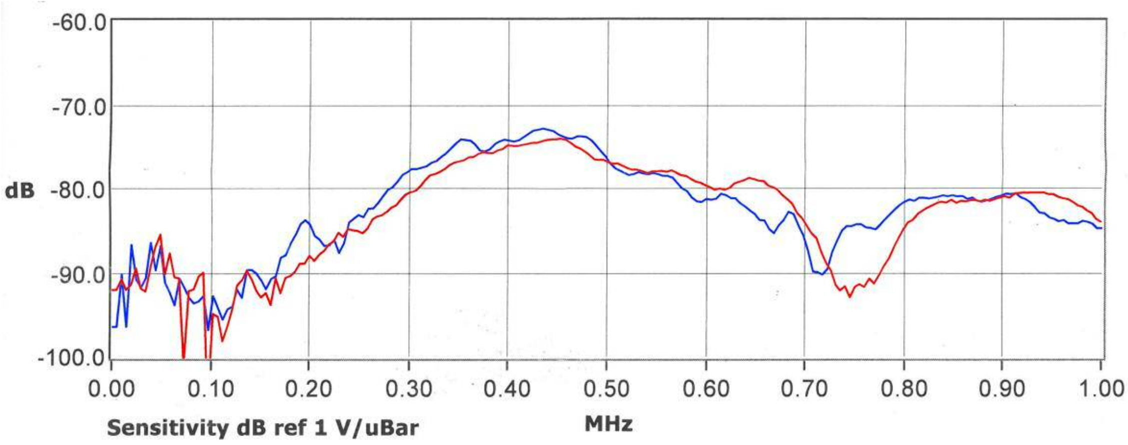


Figure 3-12 Calibration curves for the two PAC Pico sensors used to measure the AE. (Red) the rotating sensor and (blue) the static sensor.

Both sensors were wired to pre-amplifiers (pre-amps) with selectable gains of 0, 20 or 40 dB and 20 kHz high pass filters. A gain of 40 dB was used for all experiments presented here. It was not anticipated that the signals would be subject to interference or large attenuation *downstream* of their pre-amp so the specific gain was not considered important, however keeping it consistent was. The pre-amp for the rotating sensor was connected downstream of the slip ring so any noise introduced by this was also amplified. It may be preferable to place the pre-amp upstream of the slip ring to avoid this, unfortunately the off-the-self pre-amps available were too large to be placed on the shaft.

The pre-amplified signals were connected to a computer with hardware and software specifically designed for conditioning and recording AE signals. The hardware can perform high speed analogue to digital conversion (ADC), i.e. continuous sampling, of the signal and can further filter it. The maximum continual sample rate is 10 MHz, and this is the maximum for all channels in use, not per channel. The hardware also has an input for a trigger signal so that sampling can be started by an external source. The software has many inbuilt acquisition and post-processing modes specifically to help interpret AE signals, however these were designed primarily for structural health monitoring and are of little use for monitoring of rotating machinery. For this thesis the AE computer was set-up so that all that was required of it was to record the AE signals at a pre-set ADC rate and for a pre-set duration upon repeated triggers from the rig computer. (The acquisition methodology is discussed in greater detail shortly). The AE signals were not assessed in real time.

3.4.6 **Sample timing.**

This section describes the sampling methodology. Because the acoustic emission and contact voltage were expected to vary cyclically with the specific contact topography it was beneficial to synchronise sampling with this period. Synchronisation is implemented on the twin disk rig using a pulse from fast shaft encoder repeated once per rotation. The pulse does not directly initiate a sample, which would be rather restrictive, instead it is used to time a sampling program.

Figure 3-13 illustrates the type of sampling program used in this thesis in block form. The samples are recorded in two different forms, as a mean (time-average), and as a continuous signal. The load, friction and temperatures are recorded as the mean values; the acoustic emission as a continuous signal; and the contact voltage as both. A sample is triggered when the program detects a pulse from the fast shaft, and it lasts for a pre-set duration. After the sample has been taken some overhead is required before the program is ready to detect the next start pulse. In Figure 3-13 it is set to record both the continuous and mean samples every 5th fast disk rotation using a length equal to 1.1 fast disk rotations. For all other rotations it is set to only record the mean samples for a shorter length of 0.75 fast disk rotations. The program cannot start a new sample until the previous one has finished and so depending on the length of the preceding sample and the length of the overhead (which fluctuates with processor availability) some pulses will be overshot and samples missed. The fast disk rotation count is not affected by this so missed samples do not affect the synchronisation.

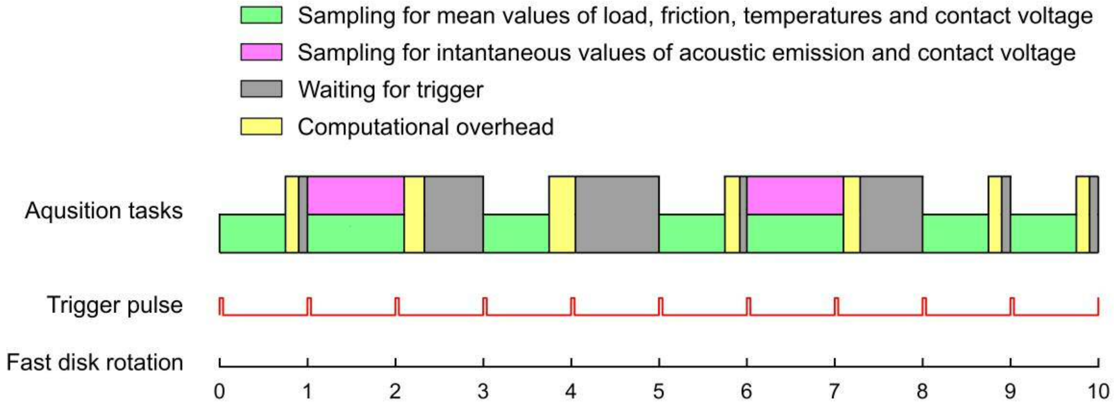


Figure 3-13 Typical sampling program

The program described above guarantees *intra*-program synchronisation, i.e. synchronisation within one continual run, however, it does not necessarily guarantee *inter*-program synchronisation, i.e. synchronisation between two or more runs. Inter-program synchronization requires control of how the program is started. At the start of this thesis the program could only be started using a basic ‘single-pulse’ method. This simply sets the first shaft pulse detected as the start of the count, i.e. the zeroth rotation, from which the sampling interval is then controlled. However for a non-zero SRR there will be multiple fast shaft pulses within a single period of the specific contact topography, thus there are multiple positions within the period when the program might start. Using the single-pulse method for a SRR of 0.5 there five different positions, and so for the program of Figure 3-13 there is only a 20 percent chance that synchronization will be maintained from one test to the next.

During this thesis a more robust ‘dual-pulse’ starting method was developed so that inter-program synchronization could be guaranteed. This required the addition of an encoder to the slow shaft so that pulses were generated from both shafts, once per respective rotation. When the dual-pulse program is started it repeatedly compares the time difference between the most recent pulse from the slow shaft and the next from the fast shaft. There will be pattern of unique time differences which repeats cyclically with a period equal to the contact topography. Figure 3-14 is an example of a pattern of unique time differences for a SRR of 0.5, in this case there are three differences per cycle. Once the program has reliably established the pattern of time differences it simply sets the zeroth fast disk rotation to coincide with the occurrence of the minimum time difference and then continues sampling as previously described. Because the pattern of time differences and the specific contact topography have the same period, inter-program synchronization is thus guaranteed.

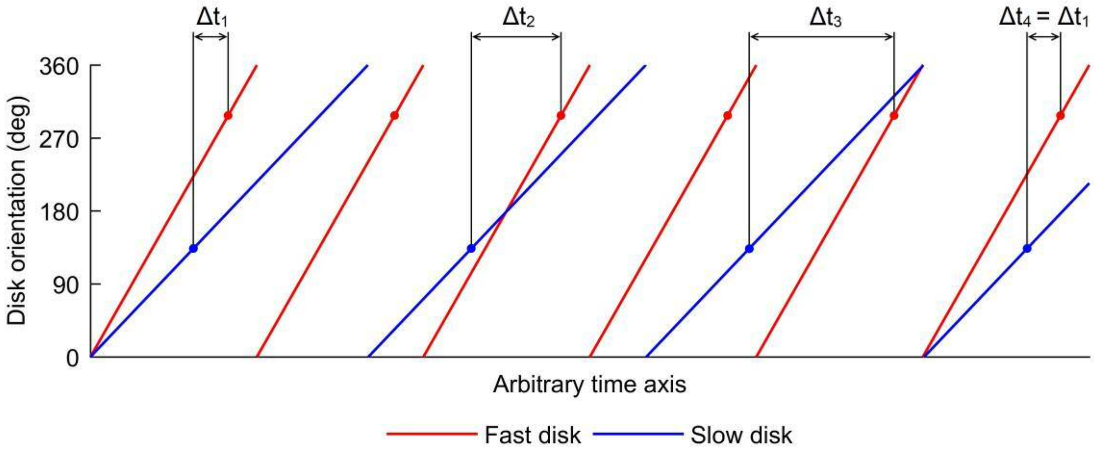


Figure 3-14 Time differences between a specific slow disk orientation and a specific fast disk orientation

4 Variable film thickness experiment

4.1 Introduction

The fundamental aim of this thesis was to explore, in detail, how the lubrication regime, i.e. the proportion of fluid film to boundary lubrication, affected the AE from the gear-like contact of the twin disk rig. This required an experiment that would, a) vary the lubrication conditions in an easily controllable and measureable way, and b) provide confirmation that the observed AE was from the test contact. As discussed in the theory (Chapter 2) the lubrication conditions of an EHL contact will vary with changes to either or both of: the uniform film thickness, and the surface roughness. This chapter presents an experiment where large variations in the uniform film thickness were deliberately induced to modify the lubrication conditions, and changes to the surface roughness were only intended to be negligible and incidental. Varying the uniform film thickness, rather than the roughness, was preferable as the process is more controllable and only requires a single disk pair to study all possible variations. Additionally the uniform film thickness can be measured continuously, i.e. whenever the contact is operating, whereas the surface roughness cannot.

Two pairs of disks were tested using the variable film thickness procedure: a rough (as-ground) disk pair, and a smooth (superfinished) disk pair. The test on the rough disk pair was the most important and is presented first. The purpose of this was to generate a range of conditions within the mixed regime, i.e. conditions where *both* the uniform film thickness and amount of asperity contact varied. Broadband AE observations were then studied to identify components that had the most plausible relationship with the amount of asperity contact. This relationship was used to assess which AE sensor interface: rotating or static, was the most effective; and finally the form of the relationship was examined in detail.

A limitation of the rough disk pair test, on its own, was that discrimination of AE components related to asperity contact was based on *assumptions* of what was, and was not, a plausible relationship between them. The purpose of the test on the smooth disk pair was primarily to provide a means of discrimination that did not require these assumptions. This test replicated the changes in uniform film thickness of the rough disks, but without the associated changes in asperity contact. Comparison of the measurements from each disk pair provided a secondary method of identifying AE from asperity contacts. These comparisons are presented at the end of this chapter.

4.2 Test procedure for the rough disk pair

For convenient reference Table 4-1 presents a summary of the procedure used for testing the rough disk pair, it includes information on the disks, contact parameters, test schedule, and measurement settings.

Table 4-1 Summary of the test procedure

Disks	
Surface finish	Ground
Operational history	Significant previous testing with following contact parameters: Nominal load: 1460 kN Slide roll ratios: 0.25 and 0.5 Oil bath temperatures: Ambient to 85°C Fast disk Speeds: 200 to 2000 rpm

Contact parameters		
Constants	Slide roll ratio, [Gear ratio]:	0.5, [5/3]
	Specific contact topography (assembly orientation of disks):	Fixed
	Load calibrated at 100°C:	1460 kN
	Maximum Hertzian contact pressure at calibrated load:	1.2 GPa
	Hertzian contact dimensions at calibrated load:	0.78 x 3.0 mm
	Oil:	OEP-80
Variables	Fast disk speed:	200 to 2000 rpm
	Oil bath temperature:	Ambient to ~ 110°C

Test schedule				
Test stage	*Fast disk speed (rpm)*	*Oil bath temperature (°C)*	*Duration of test contact*	*Uniform film thickness (μm)*
Running-in	200	110	4 minutes	0.06
1	300	Ambient – 110	Not pre-determined	0.7 – 0.07
2	500			1.0 – 0.10
3	1000			1.6 – 0.17
4	1500			2.1 – 0.22
5	2000			2.5 – 0.27

Sample timing		
Start up:	Single pulse method. Intra stage synchronisation only.	
Means:	*Data:*	Speed, load, temperatures, contact friction and contact voltage.
	Frequency:	Once per fast disk rotation.
	Duration:	Three quarters of a fast disk rotation.
Continuous signals:	*Data:*	Acoustic emission, contact voltage.
	Frequency:	Every 25th fast disk rotation.
	Duration:	1.1 fast disk rotations.

AE measurement settings	
ADC:	5 MHz per sensor.
Pre-amp:	40 dB.
Filtering:	20 kHz high pass.

Profile measurements	
Frequency:	At beginning and end of test schedule and in-between each stage.
Positions:	Six positions per disk: three equispaced circumferential positions, and two axial: the disk centre-line and near the edge of the contact path.
Length:	12 mm.
Filtering:	0.8 mm cut off length Gaussian filter.

The test procedure was designed to induce continuous variation in the proportion of fluid film to boundary lubrication, for contact between two as-ground, i.e. rough surface, test disks. As was discussed in the introduction, the intention was to test with variable uniform film thickness. Many parameters affect the uniform film thickness, of these, speed and temperature were chosen as variables, and the others as nominal constants. Speed and temperature were chosen for reasons of consistency and convenience. On the twin disk rig they are the only parameters which can be varied whilst maintaining a consistent specific contact topography[1]. They are also the only ones which can effect both a continuous and predictable change in the uniform film thickness. However, because of the predicted *direct* effect of speed on the AE (the propensity for an increase in speed to cause an increase in AE amplitude unrelated to changes in the amount of asperity contact), the procedure did not use continuous speed variation, rather, there were five tests at different discrete speeds. For each of these tests the film thickness was continuously reduced by heating the lubricating oil from the ambient temperature of the lab to approximately 110°C.

Of the constant parameters only the load and SRR were easily selectable, the others, e.g. contact geometry and oil type, were fixed by design / the inconvenience of changing them. A load corresponding to a maximum Hertzian contact pressure of 1.2 GPa was tested. This value is at the lower end of the range typically tested on the twin disk rig, and was selected so that the contact area would be relatively small, and thus allow the effects of short-length fluctuations in the specific contact topography to be studied, i.e. to keep area-averaging low[2]. The SRR was set at 0.5, this value is in the middle of the range of current options for the twin disk rig, and simulates the portion of a gear tooth meshing cycle away from the pitch point, i.e. towards the tooth root / tip where micro-pitting typically occurs.

The ground disk pair used for this experiment were not new, they had been used for extensive prior testing and were well run-in up to the selected contact pressure of 1.2 GPa. Nevertheless, it was considered likely that, because the temperature was expected to reach new highs, the procedure would induce thinner oil films than previously tested, potentially initiating a new phase of running-in. To preclude the possibility of this affecting the results, there was a dedicated running-in stage at the start of the schedule. The contact parameters of this stage were set to induce a thinner film, i.e. more aggressive lubrication conditions, than any to be subsequently tested. However, given the propensity of direct asperity contacts to cause wear, it was expected that some was inevitable in the main procedure. It was predicted that the slower the speed of a test, the more wear prone it would be, due to lower average film thickness and thus greater asperity contact. For this reason the tests were sequenced in increasing speed with the expectation that the propensity for wear would reduce rather than increase as testing progressed. The summary in Table 4-1 shows the test sequence and the predicted range of the uniform film

---

[1] Arguably the load affects the specific contact topography by changing the width of the contact, i.e. which roughness features pass through it.

[2] As it turned out short-length effects were not studied, primarily due to time constraints.

thickness for each stage. The range of speeds and temperatures tested represented the full range capable of the rig at the time.[1] For the running-in the rig was pre-heated to the maximum test temperature and the load was applied for four minutes. For the rest of the test stages the rig was started cold by refraining from turning the oil bath heating on until immediately after the load was applied. The test was then run until the bath reached a temperature of ~ 110°C.

For the SRR of 0.5 the specific contact topography had a cyclic period of five fast disk rotations. So the continuous AE and contact voltage signals were sampled every 25th fast disk rotation, which ensured synchronisation with the topography. (The maximum synchronous sample rate, every 5th fast disk rotation, was not used to avoid generating unwieldy quantities of data). The sample lengths were always equal to the duration of one fast disk rotation, which meant they were proportional to speed. At this point in the research, the third AE sensor had not been added to the bearing housing, and so the ADC was set to 5 MHz per sensor, the maximum for two channels. The speed, load, temperature, contact friction and contact voltage means were sampled once per detected fast disk rotation, and over a period of three quarters of the rotation. The dual-pulse starting method had not been developed by this point in the research so the continuous signal sampling was only synchronised *intra* test stage, not *inter* test stage.

In-situ and repeat-position profile measurements were made before and after each test stage. This was done so that the effects of any wear could be accommodated when calculating the specific film thickness. Profiles were measured on the disk centre-line, and at an offset of 1.28 mm on one side. This offset corresponds to the point at which the Hertzian pressure is half that on the centre-line. These centre-line and side measurements were made at three equispaced circumferential positions on each disk. The roughness profiles were extracted using a Gaussian filter with a cut-off of 0.8 mm. This cut-off was chosen as it was the close to the length of the contact in the direction of motion, i.e. the width of the ellipse. It was reasoned that a smaller cut-off might remove variations that affected the contact conditions.

The coefficient of friction, uniform film thickness and specific film thickness were calculated for each fast disk rotation using the mean samples. The coefficient of friction was calculated as the ratio between the contact friction, (after calibration described in Section 3.4.2), and the measured load. The uniform / central film thickness was calculated with the CDDT formula (Equation 7) using the mean temperature of the two disks, the measured load and the speed. The specific film thickness was then calculated using this and the mean of the pre-test and post-test (RMS roughness) of each disk.

---

[1] In fact to reach the maximum temperature of 110°C in reasonable time a second temporary heater had to be added to the oil bath to double the heating power as the rig is not well insulated.

4.3 Roughness and wear

This section presents the results of the roughness measurements and using them, an assessment of the impact of wear on the lubrication conditions. Note: in the context of this analysis wear is defined as any irreversible modification to the roughness features of the surface, whether having a beneficial or adverse effect on the subsequent lubrication conditions.

Figure 4-1 presents (RMS roughness) of each profile for each point in the test sequence. First consider the measurements from before and after the short pre running-in. It can be seen that the decreased at the centre-line of the contact, but not at its edge, this indicates that there was some running-in, but that it was confined to the centre of the contact. This not wholly unexpected as, during calibration tests (not presented in this thesis) it was noted that the edges of the contact ran-in faster than the centre; likely because transverse leakage of oil caused more severe lubrication conditions at the sides. It is thus reasoned that no side running-in was observed in this experiment because the sides were already run in for the test conditions whereas the centre was not. Although the only purpose of the pre running-in was to condition the surfaces ready for the main experiment, the evidence of some running-in *despite* significant operational history highlights that running-in is not necessarily confined to the *first* incidence of contact.

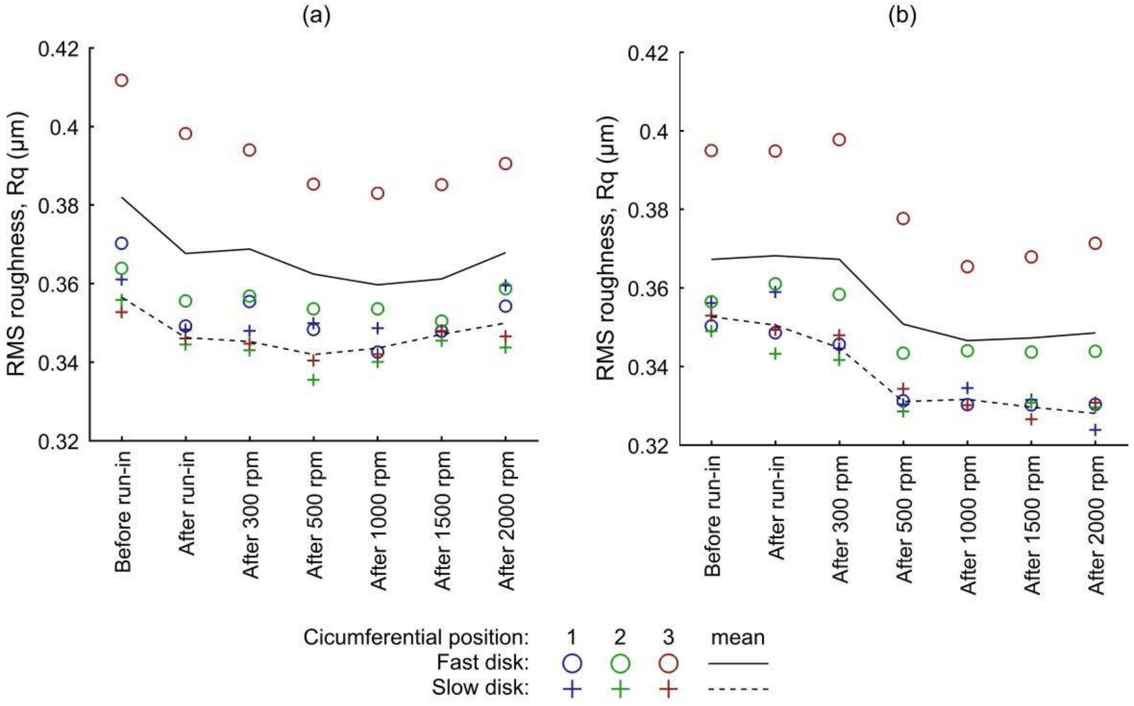


Figure 4-1 of the ground disk pair for, (a) the centre-line, and (b) the edge of the contact path.

Now considering the results from the main experiment, i.e. all test stages after running-in. For the centre-line Figure 4-1 (a) the at each circumferential position remained relatively constant for both disks. All were within a range of ±0.009 μm, and there is no obvious trend, either increasing or decreasing, that would indicate significant wear. In contrast for the measurements made at the edge of the contact path Figure 4-1 (b) there is a small but definite decrease in after testing at 500 rpm. The decrease occurred at all three measurement positions on both disks and the did not subsequently recover; this is a good indication of wear. Again, for reasons of side leakage, wear limited to the contact's edge is not surprising, but what is surprising, is the timing of the wear. It seems to be confined to one test mid sequence, and there is no obvious reason for this. It may be due to the competing effects of average film thickness, and the number of cycles per test stage. Over the test sequence both increased but the former should have reduced propensity for wear whereas the latter should have increased it. It is thus possible that the maximum propensity for wear was at an intermediate stage in the sequence.

As well as probable wear Figure 4-1 also indicates something of the circumferential variation in roughness. For every point in the test sequence, RMS measurements at all but one of the six circumferential positions are spread over a range of approximately ±0.01 μm. However all measurements made at position three on the fast disk fall well above this, i.e. it is an outlier position. Although the was significantly greater than at the other positions it does not necessarily imply significantly more asperity contact. Because running-in acts to remove prominent asperities but has negligible effect on prominent valleys it is probable that the greater is due to a preponderance of outlier valleys in position three. Given the spread of the results, it is unlikely that sampling at only three circumferential positions, which together represent only 15 % of the total circumference, was sufficient to confidently gauge the disks' average roughness. In other words, it is possible that had more or different circumferential positions been measured then their *mean* might have shifted. However this is of little consequence here, because regardless of any circumferential variation in roughness, the fixed measurement positions ensured *consistency,* which was more important for this experiment.

Although the specific film thickness calculation only required the the aligned profiles were also examined in order to assess the stability of the roughness *topography*, i.e. to check for wear not indicated by the . Figure 4-2 presents the aligned profiles from the edge of contact path at position three on the fast disk. This position has been presented as its showed the greatest variation over the test sequence. The profile measured before the running-in is not included. The first thing to note is the generally well rounded nature of the asperities which is typical of well run-in surfaces. The profiles also generally overlay well confirming no major wear over the test sequence. However there are numerous but small localised deviations, (some are highlighted by black arrows), where some profiles have a valley and others an asperity. This behaviour is indicative of a surface that has undergone or is undergoing micro-pitting fatigue. Unfortunately it is difficult to visually assess *when* a particular micro-pit formed using 2D profile alignment, as it is not possible to distinguish between true pit growth and the *appearance* of pit growth due to axial alignment error. This latter gives the illusion of healing, or fluctuating, pits, as highlighted by magenta arrows. Nevertheless a qualitative assessment of the entire profile length does suggest more asperities in place of pits at the start of the test sequence (as evidenced by a preponderance of blue asperities). This suggests that micro-pitting occurred at some point *during* the test sequence and, coupled with the results, with greatest likelihood during the test at 500 rpm.

In summary both the and realignment of the roughness profiles have shown that some subtle wear did occur during in the test procedure. However because it only effected small changes in the it is concluded that the use of a simple pre and post test average did not introduce significant error into the specific film thickness calculation. The implications of the identified wear will be discussed after the oil film thickness variation has been presented in the next section.

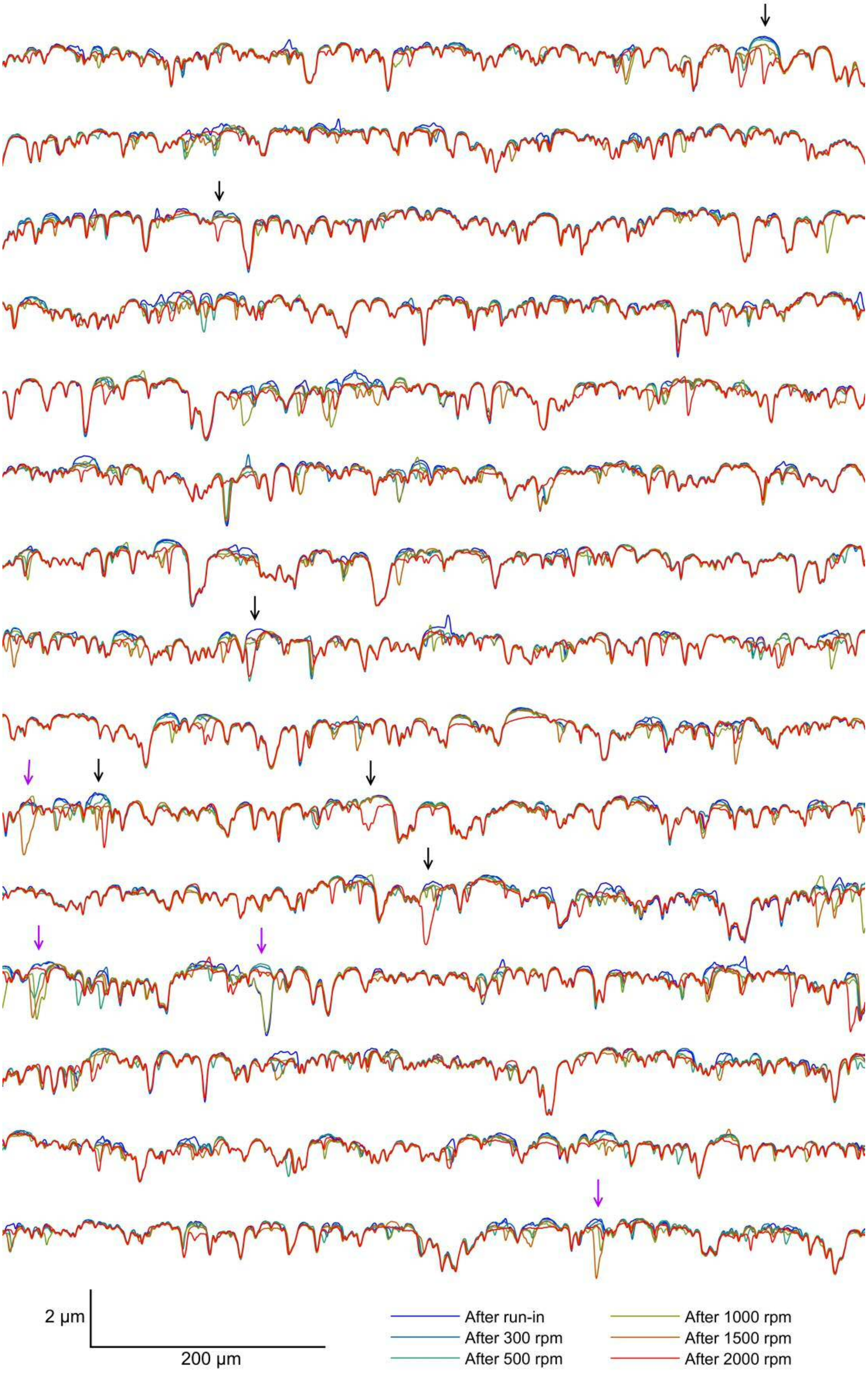


Figure 4-2 Aligned profile measurements from the contact edge of the fast disk at circumferential position three. The profiles are from continuous 12 mm measurements, they have been sectioned and vertically stacked for efficient presentation. The disk material is below the traces.

4.4 Temperatures, uniform film thickness and load

Using the temperature, uniform film thickness and load, this section describes the change in lubrication conditions over individual test stages, and between them. Figure 4-3 shows the chronological variation, of temperature, uniform film thickness, load and coefficient of friction for the entire test sequence. (The coefficient of friction is not discussed here, but is presented here for future reference).

4.4.1 Temperature

Considering the temperature first. Due to the interplay of heating from both the oil bath and the contact, the temperature rise of the disks has some quirks, however it is broadly of a similar, and tapering, form all test stages. The response for the slowest test is now explained in detail. When the load is applied, at zero minutes, there is an immediate and rapid increase in the disk temperatures due to contact heating. During this period the lubricant temperature remains relatively steady and below that of the disk as the start-up response of the oil bath heater is much slower. The rate of increase in disk temperatures starts to drop as the limited effects of contact heating cause them to quickly approach a new, marginally hotter, equilibrium. However at this point, about 2 minutes into the test, the temperature of the oil, and thus the test enclosure, begins to increase significantly which causes the disk temperatures to start to increase again. For a time the lubricant temperature matches that of the disks, but as the test enclosure heats and approaches equilibrium, the small amount of contact heating at the disks causes their temperatures to again become slightly elevated above that of the lubricant.

For the tests at faster speeds the contact heating is more pronounced, and as a result the initial temperature increase does not begin to tail off before the effects of oil heating take over. Additionally the offset between the lubricant and disk temperatures in the latter part of the stage is greater. But because the majority of heating was from the oil heater, which had constant power, all test stages took approximately the same amount of time to reach the cut-off temperature of approximately 110°C.

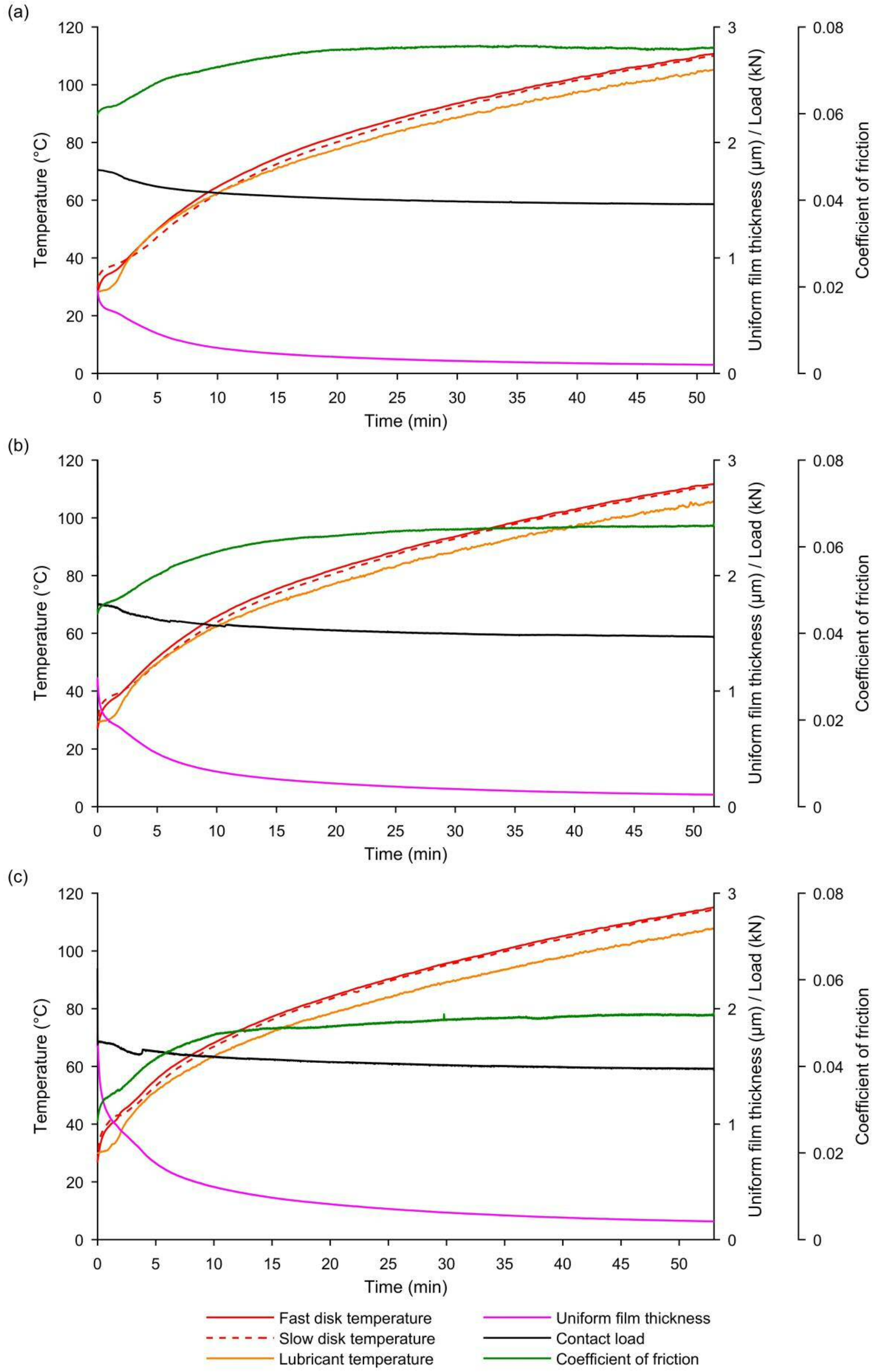
4.4.2 Uniform film thickness

Now consider the uniform film thickness. Over each test stage the decrease in this was almost entirely temperature driven and consequently the uniform film thickness has an inverse-like relationship with temperature. However the rate of change, i.e. the gradient, of the uniform film thickness decreases more dramatically than that of the temperature. This is caused by the non-linear relationship between the oil viscosity and temperature. At the start of a stage, the gradient of the uniform film thickness is relatively steep as the effects of *rapid* temperature change at *low* temperatures compound. Conversely at the end of a stage the gradient is quite shallow, almost zero, as the effects of *slow* temperature change at *high* temperatures compound. Thus for each stage, most of the change from the maximum to minimum uniform film thickness occurred during a short period at the start. Comparing the results of different test stages it can be seen that as the speed increased so did the minimum, maximum and range of the uniform film thickness.

4.4.3 **Load**

Considering the load. Despite this being nominally set as a constant it can be seen that there was a small decrease over the course of testing. Whilst not desired, this was an inevitable consequence of the increase in lubricant temperature, as discussed in Section 3.3.1. Thus, as with the uniform film thickness, the majority of the decrease occurred at the start of the test. To calculate the uniform film thickness the measured, rather than the nominal, load was used, so no error resulted. And it is worth noting, such is the insensitivity of the uniform film thickness to load, that had the nominal value been used instead, the maximum relative error would have been less than 2 %.

Whilst the load variation had negligible effect on the film thickness it will have had an effect on the contact area. The load decreased similarly for all test stages, by ~ 15 % from start to finish. Using Hertzian analysis this corresponds to a reduction in contact area of ~ 10 %, and a reduction in width of ~ 5 %. This reduction in area may have decreased the probability of asperity contact independently of the change in uniform film thickness. The implication of this is discussed once the relationship between AE and specific film thickness has been introduced.


(a)
(b)
(c)
Temperature (°C)
0
20
40
60
80
100
120
Uniform film thickness (µm) / Load (kN)
0
1
2
3
Coefficient of friction
0
0.02
0.04
0.06
0.08
Time (min)
0
5
10
15
20
25
30
35
40
45
50
Fast disk temperature
Slow disk temperature
Lubricant temperature
Uniform film thickness
Contact load
Coefficient of friction

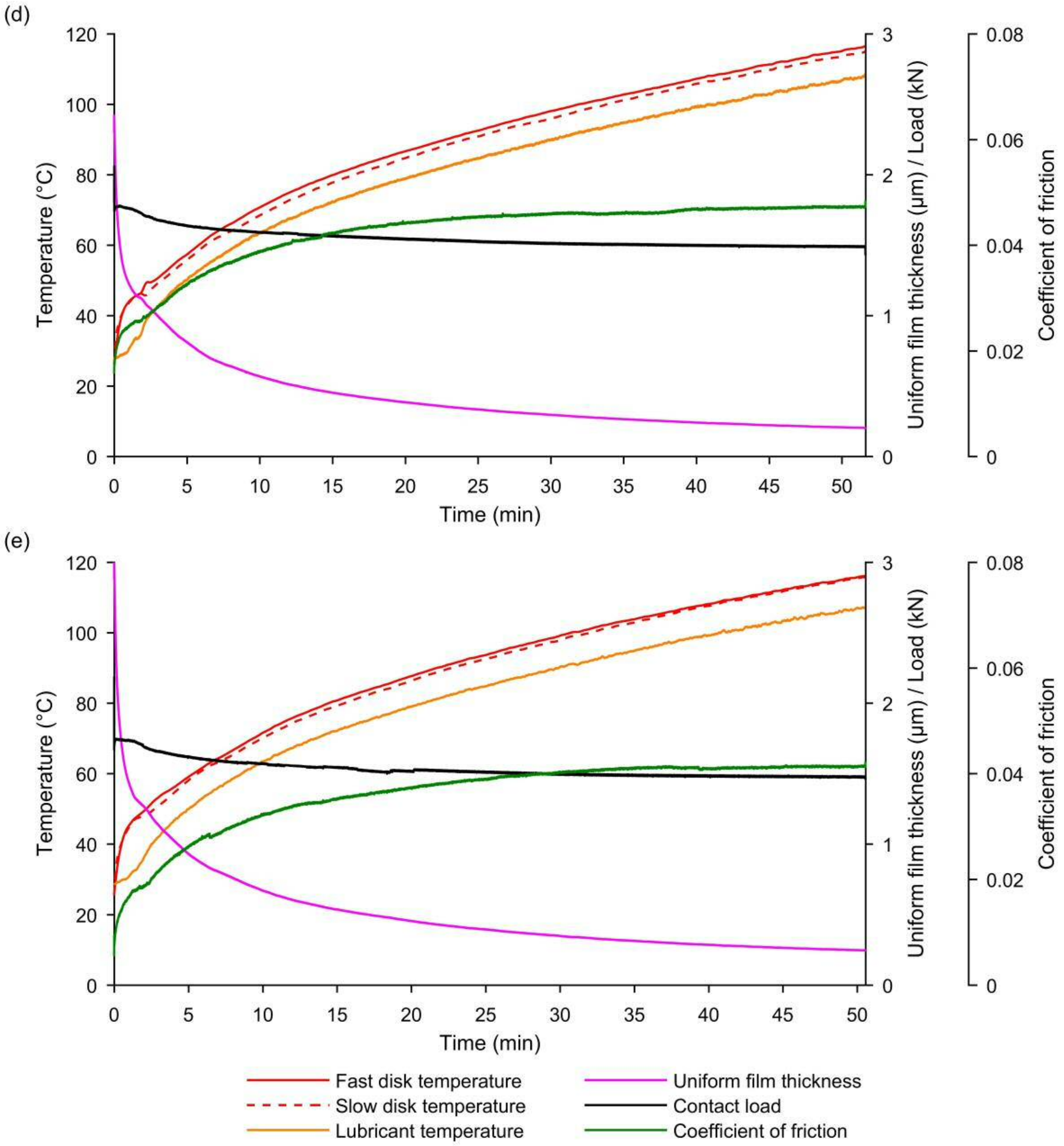


Figure 4-3 Temperatures, load, friction and uniform film thickness vs. time for fast disk speeds of (a) 300, (b) 500, (c) 1000, (d) 1500 and (e) 2000 rpm.

4.5 Roughness and uniform film thickness

Now that the scale of changes in both uniform film thickness and (RMS roughness) have been introduced, their relative influences on changes of the specific film thickness, i.e. the lubrication conditions can be assessed. The specific film thickness is simply the ratio of the uniform film thickness and the , and thus a comparison of their *range*s reveals the influence of each. As already discussed, the most significant change in the occurred during the 500 rpm test stage, so this is used as an example. For this, the at the edge of the contact path dropped by ~ 0.015 µm for both disks. In comparison the drop in uniform film thickness, was ~ 0.9 µm, a change almost two orders of magnitude greater. The implication, of this is that, the drop in had a relatively negligible influence on the specific film thickness, thus it is reasonable to treat it a constant and the surfaces as steady-state. It follows for this experiment the probability of asperity contact was only significantly affected by the changing oil film thickness, rather than wear, which was intended.

4.6 Contact voltage vs. specific film thickness

In this section the relationship between the mean contact voltage and specific film thickness is presented. Figure 4-4 shows, for each test stage, a scatter plot of the mean contact voltage per fast disk rotation vs. the specific film thickness. The data points form curves induced by the continuous change of film thickness. Both temperature and time increase, albeit non-linearly, with *decreasing* specific film thickness, i.e. as the curves are read from right to left. As well as the plots for each individual test stage, there is one showing all combined, which uses a semi-log scale to emphasize change at low specific film thickness.

The contact voltage measurements demonstrate that both mixed and full film lubrication conditions were generated by the test procedure. Considering all test stages collectively: As the specific film thickness decreased from its maximum of ~ 6, to ~ 1.5 the contact voltage did not drop, instead remaining near its maximum possible value. This indicates a relatively steady and minimal amount of asperity contact, i.e full film, or close to full film, conditions. But as the specific film thickness decreased below ~ 1.5 the contact voltage behaviour changed dramatically, dropping to mid-range values at an increasingly steep rate. This indicates a rapid increase in the amount of asperity contact, and high sensitivity to changes in the specific film thickness, i.e. a transition into mixed lubrication conditions. Because higher speeds (at otherwise equivalent conditions) induce thicker films, the high speed stages tested further into the full film regime, and conversely the low speed stages further into the mixed regime. But even at the slowest speed and highest temperature the contact voltage did not approach its minimum possible value (zero) which implies that full boundary conditions were not attained.

The transition from near full film to mixed lubrication conditions at a specific film thickness of around 1.5 to 1 is consistent with previous work on the twin disk [57] however not with some other authors' expectations. As discussed in Section 2.8 they have described a specific film thickness of less than one or two as indicative of the full boundary regime but this is not the case here. This inconsistency highlights the need to reference the specific film against a direct measure of the lubrication conditions.

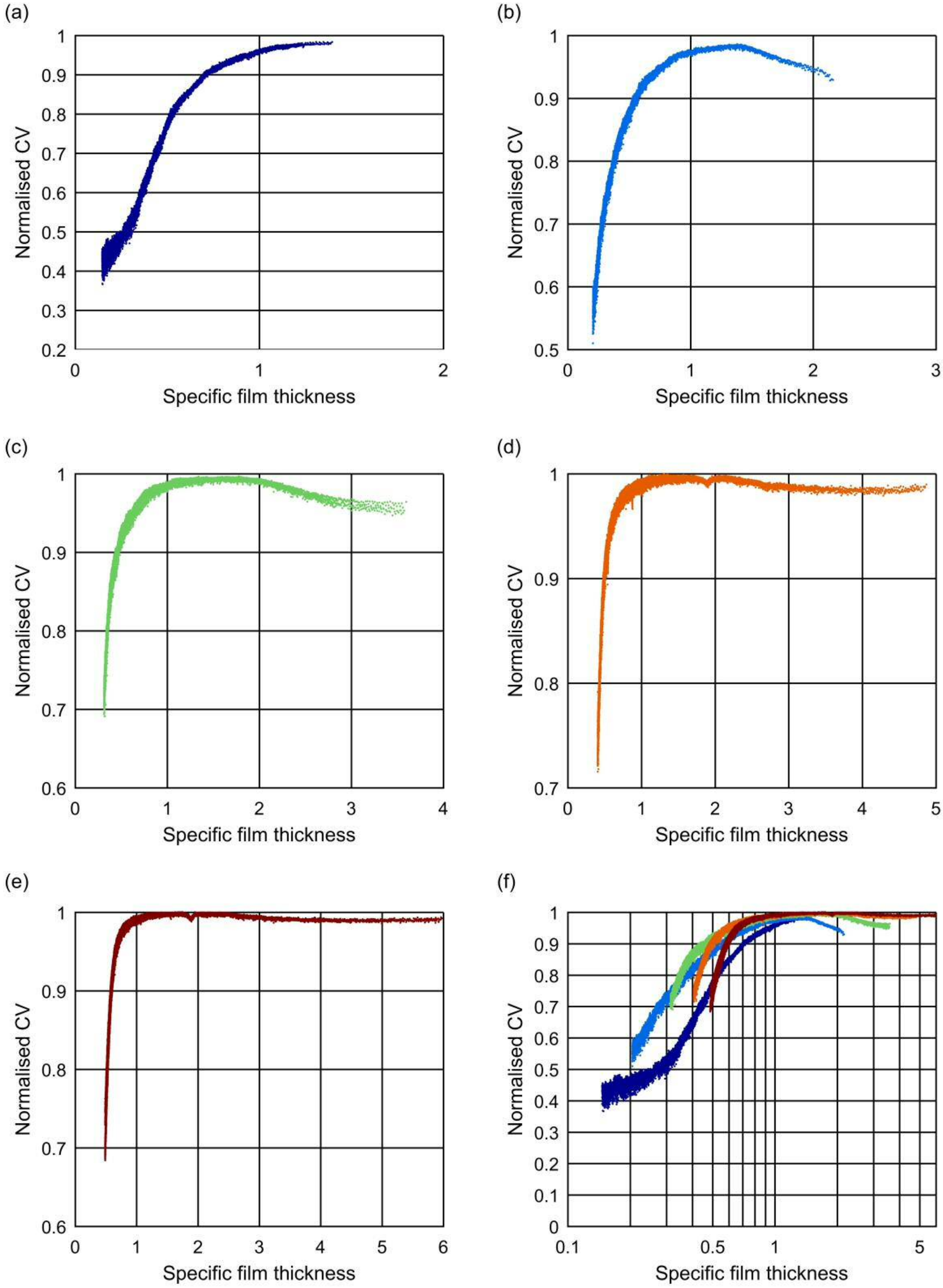


Figure 4-4 Contact voltage vs. the specific film thickness for (a) 300, (b) 500, (c) 1000, (d) 1500 and (e) 2000 rpm. Also (f) the results for all speeds on a single, semi-log, plot. Note: the different colours correspond to the different speeds and are used to differentiate between them on the combined plot.

Two interesting details of the contact voltage measurements are now discussed. The first is that as the specific film thickness decreased, and before the transition from full film to mixed lubrication, there were periods when the contact voltage increased, i.e. it had a negative gradient, which implies decreasing asperity contact. This occurred in all test stages except the slowest (which did not have a significant period in full film conditions), but the phenomenon became decreasingly prominent with increasing speed.

The decreasing asperity contact occurred *despite* decreasing specific film thickness so some other contact variable must have been responsible. There are various possible candidates, such as the reduction in contact area over each stage (as discussed in Section 4.4.3) or the removal of prominent asperities by micro-pitting (as discussed in Section 4.3). Alternatively it is possible that the EP additives in the OEP-80 oil caused progressive growth of tribofilms or changes to the surface chemistry which increased the electrical resistance of asperity contacts. If this was the case there must have been some mechanism that ‘reset’ the surface conditions in-between test stages. Possible candidates may have been the degreasing of the disks to take surface measurements, time dependent chemical reactions (e.g. oxidation), or even a ‘scouring’ away of the protective coatings caused by the more aggressive lubrication conditions at the end of each test. The phenomenon of increasing contact voltage despite decreasing specific film thickness was observed again in the experiment presented in Chapter 5 and is discussed in Section 5.6. The results of this experiment suggest the phenomenon is due to a start-up effect. The phenomenon is worthy of further investigation but was beyond the scope of this thesis.

The second interesting detail of the contact voltage measurements is that the drop in contact voltage upon entering the mixed regime was different for each test stage. Figure 4-4 shows that as the speed increased, the transition from full film to mixed lubrication appeared to become increasingly abrupt and its initiation increasingly delayed. This may have been a genuine effect of the changing speed for, as discussed in Section 2.6 and shown in Figure 2-11, the speed affects the film thickness distribution, with faster speeds causing a more prominent constriction and pressure spike and this may have affected the rate at which the transition occurred. The micro-pitting wear identified in Section 4.3 could also have played a part in changing the rate of transition. If the micro-pitting occurred progressively over the test series, preferentially removing outlier asperities and increasing the height uniformity, then the transition should be expected to become more abrupt. Finally the different transitions may simply be due to error in the specific film thickness calculation, the most likely source of which is related to the temperature. The transitions were generated by different combinations of speed and temperature. (For example a specific film thickness of 0.5 occurred at disk temperatures of ~ 60 °C for the slowest test but ~ 115 °C for the fastest. Thus any *variable* error between the actual contact temperature and the measured value, and there must have been some, would result in inconsistencies when comparing conditions of constant specific film thickness but different temperature.

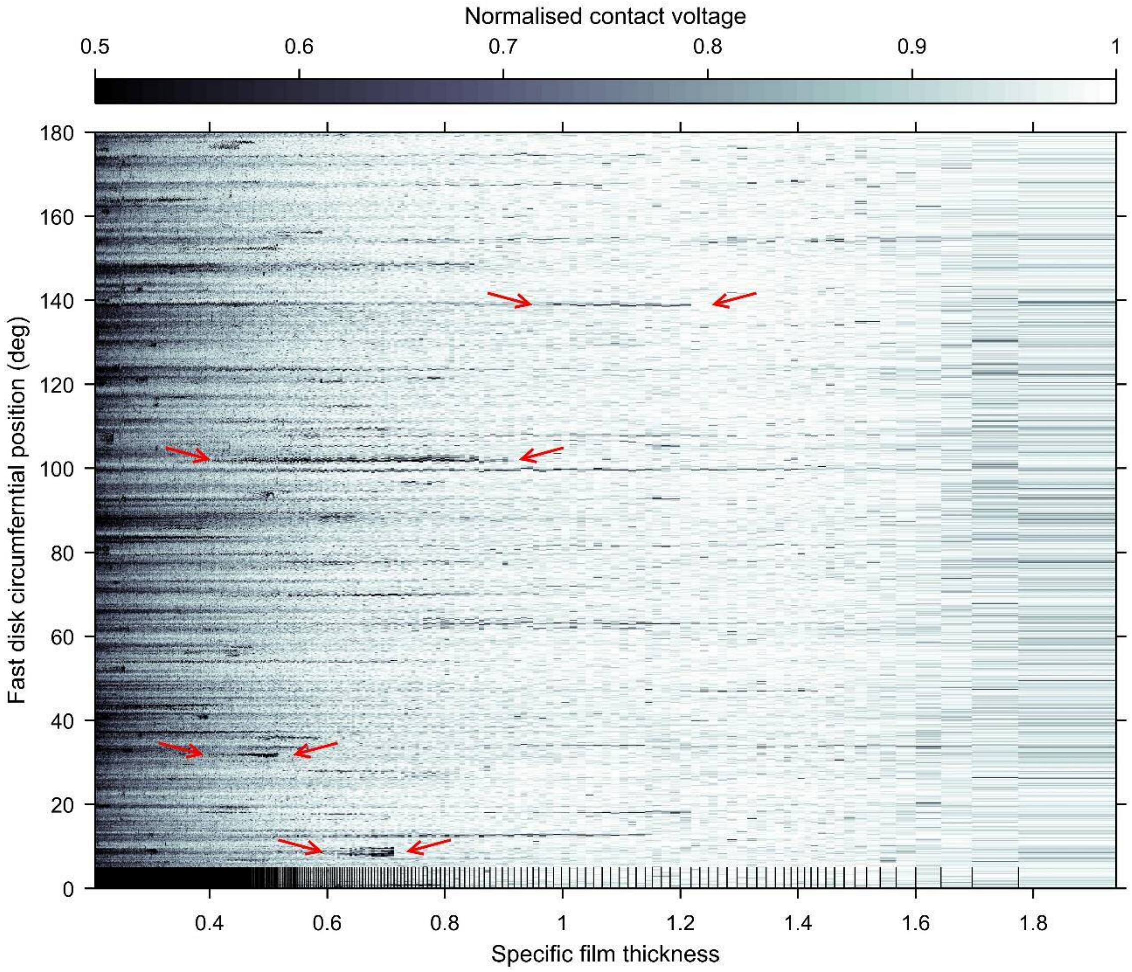


Figure 4-5 Instantaneous contact voltage for half the fast disk circumference vs. the specific film thickness, for the 500 rpm test stage. Each vertical column represents one waveform sample. The column edges are shown at the bottom of the plot. Note: the contact voltage (tone) scale has been truncated at 0.5 to highlight changes in the upper half of the measurable range, thus pure blacks can represent any value from 0 to 0.5. (A 30 kHz low pass filter was used for the instantaneous contact voltage).

The instantaneous contact voltage measurements are now considered. Figure 4-5 shows, for test stage two (500 rpm), the instantaneous contact voltage vs. the fast disk circumferential position vs. the specific film thickness. This test stage has been presented because a large proportion of it was in the mixed regime and it also generated the interesting initial increase mean contact voltage (unlike test stage one). It can be seen that at low specific film thickness there is fine horizontal banding indicating continual rapid fluctuations in asperity contact which are characteristic of the mixed regime. As the specific film thickness increases the banding becomes increasingly 'washed out' and there are only sporadic instances of significant asperity contact (grey / black lines), this indicates lubrication conditions near to full film. Unfortunately, at relatively high specific film thickness, where the initial increase in mean contact voltage occurred, there were only a very limited number of instantaneous contact voltage samples[1] and these do not reveal anything more about the phenomenon (this was also the case for the other test stages).

---

[1] Remember the instantaneous, or continuous signal, contact voltage was only recorded every 25th fast disk rotation whereas the mean contact voltage was recorded for every detected fast disk rotation.

Figure 4-5 reveals a new facet of the contact conditions not evident in the mean contact voltage results. This is, that at some circumferential positions the contact voltage *did not* increase smoothly and monotonically over time, i.e. with decreasing specific film thickness. Rather, there were spikes and transient increases as evidenced by the black speckles and bands that do not extend to the left-hand extremity of the plot. Several of these are highlighted by red arrows. These spikes and transients suggest that there was some type of surface modification that resulted in asperities with either temporarily increased prominence or temporarily increased electrical resistance. Trapped, or adhered, debris could explain the former, and changes to tribofilms the latter. Whatever the mechanism, the transients indicate that during testing the surfaces were not as stable as the before-and-after profile measurements might imply.

In summary. The contact voltage measurements demonstrate that the lubrication conditions ranged from full film to mixed and they provide a reference for the specific film thickness. But more than this they show evidence of multiple contact mechanisms (potentially tribofilms and wear debris) that affect how asperity contacts increase as the film thickness decreases.

4.7 AE vs. specific film thickness

This section presents the investigation into the relationship between the AE and the specific film thickness using the measurements from the rotating sensor. (A comparison between sensors is presented later in Section 4.8).

4.7.1 RMS of the broadband AE

The RMS of broadband AE (BAE_{RMS}), was used as a starting point due to its ease of calculation and frequent use in literature. The BAE_{RMS} was calculated over the full length of each AE sample (one fast disk rotation) and it represents the average amplitude of all frequency components from 20 kHz: the pass-band of the pre-amp, up to 2500 kHz: the Nyquist frequency[1] of the ADC. Figure 4-6 shows, for the rotating sensor measurements, the BAE_{RMS} vs. the specific film thickness.

The results from the five test stages form a set of distinct constant speed curves. The direct effect of speed on AE amplitude can clearly be seen. There is a dramatic increase in the scale of the BAE_{RMS} response as the speed increased, and for all lines of constant specific film thickness, high or low, the AE amplitude increased with speed.

For the higher speed tests, the BAE_{RMS} broadly increased in an exponential-like fashion with decreasing specific film thickness, thus suggesting that at least some of the observed AE in these tests was from asperity contacts. However the increase was not very consistent, either within or between test stages. As the specific film thickness decreased there were some localised jumps, or kinks in the BAE_{RMS} which do not represent the continuously increasing amount of asperity contact in the mixed regime (as demonstrated by the contact voltage measurements, Section 4.6). More importantly, for the slowest speed test the BAE_{RMS} as measured by the rotating sensor did not show a sustained increase at all, despite this test having the most dramatic increase in the amount of asperity contact. This demonstrates that the BAE_{RMS} is not a reliable indicator of the amount of asperity contact.

---

[1] The Nyquist frequency is always half the ADC rate.

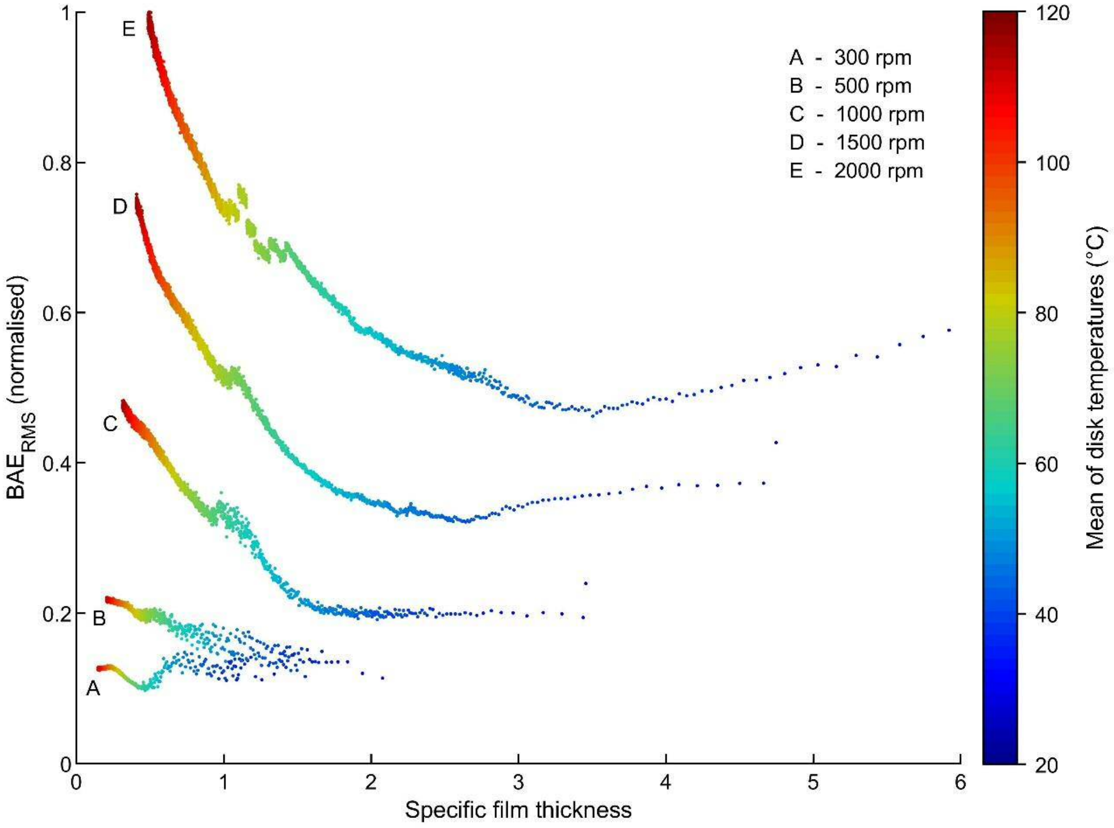


Figure 4-6 BAE_{RMS} (RMS of the broadband AE) vs. the specific film thickness, for the rotating sensor measurements. The BAE_{RMS} scale has been normalised between zero and the maximum measured value.

4.7.2 **Frequency analysis**

This section presents an analysis of the spectral characteristics of the broadband AE measurements. These were broadly similar between for both sensors and so, for brevity, only the spectral content of the rotating sensor is presented in detail.

Figure 4-7 presents the spectral content of some of the AE samples from test stage one. Each of the three panes, (a) (b) and (c), is a bar plot of the spectral content at different parts of the test stage. The panes show the average spectral content of a set of samples taken at (a) relatively high (for test stage one) specific film thickness, pane (b) at midrange specific thickness and pane (c) at low specific film thickness. Specifically, the averages were calculated from samples in the following normalised specific film thickness bins: (a) 0.8 – 0.9, (b) 0.4 – 0.5 and (c) 0 – 0.1, where 1 and 0 represent the maximum and minimum specific film thickness generated in test stage one. (The highest specific film thickness bin, 0.9 – 1, has not been presented as there were very few samples in this). To calculate the average spectral contact in each set of samples individual FFTs were performed on each separate sample and the results then averaged. Each FFT was performed over the full length of an AE sample and with a Hanning window [123] applied to mitigate the effects of spectral leakage. Various different window designs were trialled, including rectangular, but the window design was not found to affect the broad characteristics of the signals, it only affected fine details not of interest for this thesis.

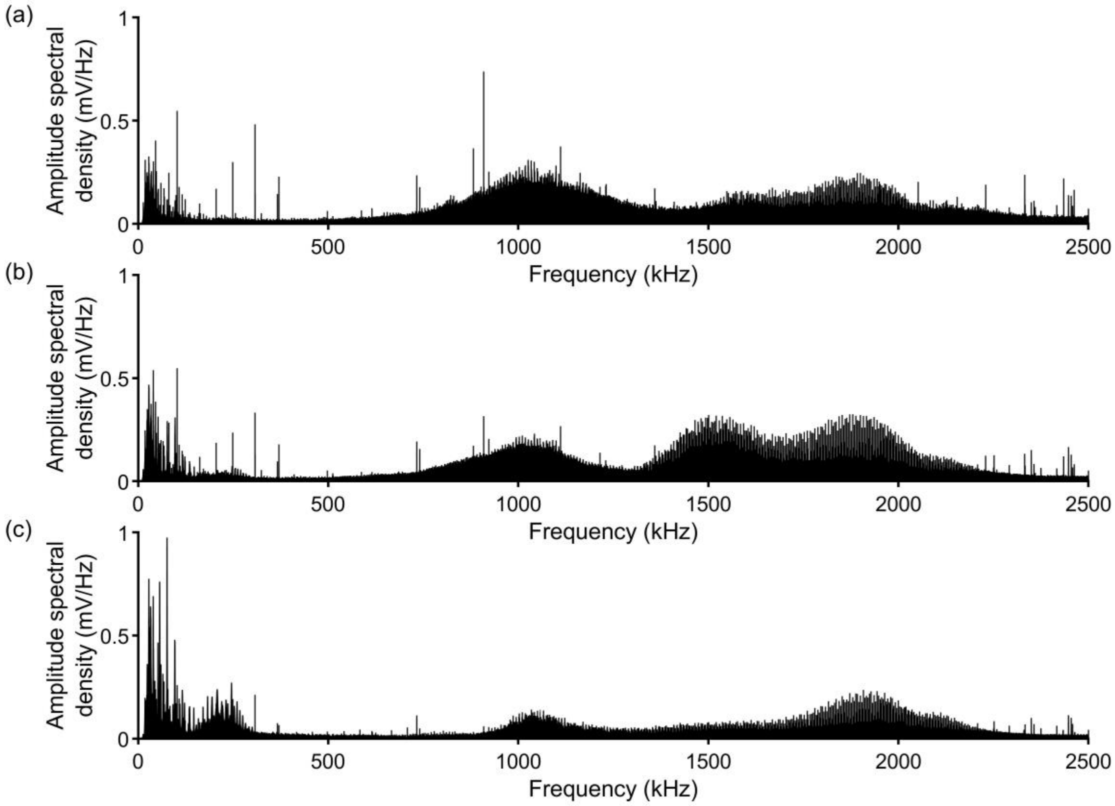


Figure 4-7 Bar plots of the average spectral content of some AE samples from test stage one (300 rpm). For samples at (a) relatively high specific film thickness, (b) midrange specific film thickness, and (c) low specific film thickness. The frequency content was calculated using an FFT with a Hanning window applied.

For comparison with test stage one, Figure 4-8 presents the spectral content of some AE samples from test stage five, calculated using the same method. The absolute values of the specific film bins are not equivalent between the two figures due to the different range of specific film thickness generated by each test stage. Additionally, the spectral resolution is different due to the different lengths of the samples, it is 4.8 mHz for test stage one, and 19 mHz for test stage five.

Figures 4-7 and 4-8 highlight the complex and varied spectral content of the AE samples. For both test stages there are numerous isolated spikes of activity overlaid with some seemingly more broadband behaviour. And there is a mixed response to the decrease in specific film thickness with activity at some frequencies increasing and some decreasing (more on this shortly). There is also a dramatic difference between test stages, with the first having relatively significant amounts of activity above ~ 500 kHz and the last having relatively negligible amounts of activity above ~ 500 kHz.

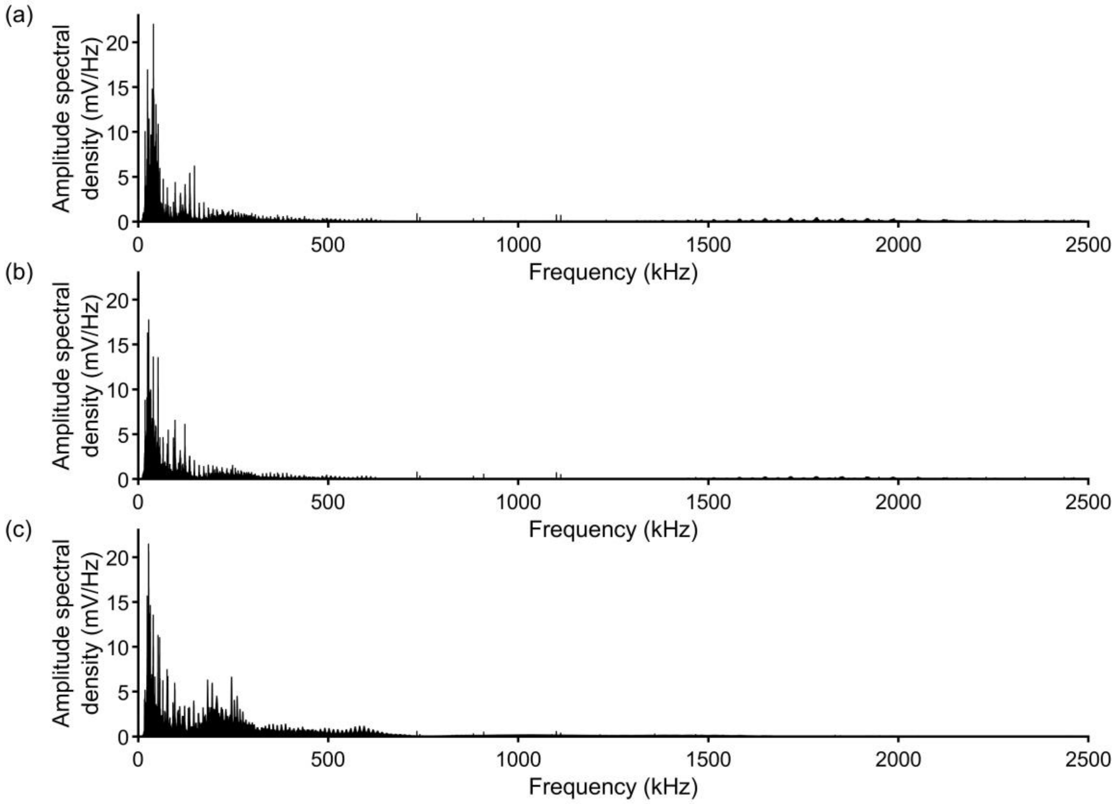


Figure 4-8 Bar plots of the average spectral content of AE samples from test stage five (2000 rpm). For samples at (a) relatively high specific film thickness, (b) midrange specific film thickness, and (c) low specific film thickness. The frequency content was calculated using an FFT with a Hanning window applied.

As an example of some of the finer detail of the spectral content Figure 4-9 presents a closer look at the low end of the spectrum, 0 – 300 kHz, for test stages one and five at their low specific film thickness ends. At this plotting resolution it can be seen that the AE activity is concentrated into numerous fine bands which appear to be sets of harmonics and are similar for both test stages. Some of these have been labelled, with arrows A to D representing harmonics spaced at ~ 20 kHz intervals, E to F at ~ 12 kHz intervals and H to J at ~ 7 kHz intervals. In fact the entire spectrums, up to 2500 kHz, are characterised by this harmonic, or finely banded, distribution. As can be seen by comparing panes (a) and (b) of Figure 4-9 the frequencies (position) of the harmonics do not alter between test stages indicating that they are not related to the speed of the rig / contact. Likewise they do not shift with changing specific film thickness indicating they are not dependent on the amount of asperity contact. The pattern of harmonics appears to be a stable system property with only amplitude, not frequency, changing with the contact conditions. Investigating the cause of this pattern was beyond the scope of this thesis but it may be related to the sensor characteristics and transmission path and material.

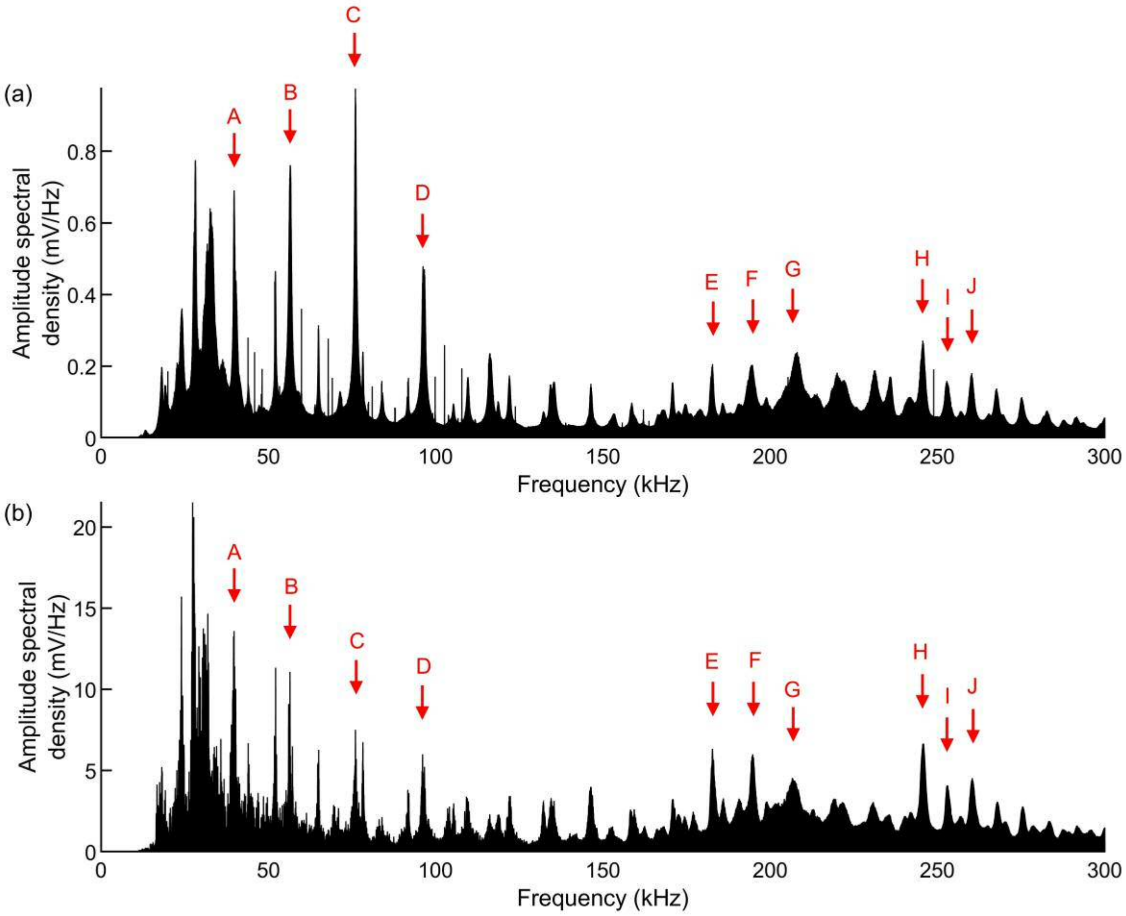


Figure 4-9 Bar plots of the average spectral content, between 0 and 300 kHz, for AE samples at relatively low specific film thickness for (a) test stage one and (b) test stage five. The frequency content was calculated using an FFT with a Hanning window applied.

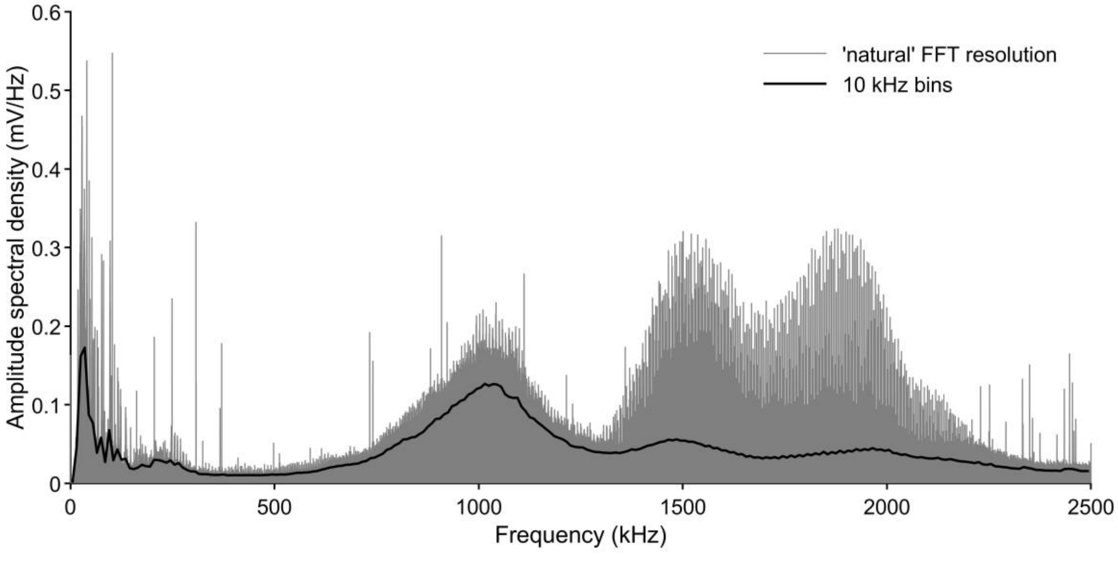


Figure 4-10 Comparison of some spectral content presented at the 'natural' resolution of the FFT and after taking the average of 10 kHz wide frequency bins. (Samples from test stage one and at midrange specific film thickness were used for this example).

The response of the AE spectrum to changing specific film thickness is now examined using a 'binned' frequency analysis. Figure 4-10 presents an example of the average spectral distribution of some AE samples at the 'natural' resolution of the FFT (in this case 4.8 mHz), and also presents the average spectral distribution with the frequency components grouped into 10 kHz wide bins and averaged. The binning removes the pattern of harmonics and makes it far easier to visually assess regional variations in amplitude.

Figure 4-11 presents binned frequency spectrograms of the AE vs. specific film thickness, for each test stage. To make these, 10 kHz binned FFTs were performed on all individual AE samples using Hanning windows. But instead of then averaging the results across specific film thickness bins, as was done for the Figures 4-8 and 4-9, they were sorted in order of specific film thickness and 'stacked' together to make the spectrograms. Each column in a spectrogram represents the spectral content of one AE sample, i.e. the AE measured during one fast disk rotation (although individual columns cannot generally be distinguished due to the size of plot). The amplitude (colour) scales are shown from zero to the maximum observed value for that individual stage.

Studying the spectrograms of Figure 4-11 it is easy to see that there was a dramatically different response to changing specific film thickness between low, 0 – 0.6 MHz, and high, 0.6 – 2.5 MHz, frequency components. Figure 4-12 presents the low frequency end of the spectrograms magnified, and it can be seen that within this there were of several sub-ranges where the frequency components behaved differently. The differences across the entire spectrum can be categorised by four frequency ranges: 20 – 150 kHz, 150 – 300 kHz, 300 – 600 kHz and 600 – 2000 MHz. Together these ranges represent almost the entire measured spectrum (remember the pre-amplifier had a 20 kHz high-pass filter). The 2000 – 2500 kHz frequency range is not considered further as the activity in this was relatively low and constant.

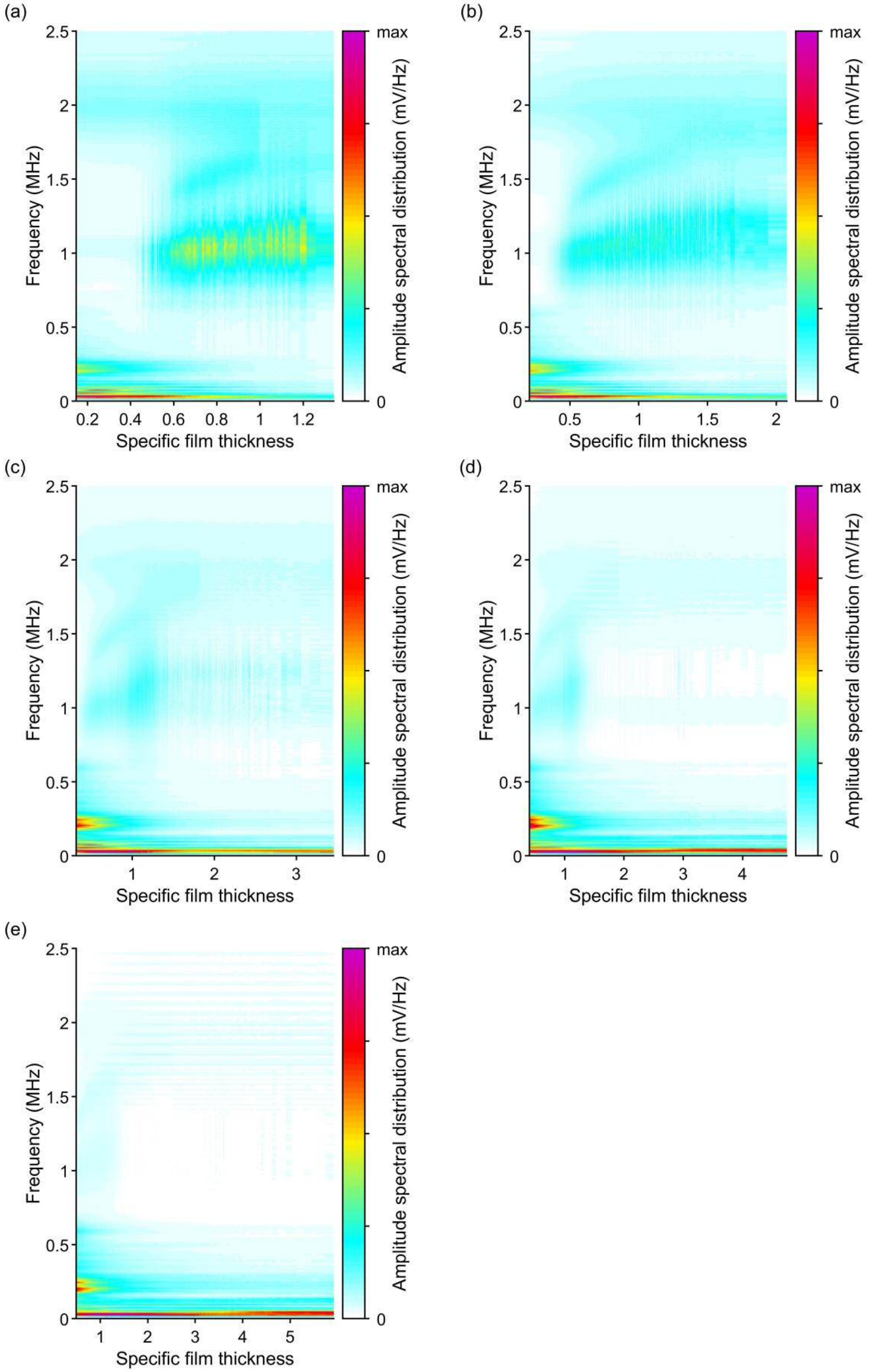


Figure 4-11 Spectrograms of the 10 kHz binned AE vs. the specific film thickness for the rotating sensor measurements, for (a) to (e): test stages one to five. Each amplitude (colour) range is from zero to the maximum observed amplitude of the individual test stage.

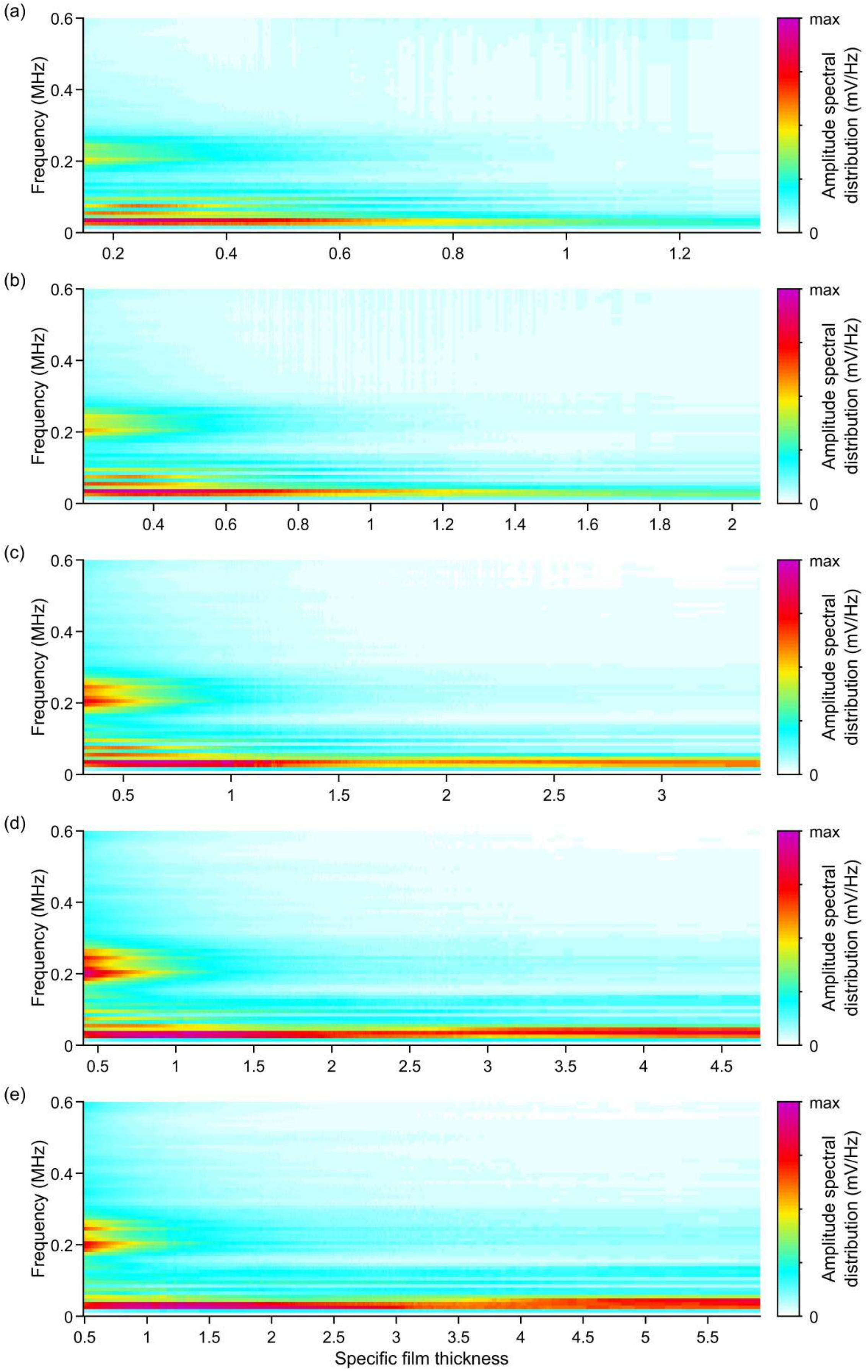


Figure 4-12 Enlarged spectrograms of 10 kHz binned AE vs. the specific film thickness for the rotating sensor measurements, for (a) to (e): test stages one to five. Each amplitude (colour) range is from zero to the maximum observed amplitude of the individual test stage.

4.7.3 **RMS of the filtered AE**

To characterise the different amplitude responses of the AE in each frequency range, the samples were digitally filtered, and the RMS of each individual sample calculated. All filters were of an equiripple design [123], with stop band attenuation of 60 dB and transition bandwidths of 10 kHz (so for the ‘150 – 300’ kHz range the 1st stop, 1st pass, 2nd pass and 2nd stop frequencies were 150, 160, 290 and 300 kHz respectively). Various different filters were trialled and it was found that the choice of design did not affect the results in a meaningful way. The RMS of the filtered AE is abbreviated to ‘FAE_{RMS}’ in general terms or ‘$FAE_{RMS\ X\text{-}X}$’ in specific terms where X-X refers to the pass band in kHz.

Figure 4-13 presents scatter plots of the FAE_{RMS} vs. specific film thickness for the four frequency ranges of interest. Consider first the $FAE_{RMS\ 150\text{-}300}$, pane (b). It can be seen that this responds to decreasing specific film thickness much more consistently than the BAE_{RMS} (the RMS of the broadband AE, Figure 4-6). The $FAE_{RMS\ 150\text{-}300}$ increases in exponential-like fashion in *all* test stages, not just the faster ones. More importantly, it increases in accordance with the changing amount of asperity contact, i.e. as the mixed regime is entered the $FAE_{RMS\ 150\text{-}300}$ increases continuously, monotonically and at increasing rate. There are no other obvious test variables that can explain this pattern of increase, so it concluded that the AE observed in the 150 – 300 kHz range was predominantly from asperity contacts at all points in the test procedure. The comparison between tests on the smooth and rough disks, Section 4.10, also confirms this. Of the four different frequency ranges the $FAE_{RMS\ 150\text{-}300}$ behaves most congruously with changes in the amount of asperity contact, and is therefore of most interest, but the other ranges are considered briefly in the rest of this section.

The response of the $FAE_{RMS\ 300\text{-}600}$, Figure 4-13 (c), is generally similar to that of the $FAE_{RMS\ 150\text{-}300}$. The need to distinguish between the two ranges is thus debatable, they could almost be considered as one consistent range. However close inspection of the $FAE_{RMS\ 300\text{-}600}$ reveals inconsistency and imprecision at the high specific film thickness ends of the two slowest tests. This does not occur for the $FAE_{RMS\ 300\text{-}600}$, and for this reason the two ranges have been kept separate. Nevertheless for large portions of the test procedure the $FAE_{RMS\ 300\text{-}600}$ did increase accordance with the changing amount of asperity contact and so it is reasoned that for most, but not all of test procedure, the AE observed in the 300 – 600 kHz range was predominantly from asperity contacts.

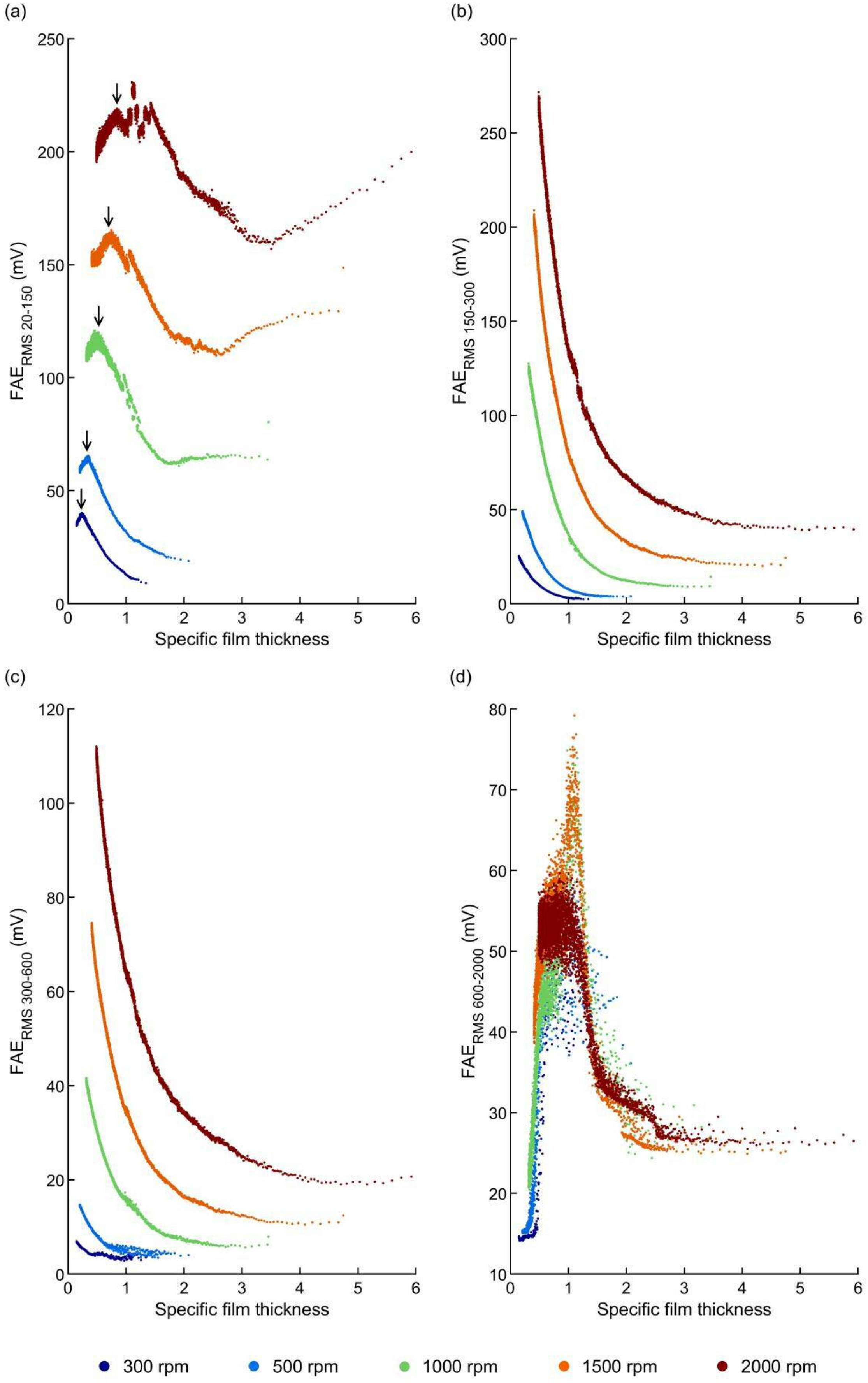


Figure 4-13 FAERMS vs. specific film thickness, for the following frequency bands: (a) 20 – 150 kHz, (b) 150 – 300 kHz, (c) 300 – 600 kHz, and (d) 600 – 2000 kHz. For the measurements from the rotating sensor.

The response of the $FAE_{RMS\ 20\text{-}150}$, Figure 4-13 (a), is significantly different and more complex than that of the two ranges already discussed. Although there are portions of each test stage where the $FAE_{RMS\ 20\text{-}150}$ increases with decreasing specific film thickness, there are notable deviations from this behaviour. In every test stage the $FAE_{RMS\ 20\text{-}150}$ peaks then starts to decrease near the bottom of the specific film thickness range. There is also, for the three fastest tests, an initial decrease in $FAE_{RMS\ 20\text{-}150}$ at the top of the specific film thickness range. The peak and subsequent decreases in $FAE_{RMS\ 20\text{-}150}$ at low specific film thickness is not at all in accordance with the changing amount of asperity contact. This does not necessarily imply the $FAE_{RMS\ 20\text{-}150}$ is unrelated to asperity contact, only that it is unrelated to the simple *amount* of asperity contact. It is interesting to note that the peak in $FAE_{RMS\ 20\text{-}150}$ does not occur at the same specific film thickness between test stages, but does occur at similar disk and lubricant temperatures. In Figure 4-13 (a) the point of every test stage where the mean disk temperature reached 90 °C is indicated by an arrow, the peak occurs at or near this temperature in all cases. This suggests that it is related to temperature rather than specific film thickness. Perhaps the most interesting hypothesis is that the $FAE_{RMS\ 20\text{-}150}$ is associated with one specific parameter of an asperity contact, such as the friction or wear, and that at a temperature threshold of ~ 90 °C there was a sudden reduction in this despite an increase in the amount of asperity contact. (The activation of additives in the oil could explain a reduction in friction or wear). A full assessment of the viability of this hypothesis is beyond the scope of this thesis but the possibility of the $FAE_{RMS\ 20\text{-}150}$ being associated with friction rather than the amount of asperity contact is discussed in Section 4.9.

The response of the $FAE_{RMS\ 600\text{-}2000,}$ Figure 4-13 (d), is the odd one out of the four filtered ranges. Firstly, the response of the $FAE_{RMS\ 600\text{-}2000}$ appears to be nearly independent of speed. The 'curves' from the five tests are close or overlap in many places to a form a roughly unified response. This is not the case for the other filtered ranges, where the increase in speed causes an in increase the scale of response. The response of the $FAE_{RMS\ 600\text{-}2000}$ is not congruous with the changing amount of asperity contact or any other measured contact parameter. Although the source of the high frequency (> 600 kHz) AE was not fully investigated for this thesis, similar high frequencies were also observed during the smooth disk testing where full-film conditions were near universal (Section 4.10.4), this rules out asperity contacts as the source. Whatever the source, this extraneous[1] AE explains the inconsistent response of the $FAE_{RMS\ 300\text{-}600}$ during the two slowest test stages. The extraneous AE actually penetrates to frequencies lower than 600 kHz and mixes with that from asperity contacts in the 300 – 600 kHz range. (The extraneous AE in the 300 – 600 kHz range can just be made out in the spectrogram of the slowest test, Figure 4-13 (a)). The extraneous AE and asperity contact AE are only of similar magnitude during the two slowest tests which is why faster tests are less affected by it.

---

[1] Extraneous in the sense that it is not related to asperity contacts.

4.7.4 Constant speed model

This section examines the mathematical relationship between the specific film thickness and the $\text{FAE}_{\text{RMS }150\text{-}300}$ (RMS of the 150 – 300 kHz pass band) as observed by the rotating sensor. Because speed affected the AE directly there were five different constant speed curves, as can be seen in (b). However it was found that each individually can be represented by a general 'constant speed model' of the form:

(10)

where is in mV, is the specific film thickness and , and are coefficients which are constant for a given speed. Table 4-2 presents fits of the model for each test stage and Figure 4-14 shows, on a semi-log plot, both the fits and the samples. Because the AE sampling rate was not uniform with respect to the specific film thickness, there were considerably more samples at the low end of the specific film thickness range of each test compared to the high. To prevent this from biasing the fits of the constant speed model the AE samples were first resampled at uniform specific film thickness intervals using a smoothing spline interpolation [124]. The degree of smoothing was chosen so that the spline faithfully followed the form of the original constant speed curve but also filtered out high frequency variations (measurement noise). The constant speed model was then fit to the smoothing spline interpolation using a non-linear least squares algorithm [125].

Table 4-2 Fits for the constant speed model (to 3 s.f.)

Speed (rpm)	**Coefficient**		
300	36.9	2.93	1.18
500	84.3	2.76	2.79
1000	231	2.10	8.70
1500	375	1.83	21.6
2000	448	1.51	41.6

To assess the accuracy of the constant speed model Figure 4-15 presents scatter plots of the specific film thickness vs. the percentage error between the fitted model and each individual sample. It can be seen that the constant speed model is broadly accurate over the entire test range. At all conditions except combinations of very low temperatures and speeds the error is within ± 10 %, and certainly there is no evidence of systematic error that would call into question the suitability of the model. The deviations of the model from the samples may to be due to measurement error in the samples rather than the model. The only instances where the model deviates significantly, are at the high specific film thickness ends of the two slowest tests. However, because of the sparsity of the samples and possible complications of start-up effects, it is difficult to be confident in the accuracy of the samples in these regions, and thus it is difficult to form a judgement on whether the model or the samples are more truthful.

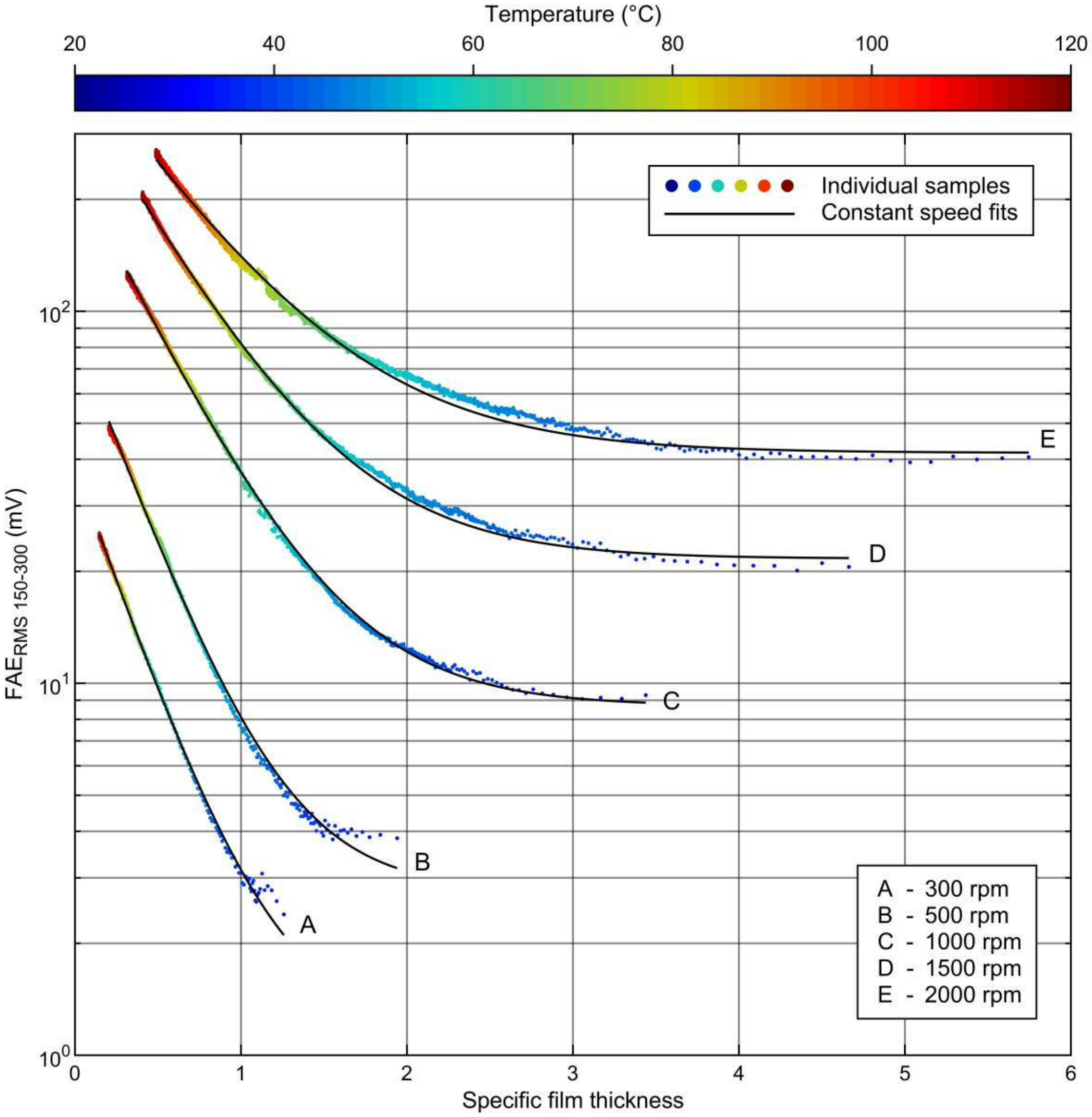


Figure 4-14 The FAE$_{RMS\ 150\text{-}300}$ from the rotating sensor vs the specific film thickness with fits according to the constant speed model, Equation 10.

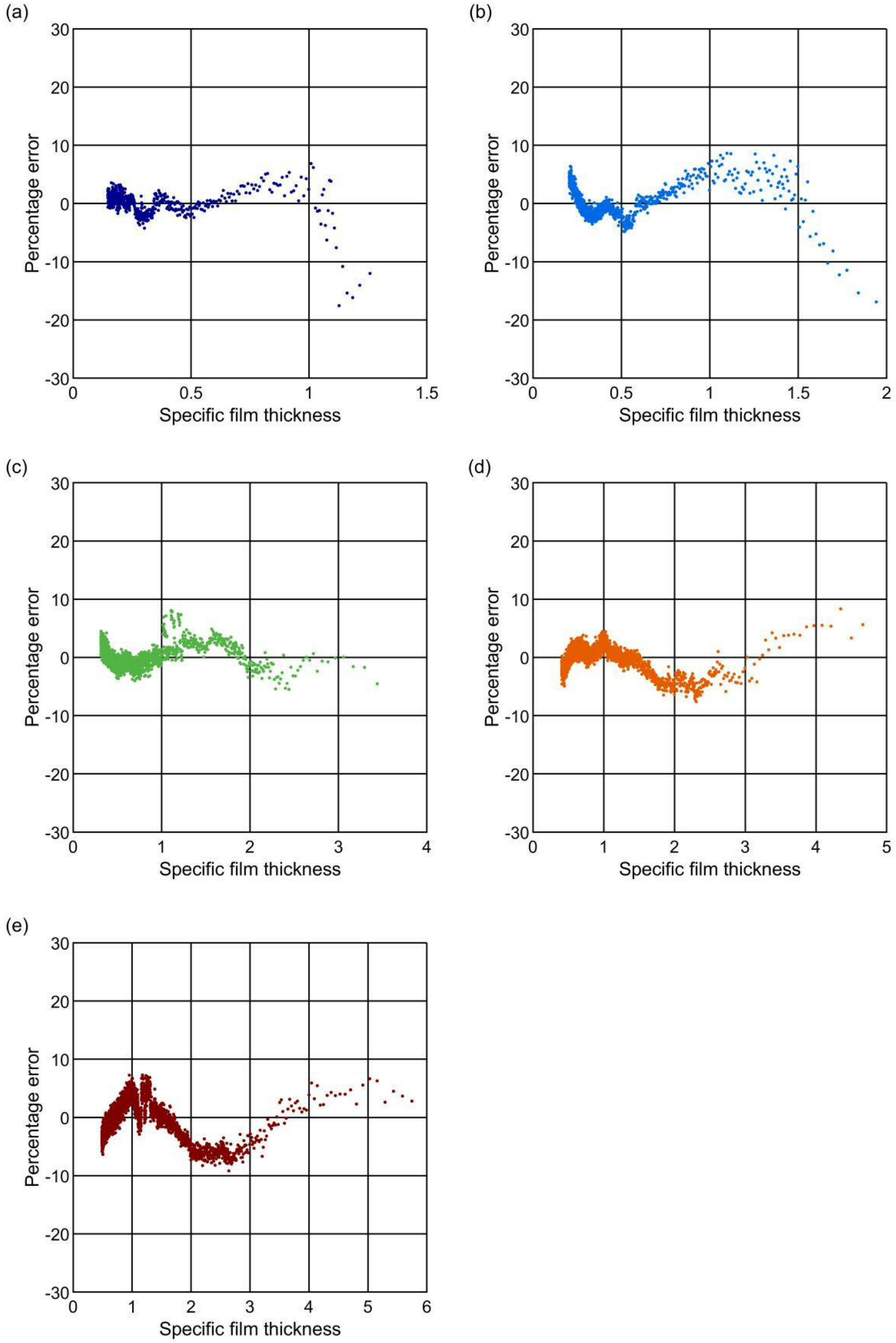


Figure 4-15 Specific film thickness vs. the percentage error between the fit / model and each measurement sample, for tests at (a) 300, (b) 500, (c) 1000, (d) 1500, and (e) 2000 rpm.

The form of the constant speed model is now discussed in detail. It consists of two terms: one an exponential function involving the specific film thickness, and one a constant. The sum of these describes the behaviour of the $FAE_{RMS\ 150\text{-}300}$ across the whole specific film thickness range tested. The transition from full film to mixed lubrication explains the need for a two termed, or summing, model. In full film conditions the constant term dominates and the influence of the exponential term, i.e. the specific film thickness is negligible. But as the lubrication regime enters the mixed regime the influence of the exponential term becomes increasingly significant and that of the constant less so.

The constant term represents the baseline, or background, level of AE that occurs in full film conditions. Interestingly, there is no indication that, if extrapolated, the different curves would converge to a *single* baseline. This implies the background AE is still directly affected by speed. It follows that the direct speed effect cannot *only* be caused by the changing rate of asperity interaction as there is no asperity interaction in full film lubrication, whatever the speed. Instead the results show that the direct speed effect is more generally applicable.

The exponential term represents the AE from increasing asperity contact at low specific film thicknesses. In fact, below a specific film thickness of unity, i.e. well into the mixed lubrication regime, the relationship between specific film thickness and AE can be represented by a *purely* exponential model. This is evident from the straight line portions of the curves in Figure 4-15 (a straight line representing an exponent on a semi-log plot). Just as there is a limit to the purely exponential relationship as the specific film thickness increases and the lubrication regime approaches full film conditions, it is logical to suppose that there would have been a limit if the lubrication regime had approached full boundary conditions. There is no evidence of this which is consistent with the results of the mean contact voltage, which only drops to intermediary values at the lowest specific film thickness tested.

At this point it is worth emphasising that the generality of the constant speed model is unknown. All that is certain at this point is that it applies to this specific procedure which only tests one control variable, the temperature[1]. It is probable, that inducing changes in the specific film thickness using other fundamental contact parameters such as roughness, geometry, elasticity etc. would result in, at least subtly, a different relationships between the AE and specific film thickness. It has already been shown this is the case for changes in speed. It is possible that that a *general* model of the relationship between acoustic emission and lubrication conditions would be complex and would need to take into many of the fundamental contact parameters, rather than just the specific film thickness, and speed.

---

[1] Consider the following joke: A mathematician, a physicist and an engineer are riding a train through Scotland. The engineer looks out the window, sees a black sheep, and exclaims, “The sheep in Scotland are black.” The physicist looks out and corrects the engineer, “All we know is that there is at least one black sheep in Scotland.” The mathematician looks out and corrects the physicist, “Strictly speaking, all we know is that at least one side of one sheep in Scotland is black.”

4.7.5 **Variable speed model**

In this section a variable speed model for the FAE$_{\mathrm{RMS}\ 150\text{-}300}$ vs. specific film thickness is presented. Unlike the constant speed model of the previous section the direct effect of speed is incorporated so that the FAE$_{\mathrm{RMS}\ 150\text{-}300}$ in any test stage can be described by a single fit where only the specific film thickness and fast disk speed are variables. The variable speed model is based on the five speed specific fits of the constant speed model given in Table 4-2. Using this data Figure 4-16 shows plots of the , and coefficients vs. speed.

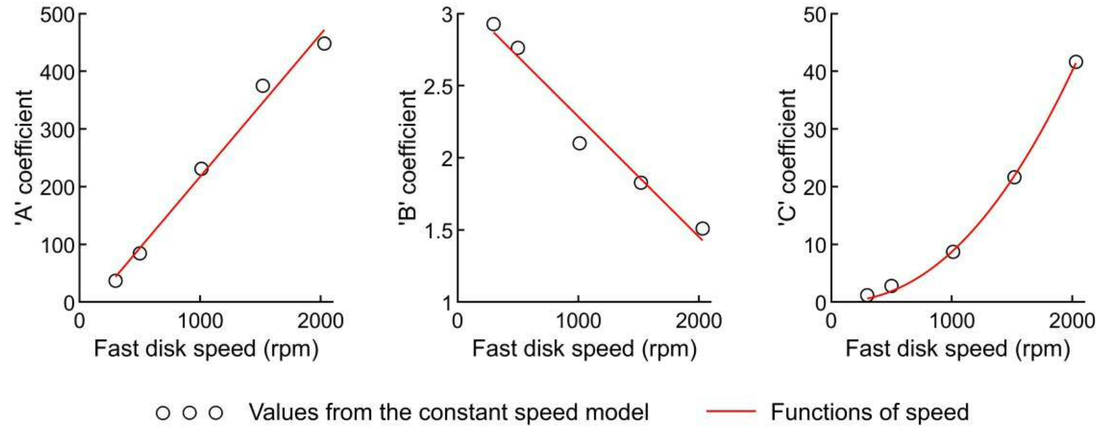


Figure 4-16 The A, B and C coefficients vs. fast disk speed for the constant speeds models of Table 4-2. Also shown are the fits for the functions of speed: Equations 12-14.

Desipte the small number of samples it can be seen that each of the coefficients has a systematic relationship with speed. Over the range tested the and *B* coefficients vary approximately linearly with speed and the *C* coefficient with a power of speed. Substituting the *B* and *C* coefficients of the constant speed model (Equation 10) with these relationships, or functions of speed, gives the following variable speed model:

(11)

where

(12)

(13)

(14)

and where, is the fast disk speed in rpm, and , …. , are coefficients from the functions of speed. Table 4-3 presents fits of the functions of speed, and these fully specify the variable speed model. The fits of each function are also plotted in Figure 4-16.

Table 4-3 Fits for the functions of speed, Equations 12-14, which specific the variable speed model (to 3 s.f.)

	-29.9		3.12	2.13	2.20

A brief discussion of the three functions of speed is now given. Consider first , Equation 12. Although the linear relationship is satisfactory across the range of speeds tested, it has unrealistic consequences if extrapolated to slower speeds. At a fast disk speed of ~ 120 rpm the fit of in Table 4-3 returns a value of zero which implies the AE becomes constant regardless of the specific film thickness. And for fast disk speeds below 120 rpm the fit returns a negative value, implying negative AE which is clearly nonsense. It would be more logical for the fit of to pass through the origin, i.e. for to equal zero, however if this is forced the accuracy of the variable speed model for the slowest tests is greatly compromised (its accuracy is discussed shortly). It thus seems likely that at speeds slower than tested, the relationship between the coefficient and speed must deviate from the linear one observed.

Next consider , Equation 13, This is satisfactory for the range of speeds tested and the fit only has unrealistic consequences if extrapolated to speeds far faster than those tested. At a fast disk speed of ~ 3700 rpm returns a value of zero, nullifying any relationship between specific film thickness and AE, and at speeds faster than this the relationship becomes reversed with decreasing specific film thickness causing a decrease in AE (rather than an increase). So it appears probable that had faster speeds been tested, the decrease in the B coefficient would have begun to taper off.

Finally consider , Equation 14, this is the least ambiguous of the three functions of speed. The coefficient has a good power law fit with the speed, and no unrealistic conditions occur if the relationship is extrapolated to faster or slower speeds. As discussed in the previous section the coefficient represents the background AE persisting in full-film conditions and the fit of suggests that this background level varies with a power of speed.

Ultimately, the speed functions and the variable speed model are exercises in data-fitting and although useful it is difficult to see how any of the speed functions might be related to a real physical mechanism, i.e. to the fundamental cause of the direct speed effect.

The accuracy of the variable speed model is now considered. Figure 4-17 shows plots of the fitted variable speed model and the original measurements. Figure 4-18 presents the mean percentage error between the model and each set of constant speed measurements. It can be seen that use of the variable speed model has introduced significantly more error than the use of multiple instances of the constant speed model. For each fit of the constant speed model, the mean percentage error, between the model and the measurements, is near zero[1] (fractions of a percent for all speeds). However when using the variable speed model the mean percentage error becomes significantly larger and inconsistent between speeds, as is shown in Figure 4-18. This occurs because of error between the functions of speed and the true , and coefficients. By biasing fits of the speed functions to have better accuracy at higher or lower speeds, it is possible to shift at which speed the variable speed model has the lowest mean percentage error, but it is not possible to reduce it universally by this method.

It should be noted that the specific variable speed model presented here is based upon fits of fits, that is: fits for the speed functions are based upon fits of constant speed model. Thus any error in the initial fits will have been carried through. It is therefore possible that error in the variable speed model could be reduced by altering the parameters of the constant speed model. A neater solution might be to fit the variable speed model directly to the measurement data, i.e. to the three dimensional sample set of AE vs. speed vs. specific film thickness. A cursory attempt at this was made using the standard fitting algorithms available in MATLAB but without any immediate success. Ultimately there was little to be gained by refining the variable speed model further. For the purpose of the thesis it was sufficient to show that the direct speed effect is consistent and predictable.

---

[1] This is hardly surprising given that minimising *total* error is the raison d'etre of the least-squares algorithm used for the constant speed fits.

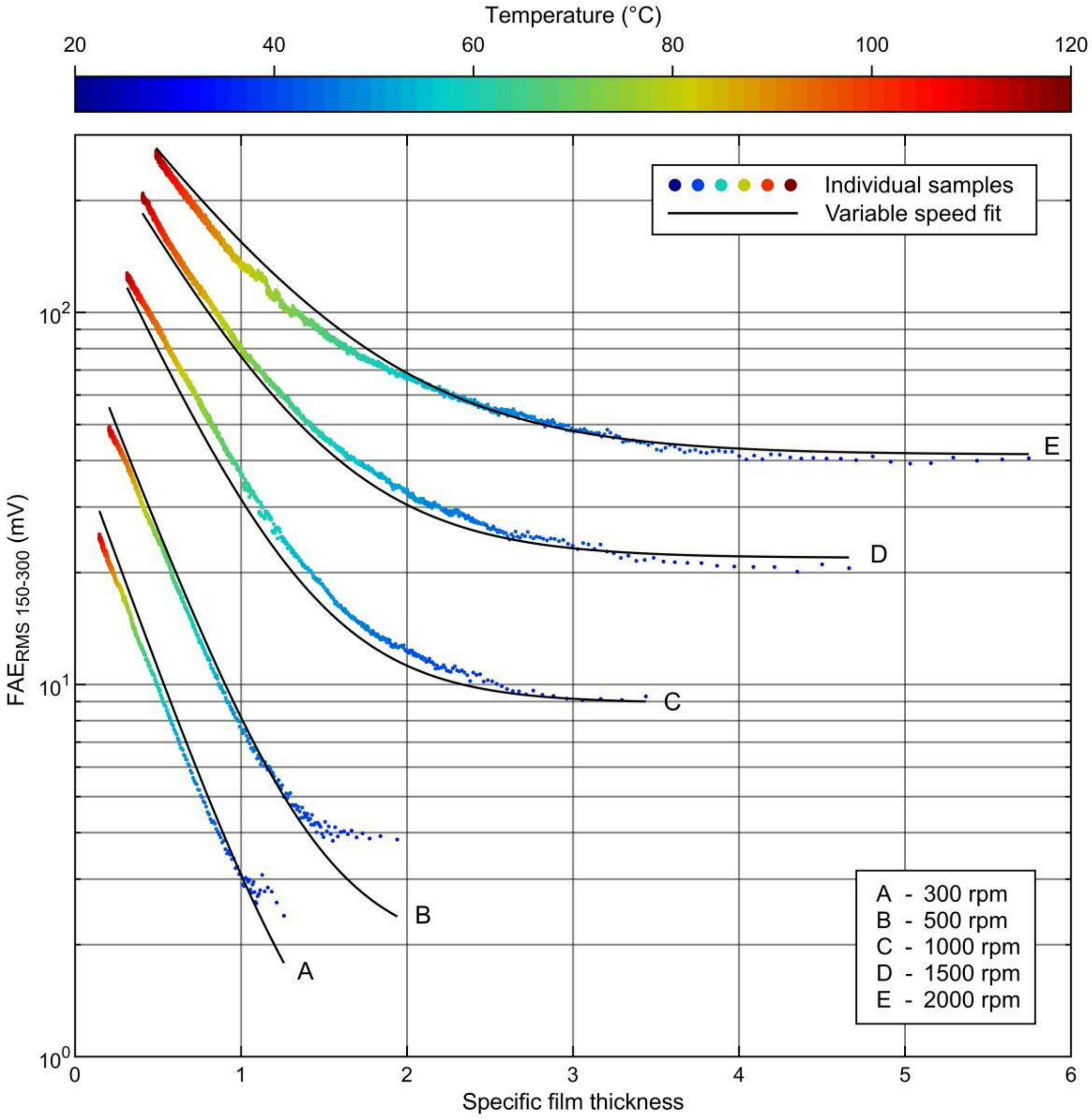


Figure 4-17 Comparison of the variable speed model / fit and the actual measurements.

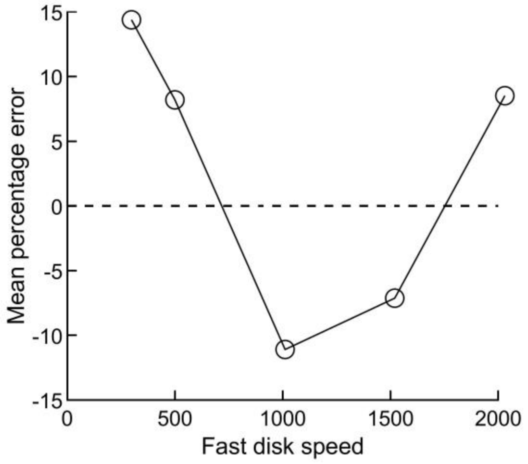


Figure 4-18 Mean percentage error between the variable speed model / fit and each set of constant speed measurements.

4.8 Comparison of sensor couplings

This section compares the AE measurements from the clamped rotating sensor and those from the oil coupled static sensor. As the sensors were of the same type, with only minor differences in response (see Figure 3-12), differences in the AE are due to the coupling method not the sensor characteristics. For the rotating sensor measurements, the AE observed in the 150 – 300 kHz frequency band had the most satisfactory relationship with the amount of asperity contact (see previous section), and so it is this range that is used in the comparison.

Figure 4-19 presents the $FAE_{RMS\ 150\text{-}300}$ from each sensor vs. the specific film thickness. (The $FAE_{RMS\ 150\text{-}300}$ was calculated as described in Section 4.7.3). The amplitude range from each sensor has been normalised between zero and the maximum value measured by each sensor individually. This occurred at the same point in testing for both: the lowest specific film thickness generated by the fastest speed. It can be seen that the relationship between the $FAE_{RMS\ 150\text{-}300}$ and specific film thickness is of a much lower quality using the static sensor measurements as opposed to those from the rotating sensor. The precision of the static sensor relationship is considerably worse and the form far less consistent both within and between test stages. Despite this, the two sets of measurements do agree in a very broad sense, i.e. the two families of curves are approximately overlaid. This shows that fundamentally both sensors were capable of measuring the AE from asperity contacts, despite the difference in coupling methods. Given that the coupling method of the static sensor, an oil filled gap, has not been tested before this is an important finding.

The un-normalised maximum amplitude from the rotating sensor was approximately twice that of the static of sensor, (1.95 times to be precise). This difference in scale reveals that the amplitude attenuation through the oil coupling of the static sensor was approximately double that of the direct contact used for the rotating sensor. This is not surprising, but despite this there is no evidence that the limits of sensor resolution were approached. Even at the lowest measured AE amplitudes, which occurred during testing at the slowest speed, the measurements from the static sensor show systematic variations, i.e. behaviour inconsistent with background noise. Thus, in the case of this testing, the greater attenuation of this interface method was of no consequence.

Although the results of both sensors are in coarse agreement, there are significant localised divergences. The results from the rotating sensor are far more likely to be an accurate measurement of the AE from the test contact, than those of the static sensor. There are two reasons for this. Firstly the measurements from this sensor are more congruous with the known changes in the lubrication conditions. Secondly the oil coupling is far more likely to be susceptible to noise than the direct coupling. The parameters of oil coupling certainly varied during testing. Both temperature and speed changes will have affected the oil flow in the interface. Additionally thermal strain in the sensor arm and fast shaft assemblies will have affected the gap width. It is speculated that changes in these coupling parameters may have induced noise by two mechanisms: variations in signal attenuation, and or generation of extraneous AE. For example, *if* the flow regime changed so that there were more or less voids, or bubbles, in the oil coupling then the attenuation would likely have changed. And *if* the gap closed to the extent that the sensor interface entered the mixed regime, then asperity contact would likely have caused significant extraneous AE.

Because, as is shown by Figure 4-19, the quality of the measurements from the rotating sensor were so much higher than those of the static sensor, the latter are not utilised any further in this or subsequent chapters. Despite this, the static sensor results are still important as they suggest that *fundamentally* an oil interface can work. The specific design of oil interface used in this thesis was relatively crude so high quality results were not expected. It is possible that the quality of measurements might be improved with refinements to the design such as the use of a hydrodynamic wedge or hydrostatic bearing to control the oil flow, as well as better geometric tolerances and smoother surfaces to ensure no asperity contact.

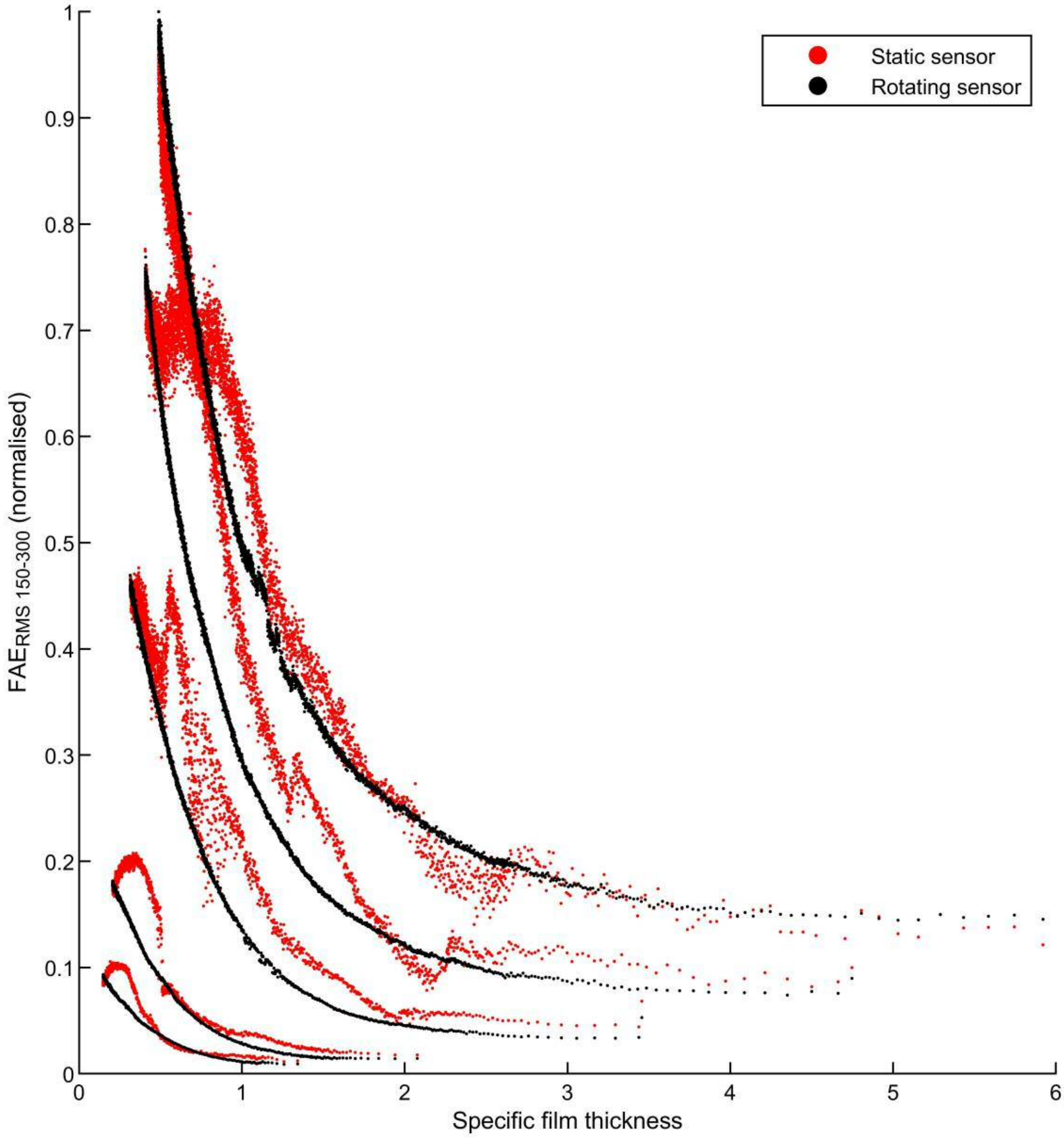


Figure 4-19 Comparison of the $FAE_{RMS\ 150\text{-}300}$ *observed by the static and the rotating sensor for the five constant speed tests. The amplitude has been normalised between 0 and the maximum measured amplitude and increases with test speed. The scaling factor between the AE amplitude for the rotating and static sensor is 1.95.*

4.9 Coefficient of friction

This section presents a brief analysis of Coefficient of Friction (CoF) measurements and their relationship to the specific film thickness and AE.

Figure 4-20 presents the CoF vs. the specific film thickness for all test stages. (The colour-scale indicates the disk temperatures and is shown as an aid to interpret the results). Each data point is the average CoF of a single fast disk rotation. These data points merge to form five distinct constant speed curves. Studying the detail reveals an apparent split at the low temperature end of each curve into two distinct and slightly offset sub-curves. Closer inspection of the curves using a larger plot and smaller data points reveals that each curve is actually formed of up to at least four sub-curves over its entire length. It is speculated that this was due the cyclic period of the contact, which was five fast disk rotations long (see Figure 3-5). For each fast disk rotation within the cyclic period the orientation of the slow disk was different and it is likely this induced minor differences in contact conditions and thus up to five subtly different sets of results: the five sub-curves.

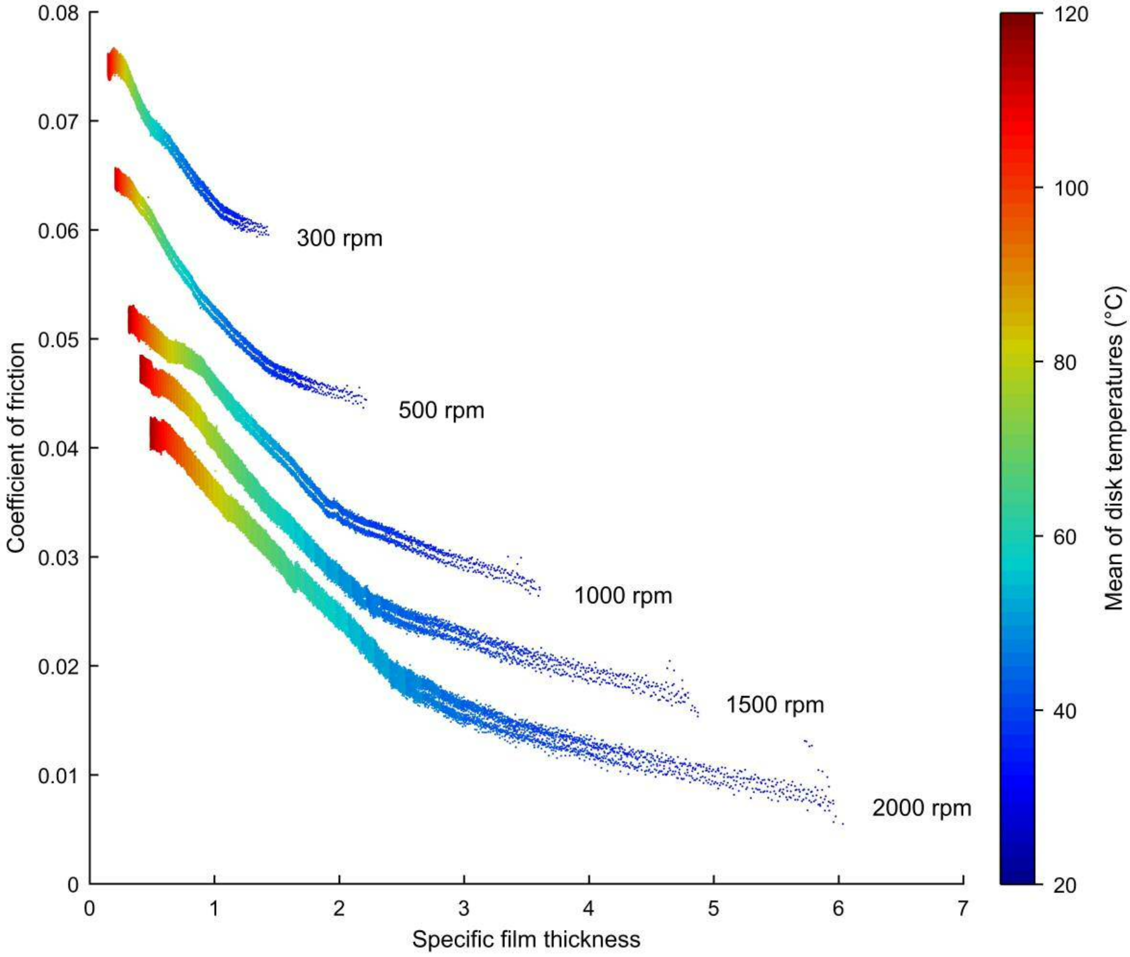


Figure 4-20 Coefficient of friction vs. the specific film thickness for all test stages

First consider the change in CoF within individual test stages. Figure 4-20 shows that as the specific film thickness decreased the CoF increased. However the increase was not in direct proportion to the changes in the amount of asperity contact. The CV (Figure 4-4) and the $FAE_{RMS\ 150\text{-}300}$ (Figure 4-14) both indicate that with respect to the specific film thickness, the increase in the amount of asperity contact was highly non-linear. In contrast the CoF increased much more linearly. For example: At the fastest speed, there was a linear increase in CoF as the specific film decreased from ~ 2.5 to its minimum ~ 0.5, but the gradient of the CV transitions from near 0 to near -1 over the same range. (The transition in gradient of the $FAE_{RMS\ 150\text{-}300}$ was less dramatic than the CV but was still far from linear). This difference is to be expected because asperity contacts are not the sole source of friction within a lubricated contact. As demonstrated by Stribeck [30], both boundary friction caused by asperity contacts and viscous shear in the fluid film affect the overall friction of a contact. In this case, for each test stage, the increase in asperity contacts was accompanied by a reduction in lubricant viscosity, and while the former acts to increase the total contact friction the later acts to decrease it. So it is likely that the increase in CoF over each test stage was tempered by the reduction in lubricant viscosity. There is also the possibility that changes in the nature of the boundary lubrication, such as the activation of friction modifiers in the lubricant, helped limit the increase in the CoF.

Now consider the change in CoF between test stages. Examining lines of constant specific film thickness (vertical lines) on Figure 4-20, shows that as the speed and temperature increased in conjunction, the CoF decreased. Again this can be explained by the temperature induced reduction in the viscosity of the fluid film. The effects of temperature on the fluid film friction were previously observed by Weeks [57] during his experiments on the same rig, but are more clearly visible in these results. Weeks suggested that the influence of changes in fluid film friction were greater than that of changes in boundary friction however this is not seen here. For each constant speed curve the coefficient of friction increased with decreasing specific film thickness despite decreasing viscous friction (increasing temperature). This is even the case at the highest specific film thickness values where asperity contacts are negligible.

Figure 4-21 presents the $FAE_{RMS\ 150\text{-}300}$ and $FAE_{RMS\ 20\text{-}150}$ from the rotating sensor measurements vs. the CoF. First consider the $FAE_{RMS\ 150\text{-}300}$, pane (a). This is of interest because the $FAE_{RMS\ 150\text{-}300}$ was found to vary in accordance with the amount of asperity contact, and to have a simple and consistent relationship with the specific film thickness (Equation 10). It has a less satisfactory relationship with the CoF. For each constant speed curve the $FAE_{RMS\ 150\text{-}300}$ does increase with the CoF (right to left along the curve) however the increase is far from proportional, i.e. linear. Changes to the CoF at low values barely elicit a response in the $FAE_{RMS\ 150\text{-}300}$, whereas changes at high values elicit a relatively extreme response. This is because the $FAE_{RMS\ 150\text{-}300}$ is, in reality, not responding to changes in the general CoF, but to changes in the amount of asperity contact.

Now consider Figure 4-21 (b), the $FAE_{RMS\ 20\text{-}150}$ vs. CoF. This is of interest because, as was discussed in Section 4.7.3, despite a continually increasing amount of asperity contact, the $FAE_{RMS\ 20\text{-}150}$ peaked and then decreased, and it was thought that it could be related to some other contact parameter, such as the friction. However, as Figure 4-21 (b) shows, for all test stages except the slowest, the peak still persists in the relationship between the $FAE_{RMS\ 20\text{-}150}$ and CoF and so it seems unlikely that it is related to friction.

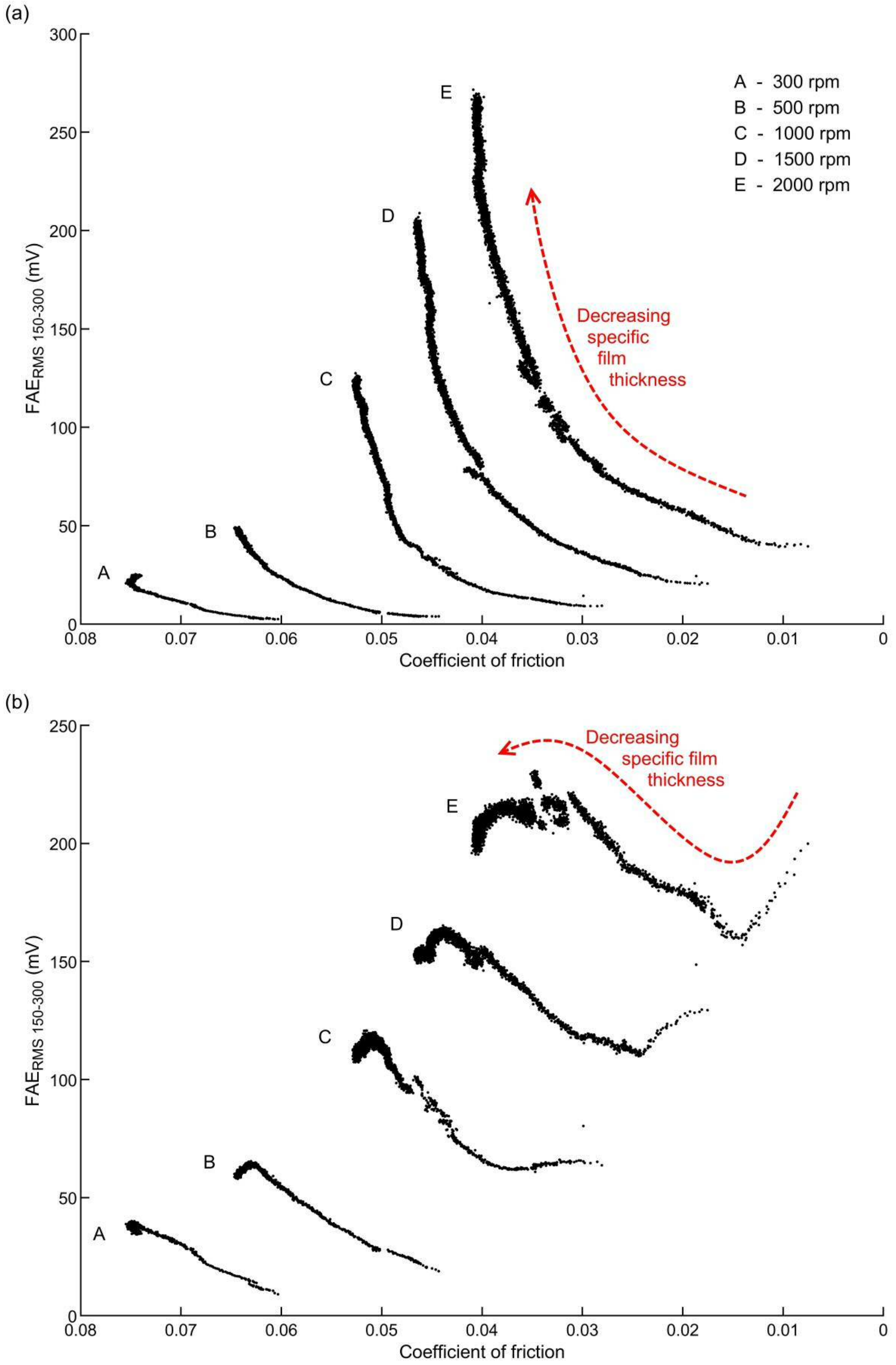


Figure 4-21 Rotating sensor AE measurements vs. the coefficient of friction for (a) the FAE$_{RMS\ 150-300}$ (RMS of the 150 – 300 kHz band pass AE) and (b) the FAE$_{RMS\ 20-150}$ (RMS of the 20 – 150 kHz band pass AE).

4.10 Rough-smooth comparison

4.10.1 Test procedure

In the analysis of the AE thus far, it has been shown that within the broadband signal there was a frequency range, 150 – 300 kHz, where the RMS of the observed AE varied in accordance with the amount of asperity contact, however there were other frequency ranges, 20 – 150 kHz and 600 – 2000 kHz where it did not. To confirm that the AE observed in the 150 – 300 kHz range was from asperity contacts, and to test if the AE observed in the other ranges was as well, it was decided to repeat the experiment, but with one major difference: a superfinished, i.e. smooth, disk pair would be substituted for the as-ground, i.e. rough, pair. This second procedure was intended to replicate all the contact parameters of the first with one exception: there would be much less asperity contact, i.e. whereas the *uniform* film thickness range would not change, the *specific* film thickness range would shift upwards. It was reasoned that with the two procedures, it would be possible to compare contact conditions where the *only* difference was the *specific* film thickness, i.e. the amount of asperity contact, and thus any differences in the AE would confirm it to be associated with this.

Although the only major change for the second procedure was the disk type there were some minor changes which were as follows: Firstly, to save time and because the propensity for wear was anticipated to be much less, the in-situ roughness profiles were not measured in-between every stage. Secondly, only one oil heater was available at the time of testing. Consequently the temperature rise was much slower and thus each stage took considerably longer than for the first procedure. Finally, by this point in the research the bearing housing AE sensor had been added to the rig. With three sensors it was only possible to perform the ADC at a rate of 2 MHz per sensor rather than 5 MHz per sensor as was the case for the first procedure.

In the next three sections the results of the two procedures are compared starting with the roughness and contact voltage measurements, which are used to verify that the lubrication conditions were different, and then finishing with an analysis of the AE measurements.

4.10.2 Roughness

This section provides a basic comparison of the (RMS roughness) between the two disk pairs. Figure 4-22 shows the mean of each disk pair at equivalent stages in their procedures. The means were calculated from all twelve profile measurement positions (six per disk: three on each centre line and three near the contact's edge). The error bars show the standard deviation between the different measurement positions. As anticipated there was no significant change in the of the smooth (superfinished) disks over the course of their testing. And it can be seen that the subtle wearing of the rough (ground) disks, which was identified in section 4.3, is of no consequence in the comparison between them. The of the smooth disk pair always remained approximately five times less than that of the rough pair. Thus at an equivalent *uniform* film thickness, the *specific* film thickness for the rough disks was always about five times greater than that of the ground disk pair. The total specific film thickness range, from hottest temperature and slowest speed, to coldest temperature and fastest speed; was approximately 1 to 30 for the smooth disks compared to 0.2 to 6 for the rough disks. This suggests, that for the smooth disk pair, there was little possibility of asperity contact for most of the test conditions. This can be easily verified using the contact voltage measurements which are the subject of the next section.

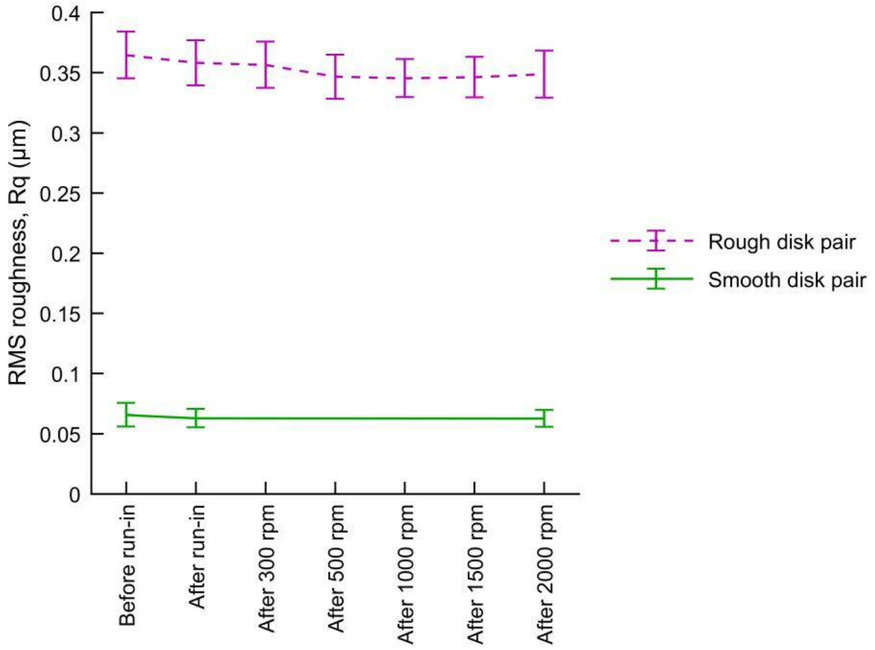


Figure 4-22 Comparison of the mean of the smooth (superfinished) and rough (ground) disk pairs. The error bars represent ± one standard deviation of the measurements from the 12 positions.

4.10.3 **Contact voltage**

The implication of the difference in , might be considered sufficient evidence that there was less asperity contact when testing the smooth disk pair. Nevertheless, it is still useful to compare the contact voltage measurements as these provide explicit confirmation. Figure 4-23 presents scatter plots of the mean contact voltage vs. the *uniform* film thickness for both disk pairs and all five test stages. It can be seen that for the rough disk pair, the mean contact voltage dropped to intermediate values at some point in all test stages, indicating the change from full film to mixed lubrication, whereas for the smooth disk pair it did not. There was a subtle drop for the slowest speed tested but this drop was not repeated, or was increasingly diminished, in subsequent test stages. This diminishing drop may have been caused by some very subtle wear or the influence of speed on the film thickness topography, but it is not important for the rough vs. smooth disk comparison. What is important is that, at low uniform film thickness the contact voltage was significantly higher, i.e. there was less asperity contact, between the smooth disks compared to the rough. Furthermore, at the faster speeds the contact voltage remained at a maximum over the entire stages and so there can have been negligible AE from asperity contact at any point in these.

As a slight aside, it is speculated that had the smooth disks been tested well into the mixed regime, the difference in the contact *topography* between the smooth and the rough pair, would have resulted in a somewhat different contact voltage and AE response. Some hint of this can be seen by comparing the instantaneous contact voltage measurements. Figure 4-24 shows, for both disk pairs, the instantaneous contact voltage vs, the uniform film thickness, at the slowest speed. Robust comparison is difficult because, as intended, the change in the lubrication regime was not consistent between the disk pairs, and also the contact voltage signal was somewhat noisy. However there is a suggestion of a difference in how the instantaneous contact voltage fluctuated. Arguably the fluctuations for the rough disk pair are more rapid, i.e. narrowly banded, and evenly distributed than for the smooth disk pair, which would indicate an effect of topography, however the difference is not certain and ideally the instantaneous contact voltage at equivalent *specific* film thickness needs to be compared. It is also interesting that for the smooth disks as well the rough there are transient increases in asperity contact at select circumferential positions (as was discussed for the rough disks in Section 3.4.3). This suggests that the surface conditions of the smooth disks also fluctuated in-test more than the pre and post test roughness measurements indicate.

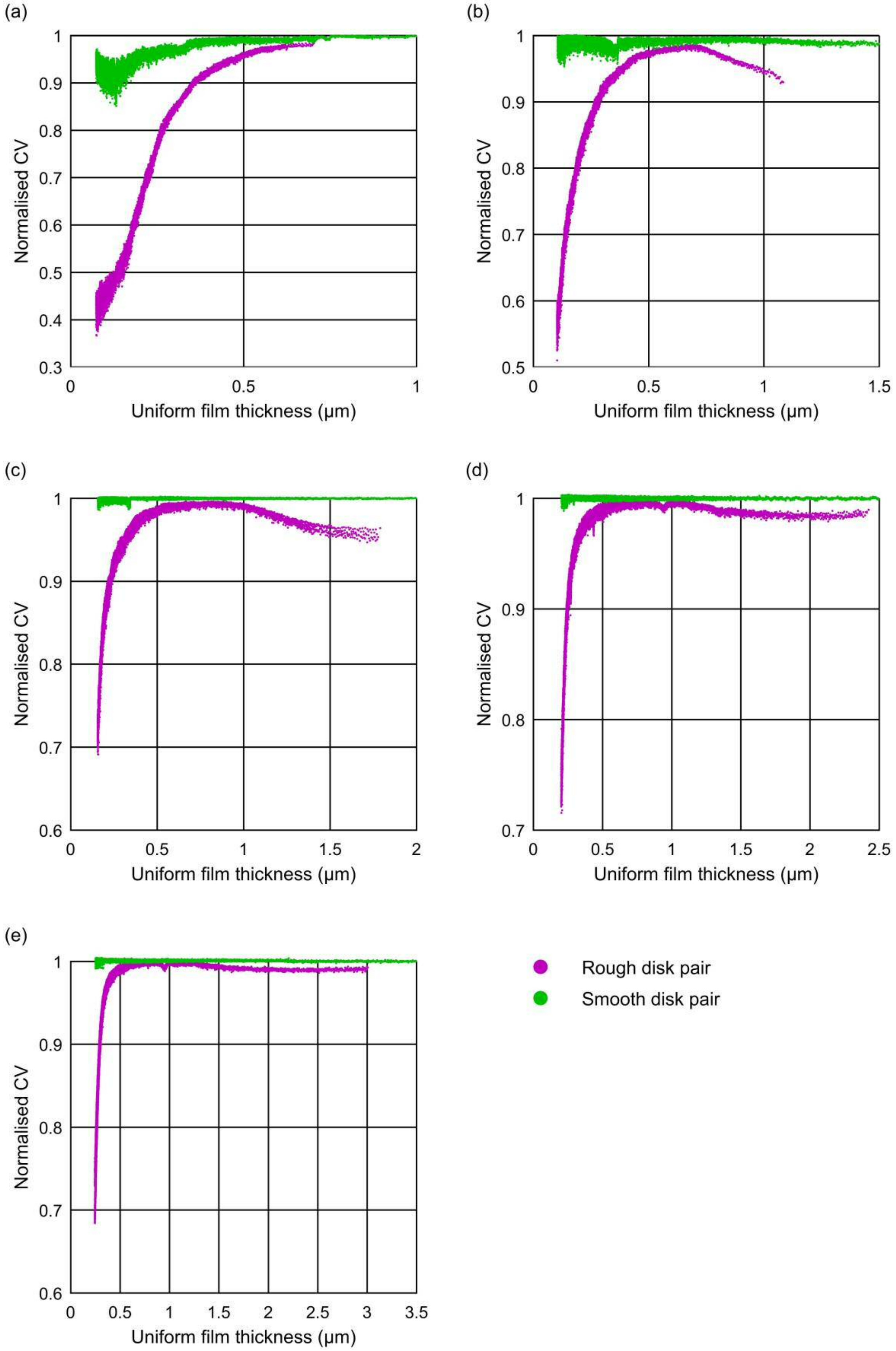


Figure 4-23 Comparison of mean contact voltage vs the uniform film thickness for the rough and smooth disks for tests at (a) 300, (b) 500, (c) 1000, (d) 1500 and (e) 2000 rpm.

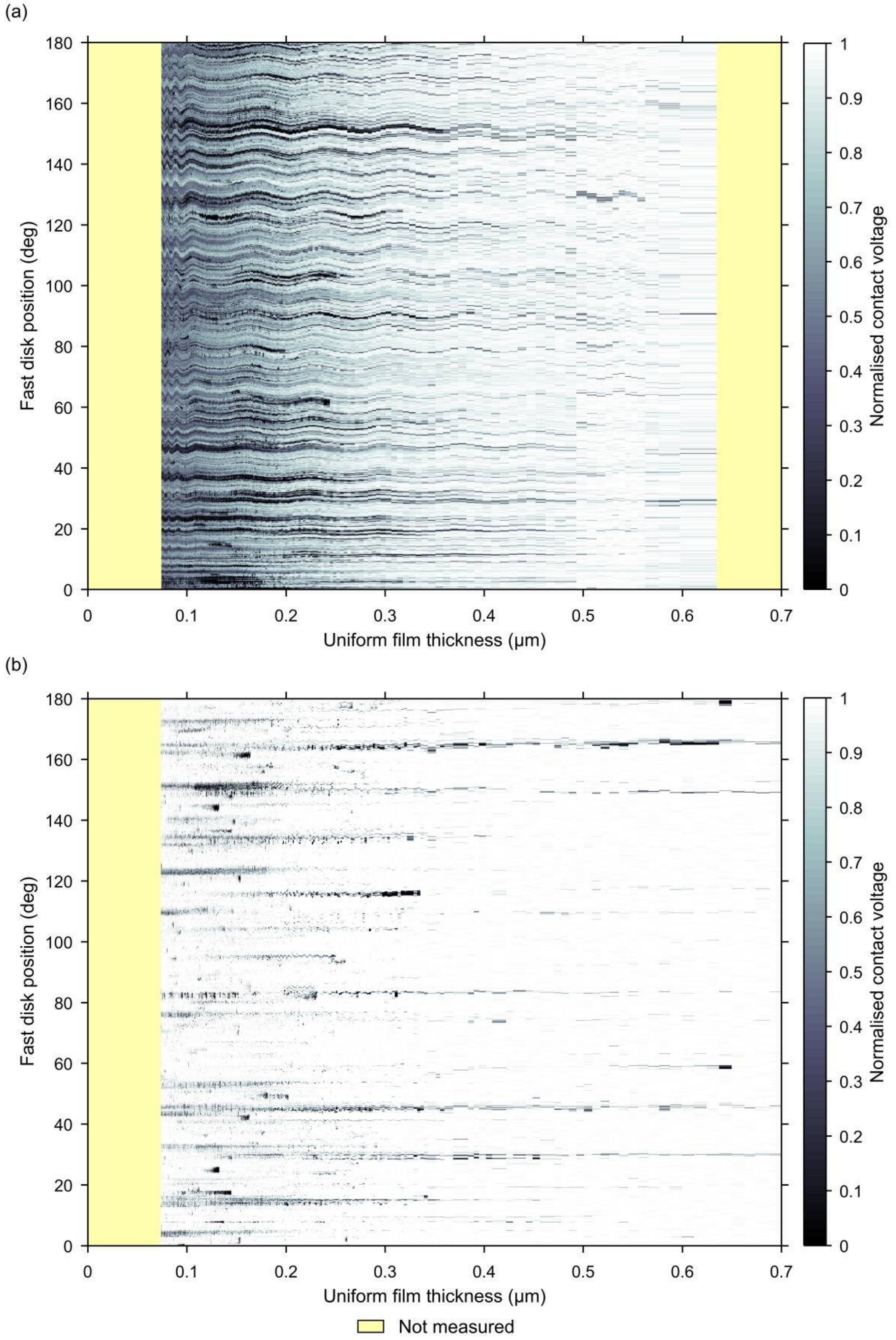


Figure 4-24 Comparison of instantaneous contact voltage vs. uniform film thickness from (a) the rough and (b) the smooth disk pairs. From the test at a fast disk speed of 300 rpm. (A 30 kHz low pass filter was used for the instantaneous contact voltage).

4.10.4 **AE**

In this section the AE measurements from the smooth and rough disks are compared. Only the rotating sensor measurements are used, for as was revealed earlier, they were of superior quality to those of the static sensor. Figure 4-25 shows the 10 kHz binned spectral distribution of the observed AE from each disk pair vs. the uniform film thickness for all five test stages. Identical amplitude (colour) scales have been used for both disk pairs. Because of the reduced ADC rate for the smooth disk samples it is only possible to compare spectral content up to a maximum frequency of 1 MHz, but as the low frequency range, 0 – 0.6 MHz, is of most interest this is not a great hindrance.

Consider the fastest test stage first Figure 4-25 (e). It can be seen that there is a dramatic difference in the AE response to decreasing uniform film thickness between the two disk sets. For the important frequency range of 150 to 300 kHz, the AE from the rough pair increased dramatically as low uniform film thicknesses were reached. In contrast for the smooth disk pair it remained constant and comparatively negligible. This is exactly the behaviour expected from the change in surface roughness. The comparison thus affirms that, as was concluded initially from the rough disk tests alone, the source of AE measured in this range was predominantly from asperity contacts. What is more interesting is that for the smooth disk pair the comparatively negligible AE response extends across almost all the low frequency range: 20 to 600 kHz. This implies that the majority of AE in this range was also from asperity contacts (although potentially indirectly). So despite the $FAE_{RMS\ 20\text{-}150}$ not varying in accordance with the amount of asperity contacts (Figure 4-13), the rough-smooth comparison shows that its response is still related to asperity contacts.

Now consider the high frequency range, 600 to 2000 kHz, of the slowest test stage, Figure 4-25 (a). As was discussed in Section 4.7.3, for the rough disk pair this was the only stage where the high frequency AE response (the $FAE_{RMS\ 600\text{-}2000}$) was of comparable magnitude to the low frequency response, and this is why it is not visible in the faster stages. Despite only being able to compare frequencies up to 1 MHz, it can be seen that the high frequency response is still evident for the smooth disk pair, in fact it is even more pronounced in the range available for comparison. This is indicates that the high frequency AE is not caused by asperity contact, or any surface roughness effect.

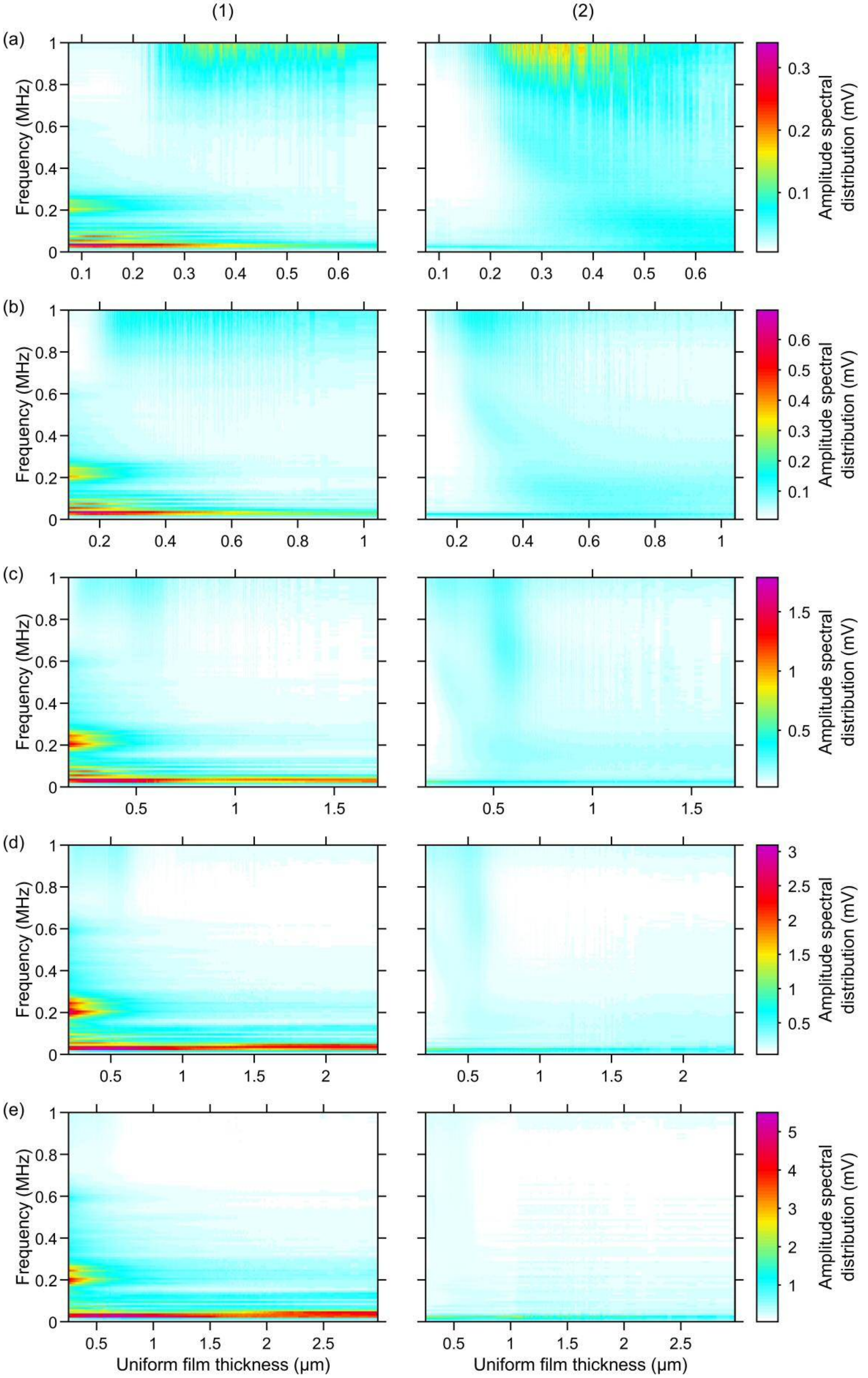


Figure 4-25 10 kHz binned AE spectral distribution vs. the uniform film thickness for columns: (1) the rough disks and (2) the smooth disks; and for rows: (a) 300, (b) 500, (c) 1000, (d) 1500, and (e) 2000 rpm.

Now consider the low frequency range, 20 to 600 kHz, of the slowest test stage, Figure 4-25 (a). As expected, at the low end of the uniform film thickness range, the AE amplitude increases for the rough disks but not for the smooth, thus indicating the difference in the amount of asperity contact. However there is a complication. At the upper end of the uniform film thickness range, the amplitude of the low frequency AE from the smooth disks was actually higher than that from the ground disks, despite there being less asperity contact. And as the uniform film thickness decreased the AE amplitude also decreased despite there being a marginal increase in the amount of asperity contact (as shown by the contact voltage measurements). This phenomenon is more clearly evident when comparing the $FAE_{RMS\ 150\text{-}300}$ between disk pairs.

Figure 4-26 presents the $FAE_{RMS\ 150\text{-}300}$ vs. the uniform film thickness for all test stages and both disk pairs. Also shown are the pre-test and post-test $FAE_{RMS\ 150\text{-}300}$ measurements. They were taken just before the load was applied to start the test, and after it was subsequently removed to end it. For these measurements the rig was at speed but the disks were completely apart, i.e. there was no EHL contact, and so they have no corresponding uniform film thickness. They are plotted at the appropriate end of the uniform film thickness ranges for ease of chronological comparison.

From Figure 4-26 it can be seen that for the rough disk pair, the pre and post test $FAE_{RMS\ 150\text{-}300}$ was always less than that measured during the testing. Additionally there was no significant difference between the pre and post test values. (As a side note this indicates that for the rough disks procedure, the difference in rig temperature at the start and end of a stage had negligible effect on the AE from the *unloaded* rig). In contrast, for the smooth disks, the unexpected decrease in $FAE_{RMS\ 150\text{-}300}$ from one end of a test stage to the other, i.e. from maximum to minimum uniform film thickness, was matched by a very similar drop from the pre to post-test $FAE_{RMS\ 150\text{-}300}$. This is unlikely to be a coincidence, and it suggests that, in the case of the smooth disks, the $FAE_{RMS\ 150\text{-}300}$ response was not related to AE from the test contact at all, i.e. it was from extraneous and unidentified sources. Despite this, it is interesting to note that the unidentified source consistently peaked at a uniform film thickness of approximately 0.5 µm for all test stages except the fastest. Ultimately the unidentified low frequency AE source, present only for the smooth disks tests, does not affect the validity of the rough-smooth comparison, but it does highlight the complexity of using AE measurements in rotating machinery, where they are many potential sources.

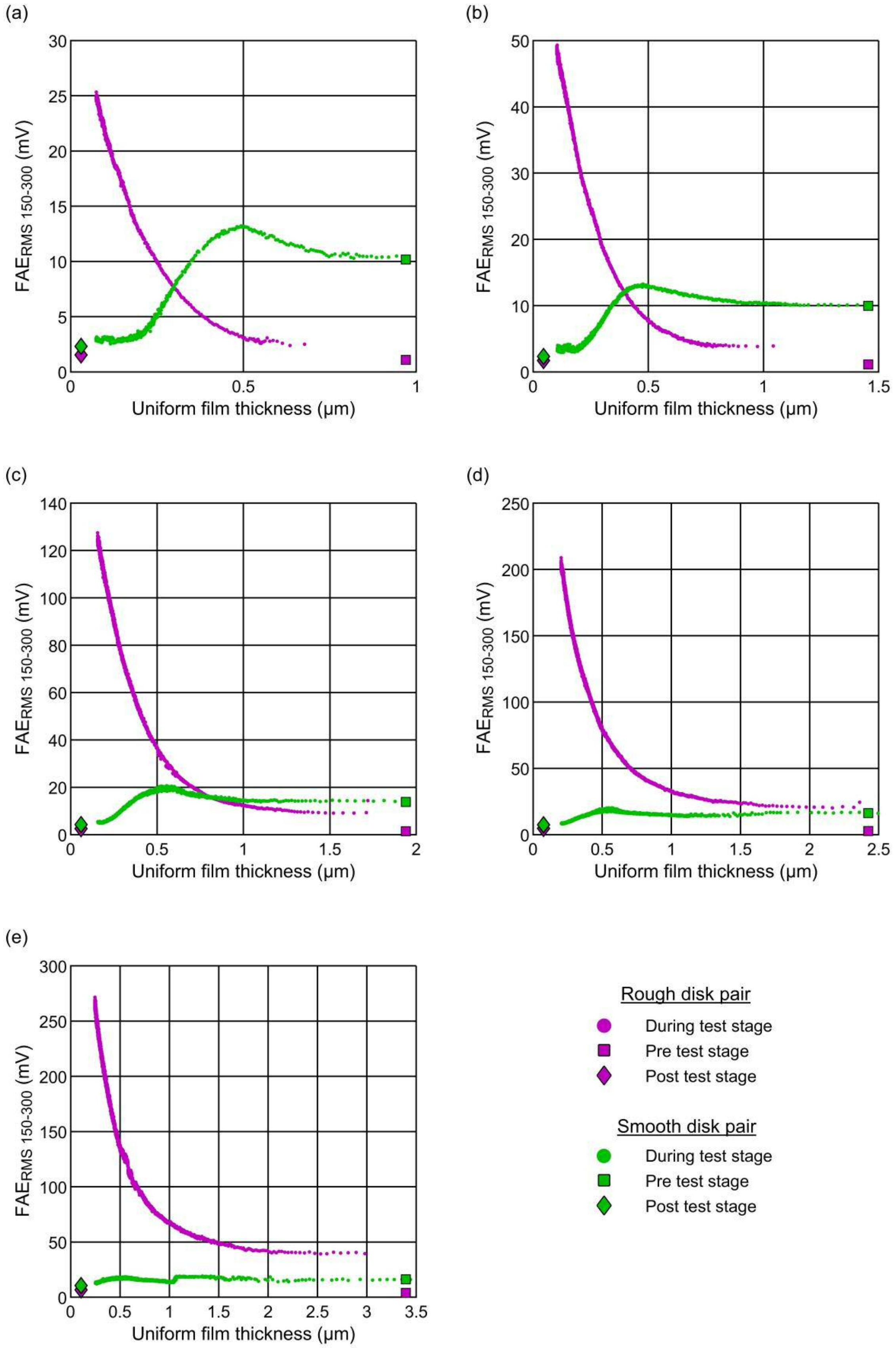


Figure 4-26 Mean of filtered (150 – 300 kHz) AE vs. the uniform film thickness for both disk pairs at (a) 300, (b) 500, (c) 1000, (d) 1500, and (e) 2000 rpm.

4.11 Review

This section presents a review of the experiment presented in this chapter.

Using a rough disk pair and by varying temperature and speed, a range of different lubrication conditions, from full film to mixed, were induced, and these were verified by the contact voltage measurements. Measurement of the RMS of the broadband AE revealed it to have an inconsistent response to changes in the amount of asperity contact, particularly at low speeds. However a spectral analysis identified a narrower frequency range whose RMS amplitude varied in accordance with the amount of asperity contact. It has been shown, that to obtain the highest quality AE measurements a suitable filter needs to be imposed. There is no reason to think this would not be the case for other similar applied contacts and yet in papers involving these there tends to be scant discussion of the filtering. Although the parameters are usually stated there is often no reason given for the specific choice of filter. This leaves open, questions of if the measurements are of optimal quality, or even worse, if they are what they purport to be.

The form of the increase in the $FAE_{RMS\ 150\text{-}300}$ (RMS of the 150 – 300 kHz pass band), as the lubrication conditions transitioned from full film to mixed, was convincing evidence that it was detecting the amount of asperity contact. And comparison with tests on smooth disks confirmed that the source of AE in this frequency range was indeed from asperity contacts. That increasing asperity contact causes increased AE, is not a new assertion, it has already been made by the Cranfield researchers and others, but it can be stated with more certainty here. By virtue of the simplified contact (disk rather than gear), and the more detailed measurement of the contact conditions, these results improve the confidence in the fundamental truth of this. Accurate and precise mathematical relationships between the AE from asperity contacts and the specific film thickness, and the AE from asperity contacts, specific film thickness and speed, have been presented.

By comparing the AE measurements from the rotating sensor with those from the static sensor it has been shown that the novel interface of the latter: an oil-coupling, is fundamentally feasible. Although, the filtered results from the static sensor could not match the precision or consistency of those from the rotating sensor they still coarsely indicated the increase in AE with increasing asperity contact.

Some limitations of the experiment are now discussed. Although changes to the lubrication conditions were successfully induced using a temperature rise, there was no control over the *rate* of rise. This led to a highly non-linear rate of decrease in the specific film thickness, it approximated an exponential decay, and as a consequence the sample density at high specific film thicknesses was relatively low. Additionally it almost certainly aggravated the errors between the measured and actual disk temperatures, which are critical for the accuracy of the specific film thickness calculation. Although the problem of varying sample density could be mitigated by tying the sample rate to the temperature rise, it is likely the overall quality of the experiment would be improved if the rate of temperature, and therefore film thickness, change could be made more uniform. To do this a method of active cooling would need to be used.

The next limitation of the experiment was the range of constant-speed lubrication conditions that were induced. At slow speeds it was not possible to induce full film conditions and at fast speeds it was not possible to proceed very far into the mixed regime. This led to only a limited range of specific film thicknesses tested at all speeds and although this was not a major issue, larger ranges would have been useful. On that note, it would be interesting to test lubrication conditions that extended further towards full-boundary lubrication conditions, to try and identify if at some near point at the AE started to approach a limit. At present it is essentially unknown how far into the full possible range of low-wear conditions has been tested.

Another limitation of the experiment is that its repeatability has not been established. Only one set of disks was tested and for this each combination of temperature and speed only once. Repeating each test stage, or the whole procedure, a number of times in succession might help identify measurement error and should reveal the scale of wear effects. It would also be interesting to test for hysteresis, i.e. does the relationship between AE and specific film thickness vary depending on if the temperature is increasing or decreasing? Again, active cooling would be required to investigate this.

5 Wear experiment

5.1 Introduction

This chapter, like the previous, explores the relationship between AE and the lubrication conditions. Previously, the changes in the lubrication condition were instigated by deliberate and large variations in the uniform film thickness, with only incidental and negligible wear. This time the changes were instigated by a procedure designed to allow a disk pair to naturally and substantially wear, with only minimal and incidental changes to the film thickness. Much of the focus of the previous chapter was on using the change in lubrication conditions to prove a good correlation between the amount of asperity contact and the AE measurements. For this chapter the relationship between AE and asperity contact is taken as accepted, and consequently, the chapter focusses on *what* the AE measurements can reveal about the wearing process, rather than simply *if* they can detect it.

As with Chapter 4, the procedure utilised both periodic surface measurements, and real-time measurements: the AE, contact voltage, temperatures and so on. After the specific procedure has been detailed, the results of the surface measurements are presented first, and are used to describe the type and rate of wear. Following this, the real-time measurements are then presented and assessed in relation to the identified wear.

5.2 Procedure

For convenient reference Table 5-1 presents a summary of the procedure, including information on the disks, contact parameters, test schedule, and measurement settings.

As discussed in Section 2.10, two types of wear[1] commonly occur in the mixed lubrication regime: running-in and micro-pitting fatigue. This procedure was designed to investigate both of these using a single, and previously unused, rough (as-ground) disk pair. For ease of analysis it was desired to keep both wear regimes very simple. Thus the entire procedure used constant control parameters and constant specific contact topography. To maintain consistency with the variable film thickness experiment, the same nominal load of 1460 N, and slide roll ratio of 0.5, were used for this procedure. This left the choice of speed and temperature to dictate the severity of the lubrication conditions, i.e. control the propensity for wear. There was no easy way of pre-determining the optimum lubrication conditions; too aggressive and there might have been the risk of catastrophic surface damage, or too benign and the rate of wear might have been too slow for study. Based on previous experience, it was decided to use a fast disk speed of 1000 rpm and an oil bath temperature of 80 °C. This combination generated a specific film thickness of approximately 0.6 during the variable film thickness experiments, so it was reasoned that it would ensure mixed lubrication in this experiment, even after running-in.

---

[1] Wear in the broad sense of any process that results in a change in roughness whether beneficial or adverse

The testing of new, or unused, disks was extremely important for this procedure for two reasons; first it guaranteed that there would be a significant geometric run-in to study, and second, it eliminated the possibility of an operational history affecting the rate of wear, i.e. there was no pre-existing fatigue. Using new disks it was certain that there would be an initial short and aggressive period of wear due to running-in, and it was hypothesised that this would be followed by a period of stability and then finally micro-pitting. To monitor the wear it was necessary to break the procedure into increments, or stages, so that surface measurements could be made.

Because running-in and micro-pitting were expected to occur over very different time scales, different sample regimes were used for each. The work of Weeks [57], showed that geometric running-in of the test disks was typically completed in 30 seconds or less. It was decided to record the AE continuously, i.e. as one long sample, over two consecutive short test stages of 3 minutes each. The first of these was to run-in the disks and the second was to verify that no further running-in would occur. For both stages the AE sample was started before the load was applied, and only stopped after it was later removed. The mean of the other measurements was calculated once for every fast disk rotation. At the time the rig control program could not record the contact voltage continuously, so only the mean contact voltages were used.

It was expected that it would take some time for micro-pitting to occur and so after the first two ‘running-in’ stages, the subsequent ones were considerably longer and only utilised periodic AE sampling. There were fifteen subsequent stages in total, with each of the first ten lasting 100 minutes and each of the last five lasting 300 minutes. The continuous AE and contact voltage were sampled every 100th fast rotation for the first ten and every 500th for the last five stages. The reason for the change in procedure is discussed once the roughness results have been introduced in the next section. Unlike the testing in the previous chapter, for this test it was desirable to minimise temperature change, so before starting each stage the oil bath was heated to temperature and the oil circulated through the test enclosure until it reached thermal equilibrium.

To record the wear history of each disk, in-situ and repeat position profile measurements were made at nine positions before and after every test stage. The measurements were made at a single circumferential position and were equispaced across a 4 mm axial width in the centre of the disk, (there was a measurement every 0.5 mm within the 4 mm range). Given that the Hertzian contact width was only 3 mm it was anticipated, and intended, that some of the measurements would fall outside the contact path.

Table 5-1 Summary of the test procedure

Disks	
Surface finish	As-ground for both
Operational history	None, new disks

Contact parameters		
Constants	Slide roll ratio, [Gear ratio]:	0.5, [5/3]
	Specific contact topography (assembly orientation of disks):	Fixed
	Fast disk speed:	1000 rpm
	Oil bath temperature:	80°C
	Load calibrated at 100°C:	1460 kN
	Maximum Hertzian contact pressure at calibrated load:	1.2 GPa
	Hertzian contact dimensions at calibrated load:	0.8 x 3.0 mm
	Oil:	OEP-80
Variables	None	

Test schedule				
Test stage/s	*Description*	*Duration per stage (min)*	*Fast disk cycles per stage (approximate)*	*Cumulative fast disk cycles by end of stage/s (approximate)*
1	Running-in	3	3000	3000
2	Running-in validation	3	3000	6000
3 to 12	Fatigue testing	100	1 x 10^5	10 x 10^5
13 to 17	Fatigue testing	300	3 x 10^5	25 x 10^5

Sample timing				
Start up	Dual pulse method. Intra and inter stage synchronisation			
Means	*Data:*	Speed, load, temperatures, contact friction and contact voltage		
	Frequency:	Once per fast disk rotation		
	Duration:	Three quarters of a fast disk rotation		
Continuous signals	*Test stages*	*Data*	*Frequency*	*Duration*
	1 and 2	AE only	Continuous	Entire stage
	3 to 12	AE and CV	Every 100th fast disk rotation	5 fast disk rotations
	13 to 17		Every 500th fast disk rotation	

AE measurement settings	
ADC:	2 MHz per sensor
Pre-amp:	40 dB
Filtering:	20 kHz high pass

Profile measurements	
Frequency:	At beginning and end of test schedule and in-between each stage
Positions:	Nine axial positions per disk, evenly distributed across a range of ± 2 mm from the disk centre-line, and at a single circumferential position.
Length:	12 mm
Filtering:	0.8 mm Gaussian

5.3 Surface roughness measurements

This section presents the surface measurements which are used to describe the rate and type of wear that occurred over the course of the experiment.

5.3.1 R_q (RMS roughness)

Figure 5-1 presents the R_q of each disk's contact path vs. the cumulative fast disk rotations. Each value was calculated from five central profiles as these were found to be well within the width of the contact. It can be seen, as evidenced by the consistent vertical offset, that the sampled circumferential position of the slow disk was inherently rougher than that of the fast disk; but more importantly that the chronological change in was similar for both disks. This is a good preliminary indication that the wear regime was consistent across both disks.

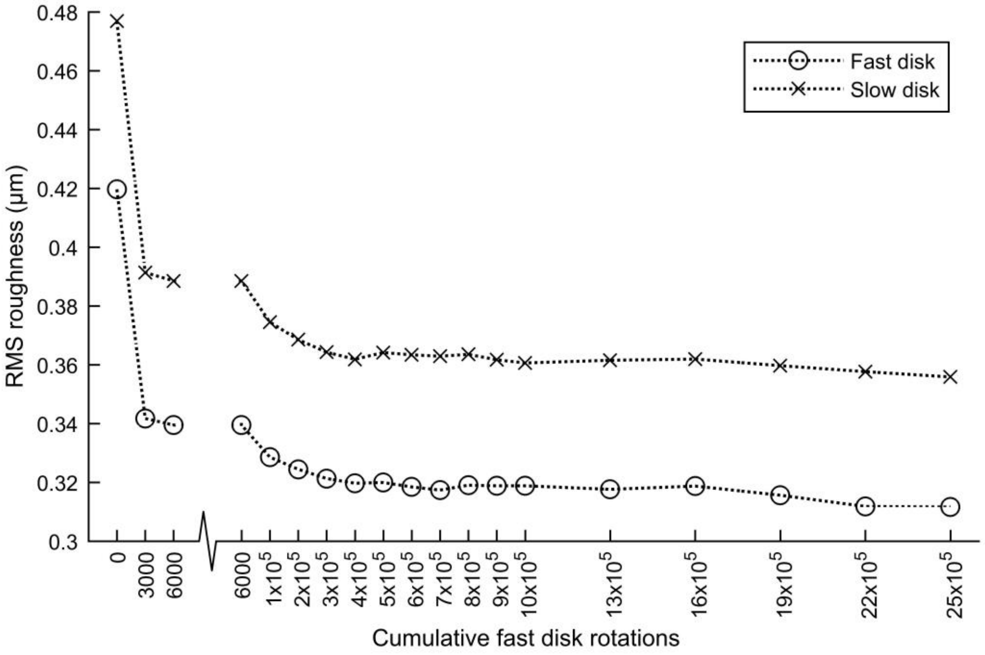


Figure 5-1 The (RMS Roughness) of the contact path vs, cumulative fast disk rotations. The was calculated using the five central profiles. Note the break in scale at 6000 cycles, this measurement is shown twice, once on either scale.

Considering now the details of the chronological change. Over the first short test stage, there was a dramatic reduction in , which is a typical effect of geometric running-in [57,61]. But then subsequently, over the next short test stage, there was only a negligible further reduction. Of the total, which corresponded to six thousand fast disk rotations, 97 % occurred in the first stage and only 3 % in the second. This confirmed that the surfaces had reached a relatively stable-state, i.e. had completed the geometric running-in, by the end of the *first* stage[1].

It was originally hypothesised that after the running-in there would be a lull in the wear before micro-pitting started however the results certainly do not reveal one. Figure 5-1 shows

---

[1] Despite the fact that the running-in was concluded by the end of the first stage both the first and second together are described as the 'running-in experiment.'

that after the initial running-in there was then a gradual asymptotic reduction in that appears to have started immediately. This does not necessarily indicate there wasn’t a lull in the wear, only that it must have been much shorter than the period of the third stage: one hundred thousand fast disk rotations. The asymptotic nature of the reduction suggests that, like the geometric running-in, the subsequent wear was of a self-limiting type. By the time one million fast disk cycles had been accumulated the rate of wear had become negligible, and so surface measurements were then taken less frequently. At 2.5 million fast disk cycles the experiment was stopped as the rate of wear was too low for further study.

Before moving on to considering the roughness profiles it is worth considering the values in a little more detail. Figure 5-2 presents the vs. axial position for every profile measurement on both disks. Colours are used to differentiate between measurements made at different stages of the experiment. For both disks, six of the central axial positions, from -1 to +1.5 mm, show a large and similar reduction in , this indicates these positions were well within the contact path. Outside of these, there are some axial positions which show a more modest reduction, and these indicate a transition region at the edge of the contact path. And finally, at the limits of the axial range measured, there are some profile measurements where there is no reduction, these are outside of the contact path. The results suggest that there was some inherent axial variation in the of the fast disk which remained in the contact path despite the wear.

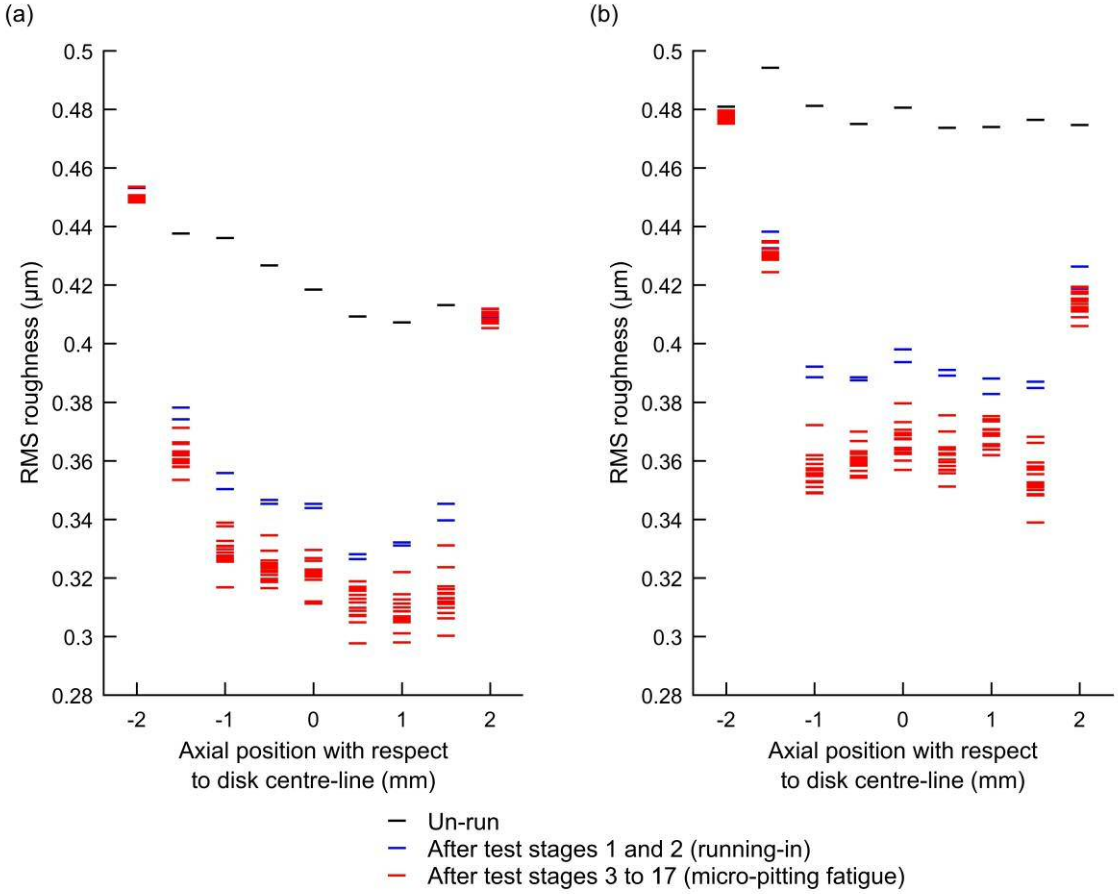


Figure 5-2 vs. axial position for (a) the fast disk and (b) the slow disk.

5.3.2 **Topography**

The measurements indicate the broad scale and rate of surface modification but provide no insight into the specific wear mechanism(s) involved. For this it was necessary to examine the specific topographical changes and these are the subject of this section.

Considering first the running-in. Figure 5-3 presents a sample of the aligned roughness profiles from surface measurements taken: in the un-run state, and after each of the first two short stages, i.e. after 0, 3000 and 6000 fast disk rotations. The sample is from the centre-line of the fast disk, however the specific disk and axial position are not important. Aligned profiles from all axial positions within the contact path and on both disks were examined, and no significant differences in topographic behaviour were observed.

The first thing to note is that there is very little difference between the profiles measured at 3000 and 6000 fast disk rotations, as shown by the lack of the green evident in the trace. This supports the confirmation, by the , that the running-in had been completed by the 3000th fast disk cycle and that subsequently the surface roughness remained relatively stable over the next 3000 rotations. In this sample there is only one significant deviation between the two run-in profiles, and this is labelled by arrow A. Here a valley appears to close-up but this is very likely an illusion caused by axial miss-alignment rather than a genuine change in topography.

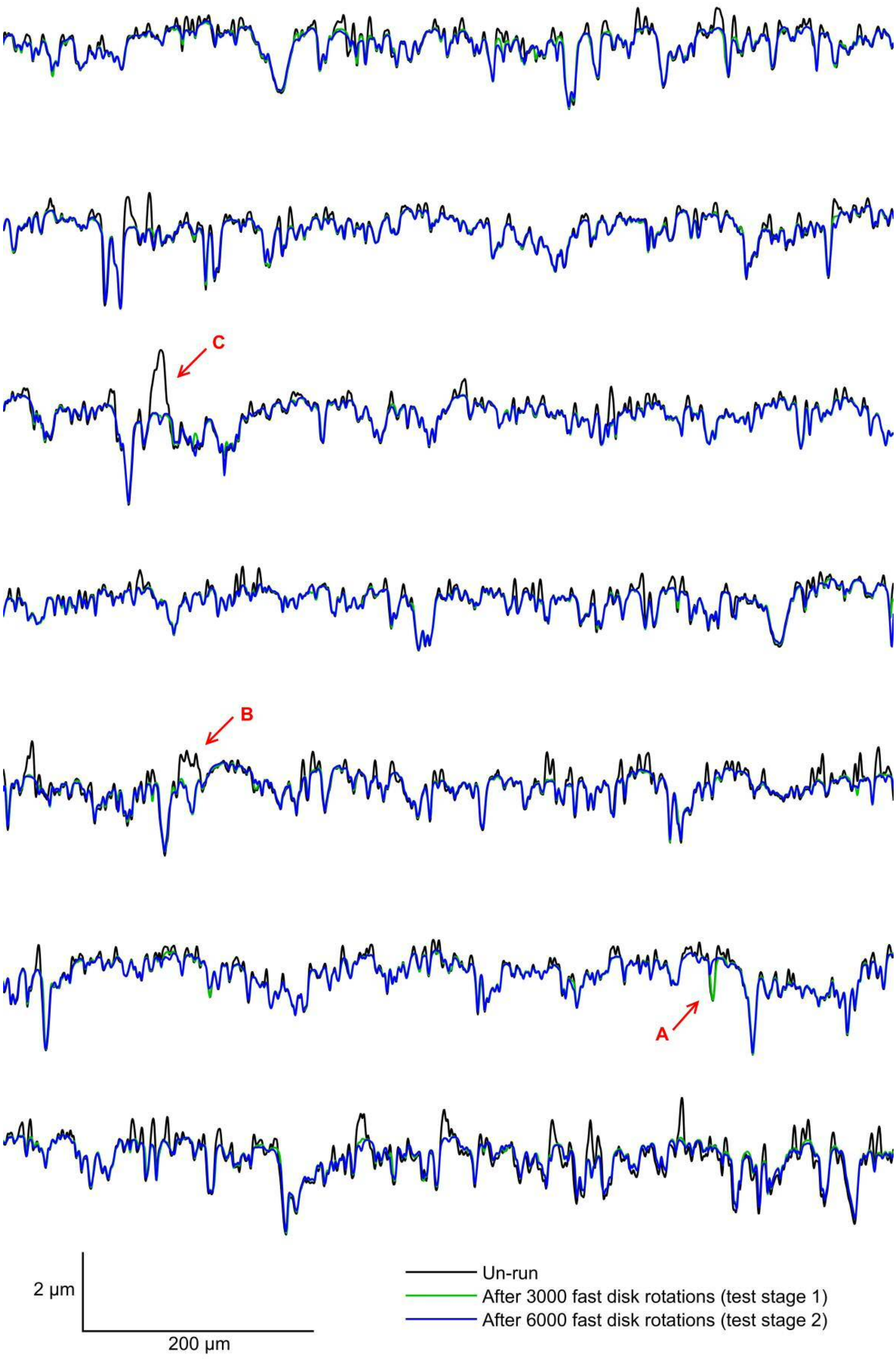


Figure 5-3 Aligned profile measurements from the centre-line of the fast disk. The profiles are from continuous measurements, they have been sectioned and vertically stacked for efficient presentation. The disk material is below the traces.

Now comparing the differences in the profiles before and after running-in. It can be seen that continuously along the sample length running-in has truncated, or rounded off, the highest ridges but has had no effect on the valleys. This type of topographical change has already been observed and described in detail by researchers in another similar running-in experiment on the twin disk rig [57,61], and they attributed the *predominant* mechanism of change to plastic deformation rather than wear, i.e. redistribution rather than removal of material. However, these results suggest that wear or removal does play a significant part in the running-in. Consider the specific run-in for two of the most prominent ridges labelled by arrows B and C. These, have not simply become more rounded or flattened versions of their previous forms, but have been replaced by significantly different forms, which include new valleys; this suggests a fracturing process. Additionally, the volume of material affected is large, (as indicated by the area bounded by the divergent profiles), and there is no evidence of it having been redistributed, instead it is simply gone. Remember, the profiles are cross-sections of ridges, not spires, and so the material cannot have deformed axially, i.e. out of the plane of the profile. These two specific ridges were checked at the other axial positions measured in the contact path and the changes were found to be similar along their lengths. The problem of where plastic deformation can have deformed the material to, is not limited to just these two specific ridges, it is merely most prominent for them. A qualitative assessment suggests a local discrepancy in material volume before and after running-in for many moderate to large sized ridges[1]. It is concluded that this running-in consisted of both material fracture and plastic deformation which both acted to remove asperities.

Now moving on to examine the post running-in topographical changes, i.e. those from the 6000th fast disk cycle onwards. First the final state of the surface will be considered. At the end of the experiment the disks were disassembled from the rig which allowed their final surfaces to be measured using 3D profilometry. Figure 5-4 (a) shows the surface roughness of a 6 x 6 mm area of the fast disk measured at the same circumferential position as the 2D in-situ profiles. This area encompasses the entire width of the contact path and includes some of the un-run surface to either side. There is a visual distinction between the contact path and the un-run surface, but it is somewhat subtle using this plotting technique. To help identify the contact path, Figure 5-4 (b) shows the vs. axial position. The axial range between approximately 1 and 4 mm has a lower, if fluctuating, which indicates the contact path. Either side of this there is a significant transition region where the gradually increases to that of the un-run surface. The existence of a transition region was earlier suggested by the axial variation of the in-situ roughness measurements (Figure 5-2), but it is much clearer here. And again evident is the slight increase in inherent roughness across the width of the disk.

---

[1] It would be interesting to mathematically compare the difference in material volume, i.e. difference in area under the profiles, but to do so a robust vertical alignment needs to be ensured.

The 3D roughness measurement presented in Figure 5-4 (a) confirms that micro-pitting had occurred on the fast disk by the end of the experiment. The micro-pits are evident as dark speckles, or dots, within the contact path. They can be found across the entire width of the contact but tend to be clustered along relatively few select ridges. Figure 5-4 (c) shows an enlargement of a set of ridges afflicted with a particularly prominent cluster of micro-pits. For comparison Figure 5-4 (d) shows an enlargement of the same set of ridges but at a position outside of the contact path. Before running-in, the length within the contact path would have been similar in appearance to this. The 3D roughness measurement of the slow disk found micro-pitting on this too. Visually it appeared to be of a similar scale and pattern to that found on the fast disk and so for brevity the slow disk measurement is not presented.

Although the final 3D surface measurements are useful since they explicitly demonstrate that micro-pitting had occurred by the end of the test, they cannot be used to identify *when* it occurred. It is tempting to assume that the asymptotic-like reduction in after running-in was caused by micro-pitting, which would imply it must have occurred at the same rate. However this is not necessarily the case. For one reason, it is as yet unknown if micro-pitting was the *exclusive* cause of the wear after running-in. For another reason it is not certain that micro-pitting would cause a reduction in the since it occurs on the scale of the pre-existing roughness features. Other simple statistical parameters including the skewness and kurtosis were examined, but when plotted against time neither of these showed any convincing systematic trend after the run-in.

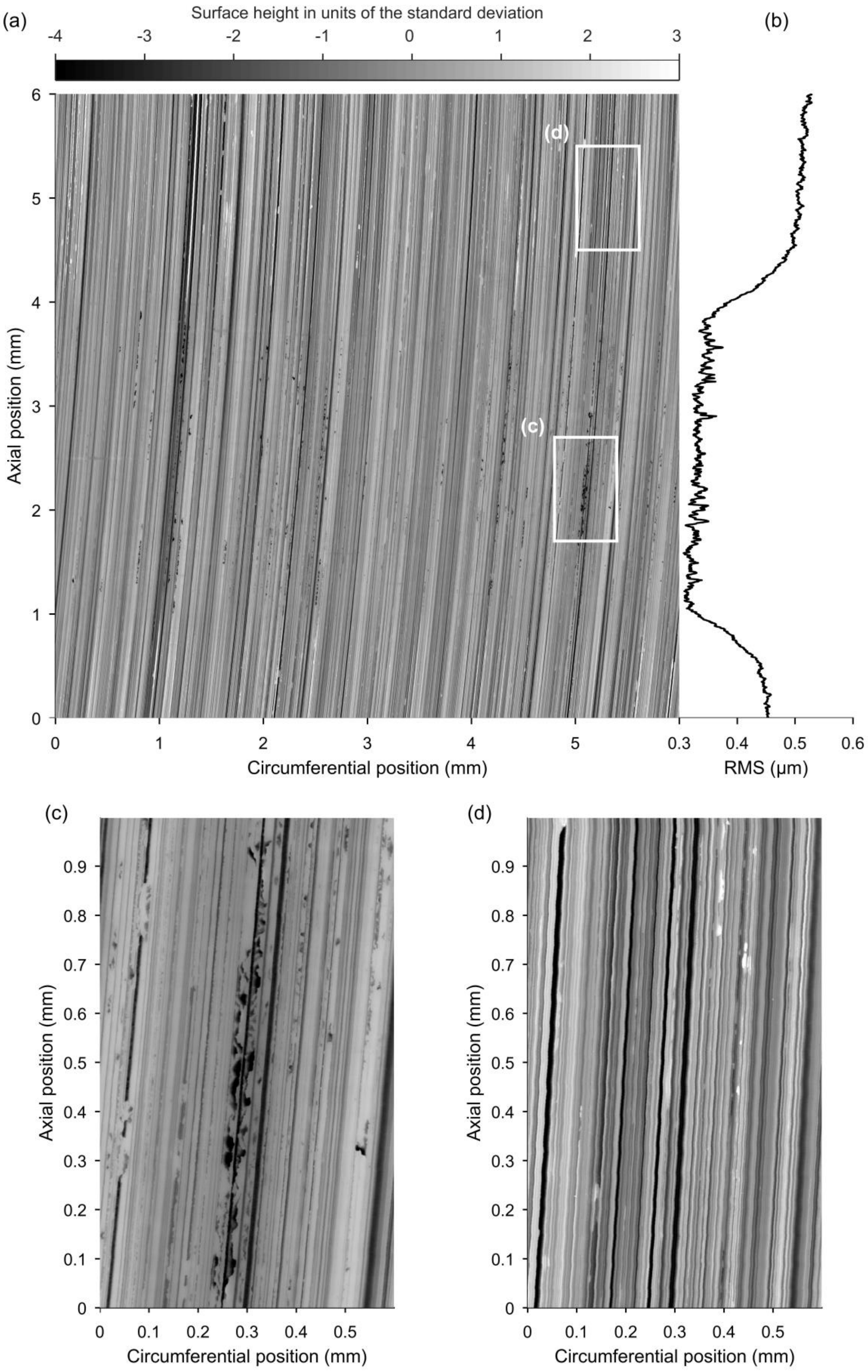


Figure 5-4 Areal surface roughness measurement of the fast disk made at the end of testing using 3D profilometry. (a) A 6 x 6 cm area at the same circumferential position as the 2D profile measurements. (b) The vs. axial position of area (a). (c) Close-up of an area in the contact path. (d) Close-up of an area outside the contact path for comparison.

To identify when the micro-pitting occurred it is necessary to examine the aligned profiles. Figure 5-5 presents a plot of the aligned profiles from all the measurements made after running-in. The profiles are overlaid in chronological order with a colour scale ranging from blue through to green, orange and then red. Shown above the profiles is the standard deviation (RMS) of the surface height chronology vs. the circumferential position. This essentially describes how much each discrete position varied with time, and the reason for including this will become apparent shortly. Figure 5-6 presents an enlargement of a short section of the aligned profiles in Figure 5-5 to illustrate the detail.

It is easy to visually identify the locations which had become prominent micro-pits by the end of the experiment. These are the short lengths where there is substantial deviation in form between the profiles leading to a confusing mess of overlaid ridges and valleys. An example of a location which micro-pitted can be seen in the centre of the enlargement (Figure 5-6), and is identified by the range labelled A. From Figure 5-5 it can be seen that micro-pitting locations are relatively infrequent, and in-between each micro-pit there are relatively long lengths of surface which remained comparatively stable after running-in. However, although the most dramatic topographical changes occurred at micro-pitting locations there is also evidence of a different type of subtle but more uniformly distributed type of wear. Upon close inspection it can be seen that many of the asperities which did not pit appear to have been progressively, but only slightly, worn down. Two examples of this can be seen in the enlargement and are identified by ranges labelled B. There is thus some uncertainty as to contribution of these two different types of wear on the reduction in .

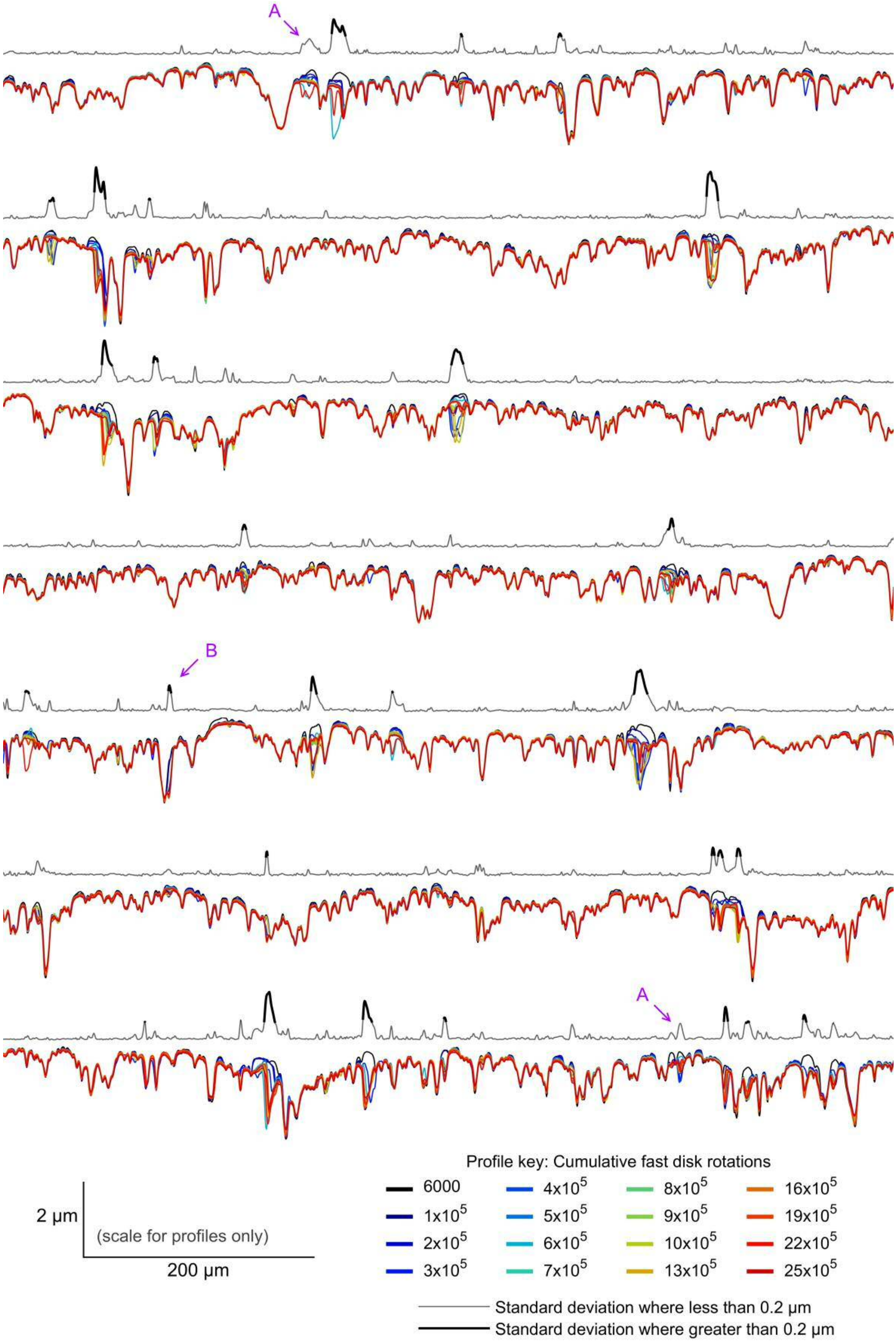


Figure 5-5 Aligned profiles from all measurements made after running-in. The section shown is from the fast disk centre line. Shown above the profile trace is the standard deviation (RMS) of the surface heights vs. circumferential position. The profiles are from continuous measurements, they have been sectioned and vertically stacked for efficient presentation. The disk material is below the traces.

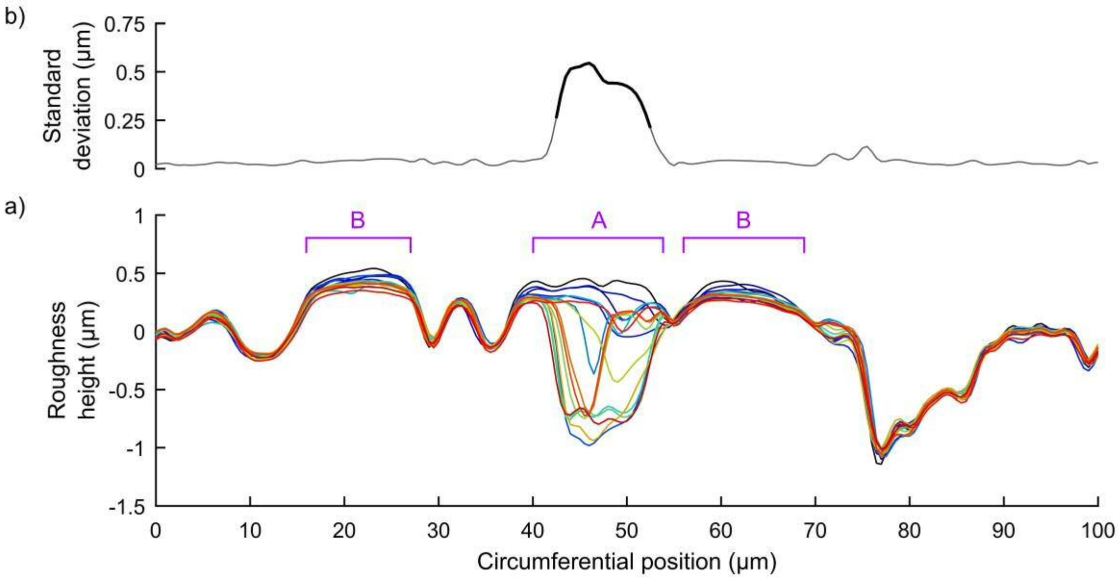


Figure 5-6 An enlarged section of the aligned profiles presented in Figure 5-5, the colour key is identical to that of Figure 5-5.

Although it is possible to use a visual assessment to identify when a *specific* micro-pit likely formed, this is of little use. And even on an individual basis this is not straightforward for, as discussed earlier in Section 4.3, axial alignment error can give the illusion of both pit growth and shrink. It would be more useful to identify *on-average* when the micro-pitting occurred, i.e. how the rate of pit creation changed over the experiment. To do this it is necessary to identify their locations algorithmically. A simple way to do this is by consideration of the standard deviation of height chronology vs. circumferential position. It can be seen in Figures 5-5 and 5-6 that there are substantial peaks in the standard deviation at the pitting locations and this can thus be used to distinguish between pitting and non-pitting locations. In Figures 5-5 and 5-6 the standard deviation is shown in bold where it is greater than or equal to a threshold of 0.2 μm. There are however some problems with the use of a simple threshold like this. Firstly, it has only identified the centre of the pitting locations, not their entire width. Secondly it has a tendency to simultaneously exclude some pitting locations and include some non-pitting locations. Examples of these are given by arrows A and B respectively in Figure 5-5. It is probable that both problems could be mitigated by the use of a more sophisticated thresholding / identification scheme but this was beyond the scope of this thesis. As a word of caution it should be noted that although the standard deviation identifies micro-pitting locations for this *specific* experiment, it may not do so for others; in short it should not be assumed to be generally effective.

Figure 5-7 shows, for each disk individually, how the average surface height of the pitting locations changed over the experiment. The averages were calculated from pitting locations extracted across the five central profiles using the standard deviation threshold of 0.2 μm. Also shown is the change in average height of the non-pitting locations and the average of the entire measurement length. Although Figure 5-7 shows the height change from the very start of the experiment, i.e. before running-in, it is important to note that the pitting locations were only identified using the standard deviation from profile measurements taken *after* the running-in.

First consider the average height of the entire profiles length, i.e. making no distinction between pitting and non-pitting locations. For both disks the average height declines during running-in and the early stages of the fatigue experiment before becoming nearly constant. As the profiles were vertically aligned using the valleys as a datum, the decline indicates an average wearing or redistribution of material from the heights to the depths which is to be expected. Now consider the average height of the prominent pitting locations. This is relatively high before running-in and remains so after. This is not surprising and is a somewhat trivial observation; it simply implies that micro-pitting occurred on the ridges rather than valleys. What is important is the form of the subsequent decrease. It can be seen that for both disks, the average surface height at pitting locations decreased dramatically, and from positive to negative values, in the initial stages of the fatigue experiment. Subsequently the average height remained at a relatively stable low level. This shows that the majority of the prominent micro-pits were formed in the early stages of the fatigue experiment, by ~ 300,000 fast disk rotations, and that very little significant pitting took place after this, which suggests that the micro-pitting was a self-limiting process. Identification of why the micro-pitting was self-limiting is not necessary for this thesis, but it is posited that it was because the asperities removed by micro-pitting caused the contacting surfaces to become progressively more conformal until there were no longer any directs contacts sufficiently aggressive to induce further micro-pitting. Alternatively it might be related to the geometric running-in. It has been shown that micro-pits preferentially form on asperities which were significantly reduced by running-in [57,61]. This suggests that the micro-pits are ‘seeded’ by the running-in process, and so once all these micro-pits have ‘germinated’ then no more will occur until a change of conditions initiates more running-in.

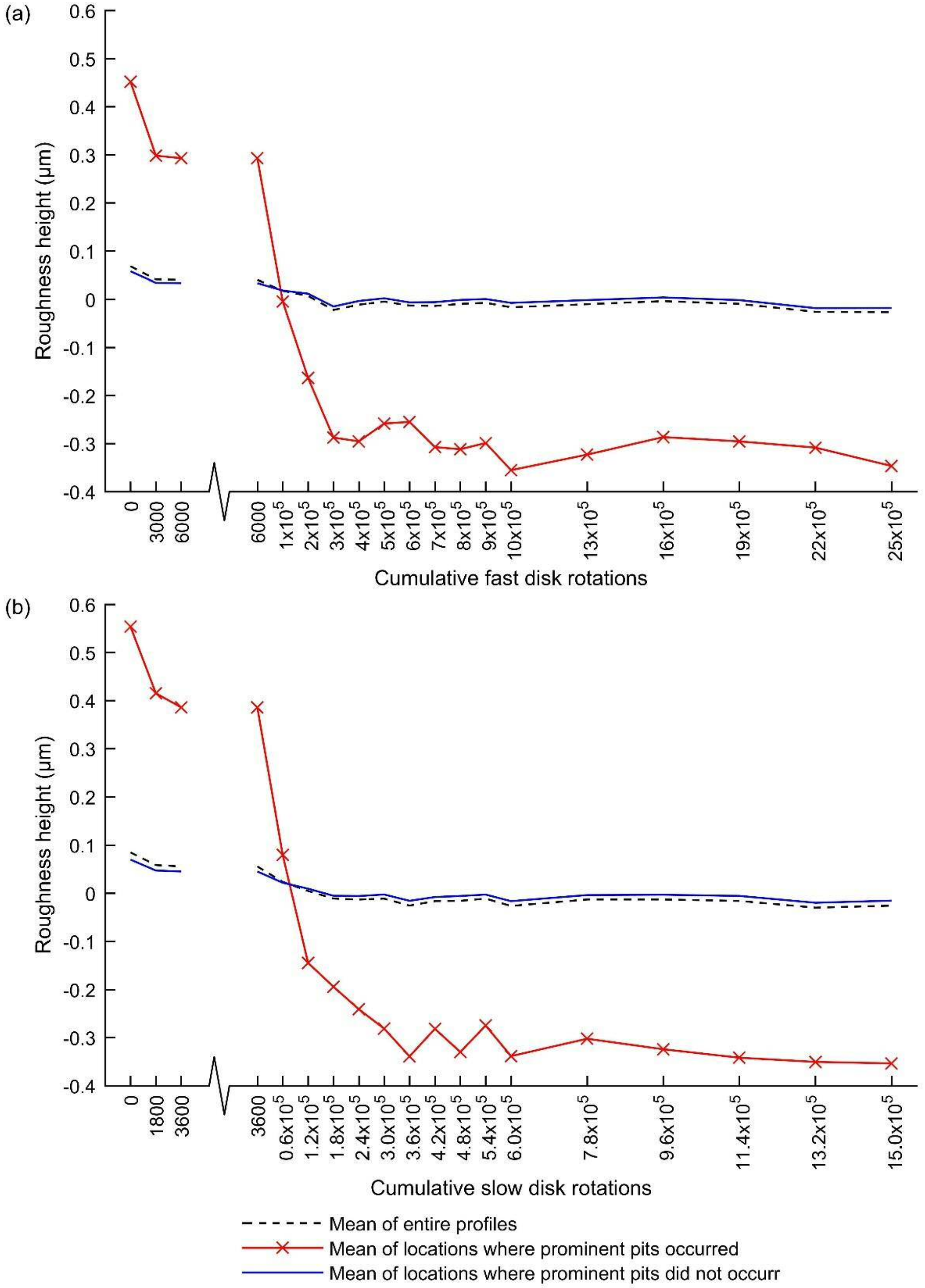


Figure 5-7 - Comparison of the mean roughness height of locations where pits developed and those where they did not. Calculated for the five central profiles of each disk using a standard deviation threshold of 0.2 µm. For (a) the fast disk and (b) the slow disk.

5.4 Running-in

This section investigates the AE measured by the rotating sensor during the running-in experiment, i.e. test stages one and two. The AE parameter used is the RMS of the 150 – 300 kHz pass band, the $FAE_{RMS\ 150\text{-}300}$. This parameter has been used as it was identified in Chapter 4 to provide a good indication of the amount asperity contact. The band-pass filter was the same as that used in Chapter 4 (details in Section 4.7.2). In Chapter 4 the $FAE_{RMS\ 150\text{-}300}$ was windowed 'naturally' by the interval based sampling programme. However for the running-in experiment a signal continuous sample was used. The $FAE_{RMS\ 150\text{-}300}$ vs. time was therefore calculated using a moving window.

Figure 5-8 presents the $FAE_{RMS\ 150\text{-}300}$, contact voltage, load, friction, and disk temperatures, vs. time for the first two short test stages. For this figure, the $FAE_{RMS\ 150\text{-}300}$ was calculated using a moving window with a width of 300 ms: equivalent to approximately five fast disk rotations, and which was moved in 15 ms increments. The measurements immediately preceding and succeeding each test stage are shown, and the start of each stage, i.e. the moment of load application, has been set as time zero. Firstly, it can be seen that, compared to the duration of each stage, the loading and unloading ramps are extremely short, so much so that they appear instantaneous. At the start of both test stages there was a transient spike in the load before it quickly settled at a steady-state value, this is an overshoot response caused by the speed of loading and it is inevitable. The overshoot for the second test stage *appears* to be significantly less than that of the first but, as will be discussed shortly, this may be a 'picket fence' illusion caused by insufficient temporal resolution.

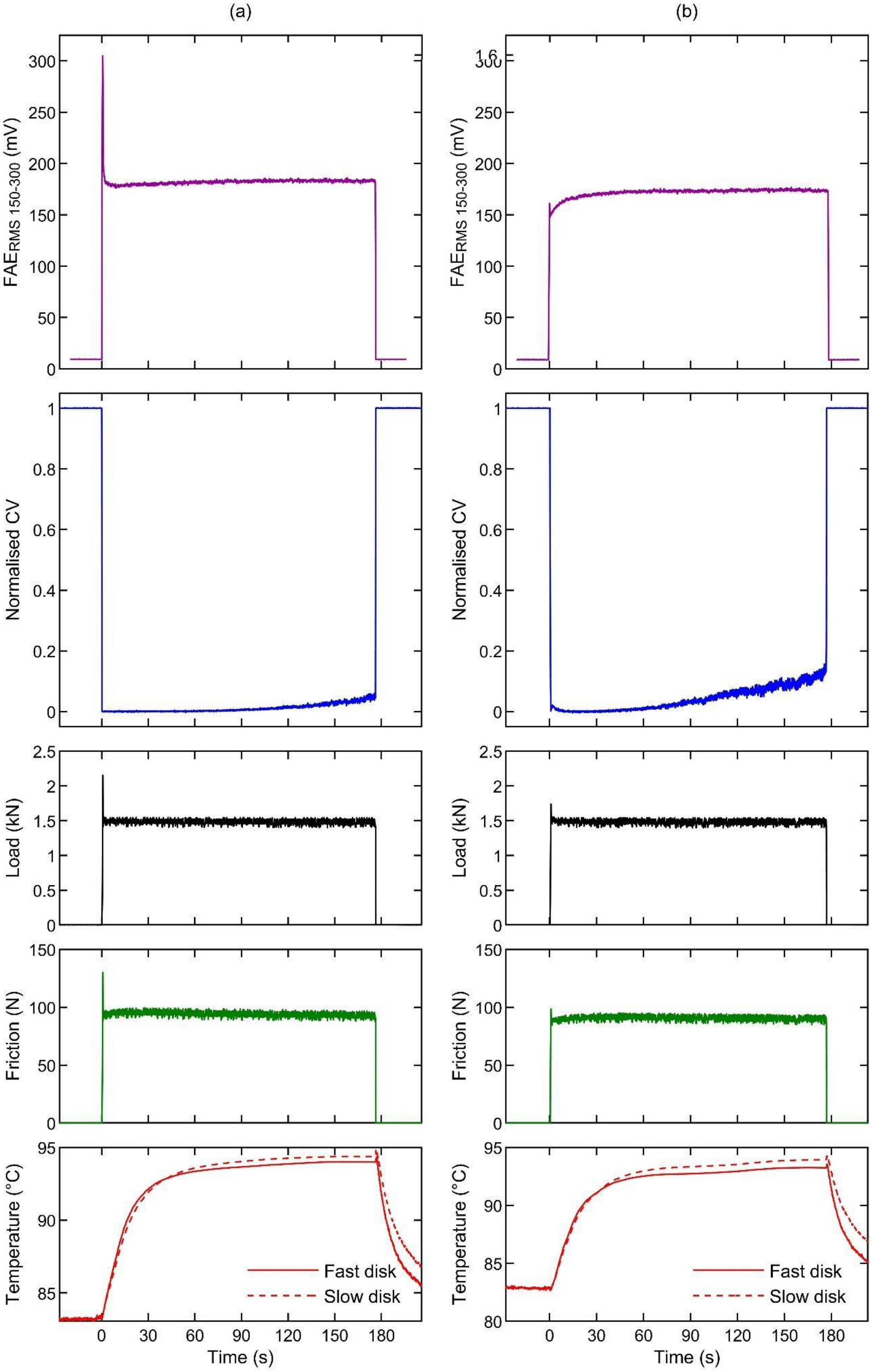


Figure 5-8 The FAE$_{RMS\ 150\text{-}300}$ (RMS of the AE signal in the 150 – 300 kHz pass band), contact voltage, load, friction and disk temperatures, vs. time for a) the first test stage: fast disk rotations 0 to 3000, and b) the second test stage: fast disk rotations 3000 to 6000. Also shown are the measurements immediately preceding and succeeding the test stages.

Considering the first test stage, that in which the geometric running-in occurred. Upon loading the contact voltage immediate goes to zero and remains at this level for approximately half the stage before increasing slightly towards the end. A value of zero indicates that there is continual, aggressive asperity contact and this is typical of the running-in which by definition must involve significant asperity contact. It should be emphasised that a constant contact voltage of zero does not indicate that the amount of asperity contact is constant, rather that it is equal to *or greater than* the measurement limit. Now consider the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$, upon loading this spikes immediately and then decays extremely rapidly to a much lower and relatively stable level. For the remainder of the test it increases very subtly. Remember, as was shown by the profile measurements, there was a substantial reduction in asperity height at some point in this test stage, and so the amount of AE from asperity contact should have subsequently dropped. So if the reduction in asperity height occurred gradually over the course of the stage, there should be a corresponding gradual decrease in the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$, however this is clearly not the case. The only significant reduction in the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ occurs due to the spike which suggests that the reduction in asperity height, i.e. the running-in, occurs on the time scale of this, and is concluded very early on in the stage. This complicates interpretation of the results as it means the effects of loading and temperature transients at the start cannot be ignored. It also necessities a more detailed examination of the start of the test stage. Before this is done however it is worth considering the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ and contact voltage measurements from the second test stage.

Remember that there was comparatively little change in surface topography during the second stage compared to the first, i.e. there was no running-in during this. Upon loading at the start of the second stage the contact voltage again drops to zero, but this time not instantly. There is also, upon close scrutiny, a very subtle transient dip which is due to the spike in load causing a corresponding spike in the asperity contact. Although the contact voltage does immediately drop to very low levels, the fact that it does not instantly reach zero is an indication of the reduction in surface roughness compared to the start of the first stage. The slight decrease towards zero is caused by a decrease in oil film thickness. After reaching zero, the contact voltage then increases gradually, which would appear to indicate a gradual wearing. It is curious that the increase is more pronounced than in the first stage where the majority of surface change occurred. It may be because the temperature of disks for the second stage was, as can be seen, slightly lower than at equivalent times in the first stage, and this would have induced a slightly thicker film, and thus less asperity contact. Also it is interesting that the increase in contact voltage appears to contradict the response of the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ and this will be discussed shortly.

The response of the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ at the start of the second stage is noticeably different to that of the first. There is still a spike, but it is diminished in prominence, and immediately after there is a tapering increase which is consistent with the expected effects of temperature on the amount of asperity contact. At the very start of the stage, the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ and contact voltage are congruous, with both indicating a spike then more gradual increase in asperity contact, but then subsequently they start to contradict each other, with the slight increase in $FAE_{RMS\ 150\text{-}300}$ indicating more asperity contact but the contact voltage indicating less. This contradiction also occurs in the latter half of the first test stage. It is difficult to explain the contradiction but the $FAE_{RMS\ 150\text{-}300}$ response certainly fits better with the temperature response and evidence of minimal wear. The results of the micro-pitting experiment, to be discussed in Section 5.6, suggest that the contradiction is most likely due to a start-up phenomenon only affecting the contact voltage.

Comparing the $FAE_{RMS\ 150\text{-}300}$ in test stages 1 and 2, it is clear that the only significant difference is at the start. This is more evidence that all the running-in occurred at the start of the first stage. Despite this, it is not clear from Figure 5-8 what proportion of the spike in AE is from running-in and what is from the load response, which also spiked. To help clarify this the beginnings of the two test stages are now examined in greater detail.

Figure 5-9 shows the $FAE_{RMS\ 150\text{-}300}$ and load, vs. time, for the initial 1.5 seconds of the first two test stages. A narrower moving window was necessary to examine the $FAE_{RMS\ 150\text{-}300}$ on this time scale, so a window width of 1.25 ms, equivalent to ~ 7.5° of a fast disk rotation, and increment of 0.25 ms were used. Whilst the window used for Figure 5-8 was sufficiently wide to filter out any AE oscillations at the rotational frequencies of the disks this is not the case for Figure 5-9. It can be seen that there were substantial oscillations in the $FAE_{RMS\ 150\text{-}300}$ and these occur at the rotational frequency of the fast disk. There are several possible causes of these oscillations, but the most likely is the cyclic variation in sensor position relative to the contact. This would cause a cyclic variation in attenuation so that the peak of the oscillation occurs when the sensor is closest to the contact, and the valley when it is farthest from it[1]. A cursory examination of the results from the static sensor revealed these to be free of the oscillations which supports this explanation, however it should be cautioned that the static sensor results were very noisy and this noise may have simply obscured the oscillations. Alternatively the oscillations of the $FAE_{RMS\ 150\text{-}300}$ might have been due to oscillations in the load caused by slight eccentricity of the disk, but as the load signal was averaged over each fast disk rotation this cannot be tested. Finally, but somewhat tenuously, they might have been due to oscillating levels of asperity contact caused by circumferential variation in the roughness of the disks. As the circumferential variation was not measured, this again cannot be tested, however it seems unlikely that the circumference of a disk would vary in such a way as to cause the triangular oscillations shown. And additionally, given that the cyclic period of the specific contact topography was five fast disk rotations some oscillatory behaviour on this wavelength would also be expected, but none is evident.

---

[1] As an aside, this might offer a way to distinguish between AE transmitted from the contact from that from other sources. If it was proved that the oscillations were due to variable sensor-contact distance then any AE that does not show the oscillatory behaviour cannot have been from the contact (for the rotating sensor measurements only!).

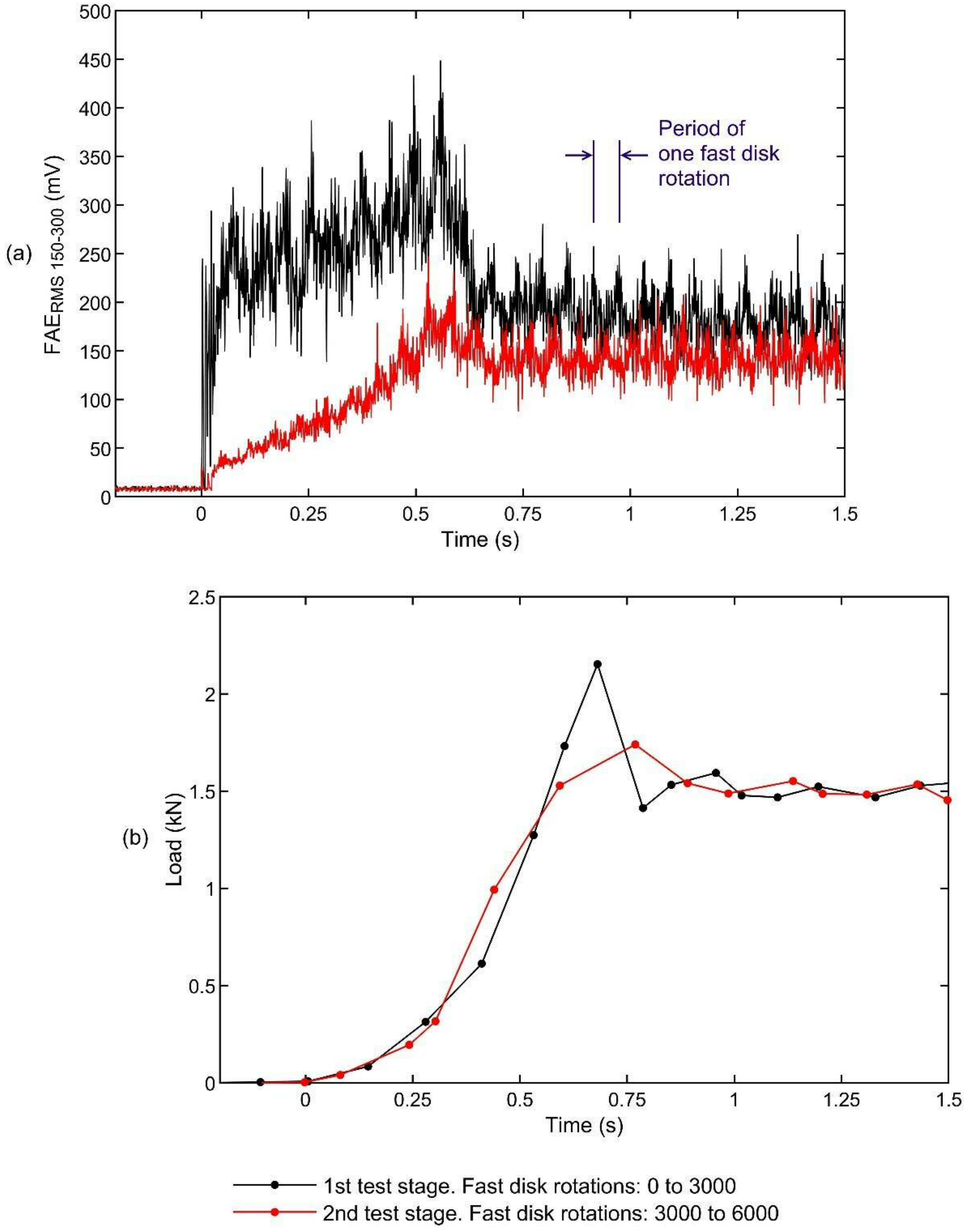


Figure 5-9 The (a) FAE$_{RMS\ 150\text{-}300}$ and (b) load, over the start of the first two test stages.

Regardless of the high frequency oscillations in FAE$_{RMS\ 150\text{-}300}$, Figure 5-9 shows that, comparing stages, there is a significant difference in the initial FAE$_{RMS\ 150\text{-}300}$ amplitude response that cannot *solely* be explained by differences in the loading response. Before this is discussed further it is worth noting that the AE and load measurements were not precisely time synchronised. Time zero has been set at the start of the increase of each *independently*, it is only highly probable, not guaranteed, that the start of the increase in FAE$_{RMS\ 150\text{-}300}$ and load were coincident. To emphasise the uncertain synchronisation the two plots have been deliberately vertically misaligned.

Comparing the loading response from each test stage, it can be seen that for both the load goes from zero to near steady-state within one second. Importantly the form of the increase at the start of each is very similar. At the top of the ramp the overshoots differ, with the load appearing to spike more for the first stage than the second, however it is possible that this difference is an artefact of the low sampling resolution. But ultimately the difference at the end of the loading ramp is not important, as it can be seen that there is a substantial difference between the $FAE_{RMS\ 150\text{-}300}$ response that lasts for longer than the duration of the loading spike and extends to the very start of the ramp.

For the first stage the $FAE_{RMS\ 150\text{-}300}$ jumps quickly, almost instantly, to a level above that of the final steady-state load, and continues to increase for a short period. In contrast the $FAE_{RMS\ 150\text{-}300}$ at the start of the second stage ramps up much more slowly. The biggest difference in the $FAE_{RMS\ 150\text{-}300}$ between test stages occurs at the very start of the loading ramp and by the time this is over, in just over two cycles of the specific contact topography the $FAE_{RMS\ 150\text{-}300}$ of each has converged towards a much closer, and steadier-state load. This indicates that the majority of running-in occurred within the first loading ramp which suggests that the initial running-in is largely independent of time. In other words, it occurs in tandem with the rate of forming the contact, and there is no delay between contact and the running-in.

An interesting question arises: is the observed AE greater during running-in because the running-in provides an additional transient source of AE: the material deformation / fracture; or it is because running-in reduces the subsequent amount of asperity interaction, and therefore the subsequent AE? In other words, is it the *mechanism* of surface modification that is being observed, or its subsequent *effects*? Of course it may be both. Due to the complications from the loading and temperature transients this is difficult to judge here but the results of the micro-pitting experiment suggest it is the *effects* of the surface modification that are observed not the mechanism. It must, however, be stated that these observations are based on the relatively simple analysis techniques used, and that more sophisticated signal processing techniques may yield information from the signals related to the mechanisms at work.

5.5 Transition between running-in and micro-pitting

Before the micro-pitting experiment is examined on its own, this section presents together, the AE measurements made during running-in, and those at the start of the subsequent micro-pitting experiment. This allows an examination of the measurement consistency from one experiment to the other, and also some fine details of the micro-pitting measurements that are obscured when the entire experiment is presented.

Figure 5-10 shows the $FAE_{RMS\ 150\text{-}300}$, contact voltage and slow disk temperature measurements made over the first three test stages, i.e. the two running-in stages and then the first stage from the micro-pitting experiment. Only the slow disk temperatures are shown as the fast disk temperature measurements became unreliable during the micro-pitting experiment[1] This is of little concern as the portions of testing where the fast disk temperature measurement was working show it to have been very similar to that of the slow disk, and there was no evidence of any divergent trend. The different stages are delineated by breaks in the curves; the first two combined: fast disks rotations 0 to 6000, are very short compared to the third: rotations 6000 to 106,000[2]. For clarity, the measurements made in the loading and unloading ramps of the first two test stages are not shown; the $FAE_{RMS\ 150\text{-}300}$ measurements start from the top of their initial spike, although for the first test stage this is outside of the plot boundary. The loading and unloading ramps for the third stage and onwards were not sampled.

Comparing the $FAE_{RMS\ 150\text{-}300}$ measurements between the end of one test stage and the start of the next is difficult because of the difference in temperature, and therefore film thickness. As has been discussed previously, because the contact's friction adds a new source of heat, there is always an initial but short-lived rise in the disk temperatures when they are initially loaded. Because the first two test stages were so short, the initial temperature rise lasted the entire length of each, and thus no clear steady-state temperature was reached by either. In contrast, because the third test stage was considerably longer the majority of this was spent at an easily identifiable steady-state temperature. It can be seen that the slow disk temperature oscillated subtly at steady-state, this was identified as a response to the heating element of the oil bath periodically switching on and off. Also of note, the starting temperature was noticeably lower for the third test stage than for the first two. This was probably because less time was allowed for the test enclosure to approach thermal equilibrium before the stage was started. The steady-state temperature of the third stage was also significantly lower than the final temperatures of the first two test stages. In

---

[1] During the long test stages the fast disk temperature would sporadically jump between two different steady-state levels ~ 10°C apart. This was totally unrealistic behaviour thought to be due to instrumentation issues.

[2] The nominal rotation counts are used for convenience, the actual rotation counts differ marginally but not by a significant amount, and not in any way that would affect the results.

part this was certainly due to the lower starting temperature, however it was also likely due to a gradual reduction in frictional heating which will be discussed in more detail in the next section.

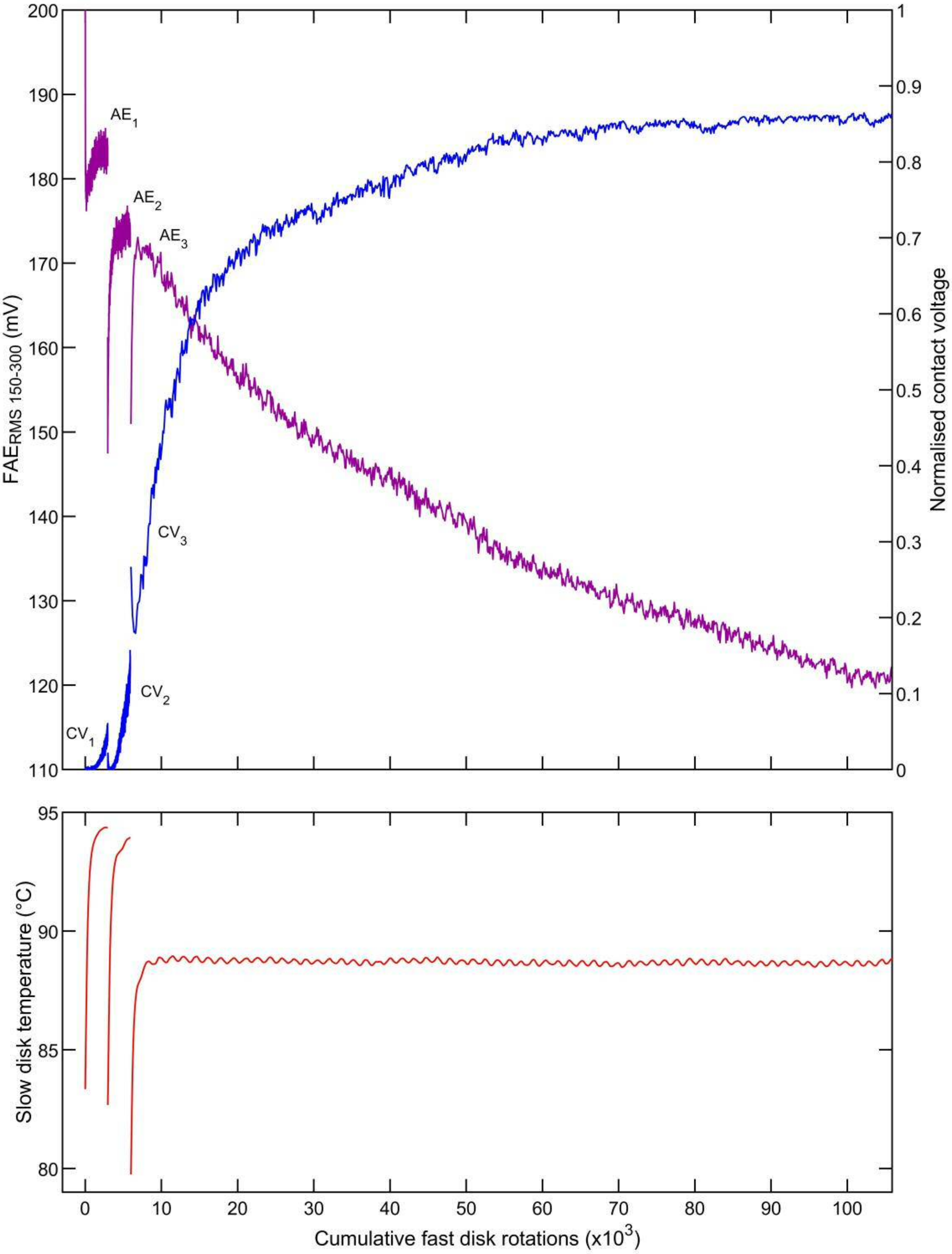


Figure 5-10 The FAE$_{RMS\ 150\text{-}300}$ (AE$_i$), contact voltage (CV$_i$) and slow disk temperature for the first three test stages. The initial ramps in FAE$_{RMS\ 150\text{-}300}$ and CV at the start of the first two test stages are not shown.

Now comparing the FAE$_{RMS\ 150\text{-}300}$ from the second to the third test stage, remembering that the running-in had been completed before both of these. It can be seen that there is reasonable consistency between the FAE$_{RMS\ 150\text{-}300}$ response and amplitude from one to the other. In the periods of increasing temperature both show an increase in $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ due to the thinning oil film and consequently greater asperity contact. After the rapid initial increase, the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ amplitude of the third stage peaks at a similar amplitude to that found at the end of the second test-stage. There is a small difference but this is to be expected given the temperature difference between these two regions.

The long steady-temperature period of the third stage is where the change in $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ becomes interesting. It can be seen that over this there is a gradual decrease in $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ so that by the end of the stage it has decreased substantially, by about 30 %, from its initial peak amplitude. This indicates a gradual decrease in asperity contact which, because the nominal film thickness was steady, can only have been caused by wear. For this test stage the contact voltage response is congruous with that of the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$, i.e. where the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ increases the contact voltage decreases and vice versa. It is interesting to note the difference in the form of the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ curve and the contact voltage curve. Both are non-linear but the contact voltage has a much more dramatic change in gradient than the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$. At the start of the steady-temperature period the rate of change of contact voltage, i.e. its gradient, was noticeably greater than that of the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$, but by the end it was noticeably less. This indicates a difference in relationship with the amount of asperity contact, which is not surprising given that each measures a fundamentally different property, i.e. one electrical and one mechanical. It should be noted that it cannot be determined from this comparison alone which has the more directly proportional, or linear, relationship with the amount of asperity contact because the *intra* test stage wear rate is uncertain.

As a final observation from Figure 5-10 it can be seen that the subtle temperature oscillations in the steady state-period affected the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$, which also oscillates with the same period. There is some suggestion of an oscillation in the contact voltage measurement but it is certainly not as clear as that for the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$. Closer inspection of the later stages of the micro-pitting experiment (not presented here) revealed the oscillations to be present in both the contact voltage and $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$. It is highly probable that they were caused by subtle temperature induced fluctuations in the amount of asperity contact[1]. This gives an indication of just how sensitive the $\mathrm{FAE}_{\mathrm{RMS}\,150\text{-}300}$ and contact voltage are to changes in the amount of asperity contact. For the steady-temperature period of the third stage, the peak to valley amplitude of the oscillation is less than half a degree Centigrade and occurs within a very limited range: between 88.5 and 89.0 °C. Using the average from before and after the third stage the specific film thickness prediction only varies between 0.358 and 0.354 across this temperature range.

---

[1] The AE measurement hardware was separate from that of the rest of the rig so it is extremely doubtful that the AE oscillations could have been due to interference from other signals rather than genuine observations.

5.6 Micro-pitting experiment

In this section the real-time measurements from the entire micro-pitting experiment are considered. Figure 5-11 presents the $\text{FAE}_{\text{RMS }150\text{-}300}$, contact voltage and slow disk temperature measurements made from the third test stage onwards to the end of the experiment. They are plotted against the cumulative fast disk rotations and the tick marks on this axis delineate the different test stages.

It can be seen that both the $\text{FAE}_{\text{RMS }150\text{-}300}$ and contact voltage underwent a broadly tapering change over the course of the entire experiment, so that, although rapidly changing at the start of the experiment, by the end they reached a comparatively steady-state. The overarching changes in both: a decrease in $\text{FAE}_{\text{RMS }150\text{-}300}$, and an increase in contact voltage, indicate a decrease in asperity contact which can only have been caused by the surface wear. The near maximum contact voltage at the end of the experiment suggests that by this point the lubrication was almost full film. It is thus probable that had the test been continued, the surfaces would have eventually failed by some mechanism other than micro-pitting, if at all. Importantly, both the AE and contact voltage measurements support the evidence from the surface measurements that the wear, including micro-pitting, was self-limiting.

The details of the results in Figure 5-11 are now discussed. As expected there was an initial temperature rise at the start of all test stages which effected both the $\text{FAE}_{\text{RMS }150\text{-}300}$ and contact voltage. It *appears* that the starting temperature was significantly higher for the last five, longer, test stages but this was not the case. It is an illusion caused by the change in sample rate from every 100th to every 500th fast disk rotation, for these stages. To explain: sampling was always started before loading, and the load was applied immediately following a sample. Thus for the last five stages there was a longer interval from loading to the first in-test sample, and thus less of the initial temperature rise is sampled.

Interestingly the temperature behaviour of the third test stage is slightly different to the others. For these, after the initial extremely rapid (vertical) rise, the disk temperatures continue to increase at a much slower and tapering rate, giving a radiused corner to the temperature profiles. This was due to the slow rise in the temperature of the test enclosure, as the entire rig gradually approached thermal equilibrium. However for the third test stage, there was no slow temperature increase. After the vertical rise the temperature immediately transitions to near steady-state and then starts to marginally decrease. It is speculated that this is due to a significant decrease in the frictional heating over this stage due to the relatively large reduction in asperity contact as indicated by the change in the $\text{FAE}_{\text{RMS }150\text{-}300}$ and contact voltage.

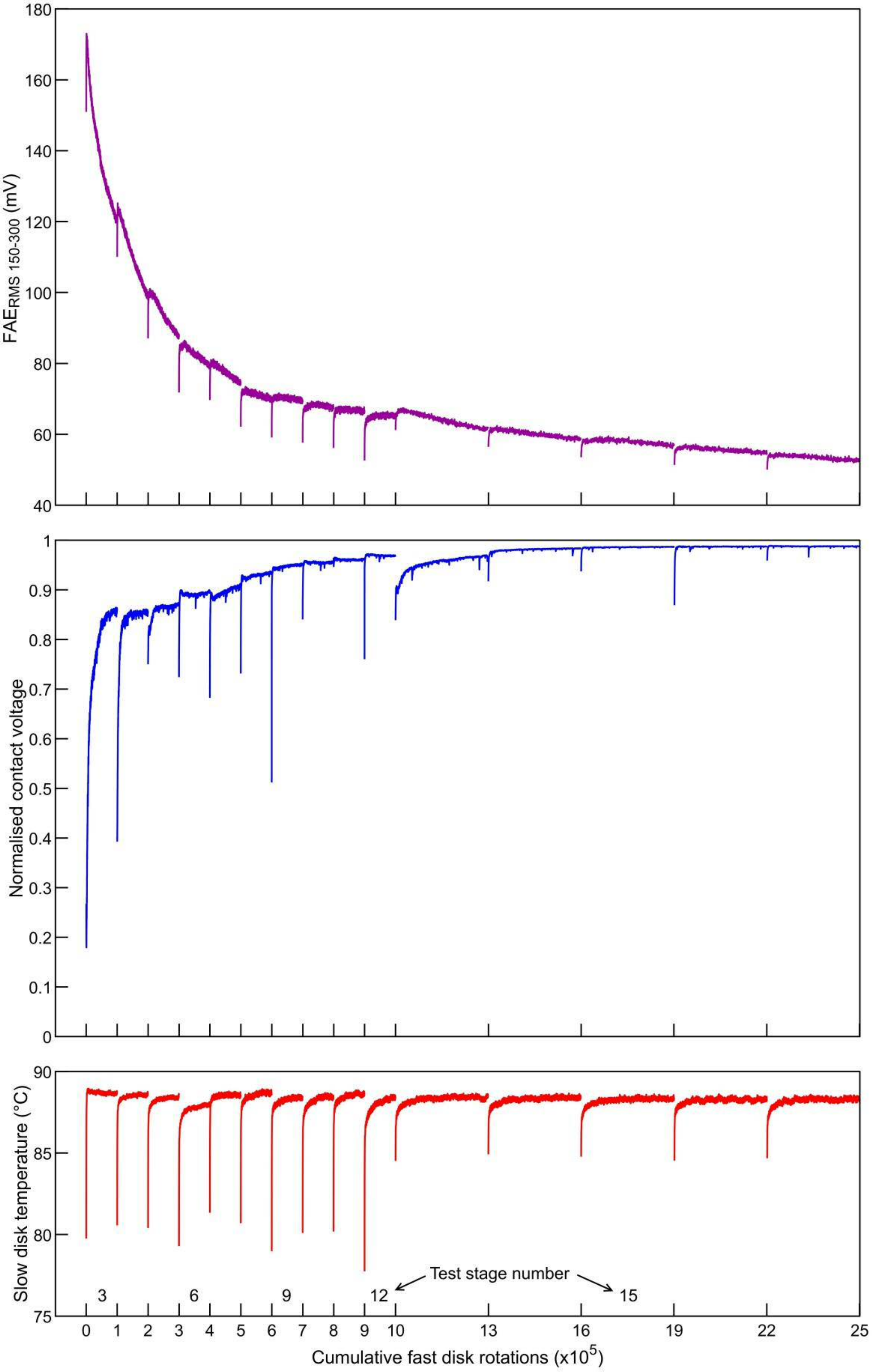


Figure 5-11 The FAE$_{RMS\ 150\text{-}300}$, contact voltage, and slow disk temperature for test stage 3 and all subsequent. The tick marks on the abscissas delineate between the different test stages, some of which are explicitly labelled on the temperature plot.

Although the changes in the $FAE_{RMS\ 150\text{-}300}$ and contact voltage are broadly harmonious in the steady-temperature periods, they are not during the initial rapid temperature rises. During all these the $FAE_{RMS\ 150\text{-}300}$ behaviour is as might be expected: it increases, indicating increasing asperity contact. Additionally, the magnitude of the increases are reasonably consistent with those of the temperature. In contrast for the contact voltage there is an initial increase which suggests *decreasing* asperity contact despite a thinning oil film. The magnitude of the increases are also much less consistent with those of the temperature, for example the initial increase in contact voltage at the start of test stage 9, (at 6 x 10^5 fast disk rotations), is triple that of the next test stage despite them having very similar temperature profiles. A similar 'increasing contact voltage despite decreasing thinning oil film' phenomenon was observed during the variable film thickness experiment (Chapter 4) and it seems likely that they are related. These results add to the evidence that it is start-up effect, and as discussed in Section 4.6 the repeated growth of a non-conductive tribofilm could explain the behaviour.

If the transients are ignored the contact voltage is mostly consistent from one stage to the next however there is one obvious exception to this: the 13th test stage, (the first long one). At the start of this the contact voltage jumps to a lower value and only recovers slowly and partially. There is no evidence that this was due to a thermal effect as the steady state temperatures of this and the preceding stage were similar. The instantaneous contact voltage measurements help to understand the step change in mean contact voltage and they are now presented.

Figure 5-12 shows the change in instantaneous contact voltage over the duration of the micro-pitting experiment. The instantaneous contact voltage is shown for one whole rotation of the fast disk, i.e. one fifth of the specific contact topography[1]. The contact voltage, (tone), scale has been truncated between 0.7 and 1 to emphasise the changes in this range, thus pure black may represent any value from 0 to 0.7. It can be seen that the contact voltage fluctuated rapidly with circumferential position which results in horizontal banding. This has been discussed before, but what is important here is the *specific pattern* of the bands. It can be seen that the pattern is consistent from test stages 5 to 12, (fast disk rotations: 2 x 10^5 to 10 x 10^5), which confirms a single, unique, specific contact topography during this period. The bands do fade over time which indicates the wearing of the surface, but fundamentally the pattern remains the same. However at the start of test stage 13 there is an instantaneous change to a *partially* different pattern which is then consistent until the end of the experiment. The pattern is described as only partially different because whilst there were new asperity contacts (black bands) in places where there were none previously, there does appear to be some continuation of the previous pattern. The change in pattern is indicative of a change in the specific contact topography, which might explain the instant drop and then slow recovery in mean contact voltage over test stage 13 as new asperity

---

[1] The instantaneous contact voltage was sampled across the entire specific contact topography but it is too cumbersome to present the entirety of this here

contacts induced more wear. There also appears to be similar but more subtle change in the pattern from test stage 3 to 4, and 4 to 5, which again seems to indicate a change in the specific contact topography. Given the evidence for this, the question becomes how did it happen? The only way is for the assembly orientation of the disks to have changed, but this was not done deliberately at any point in the experiment. The flexible couplings that link shaft lengths are secured by clamps, rather than by keys or splines, and so it is mechanically feasible that they might have slipped due to the initial loading spike, *if* they were not sufficiently tightened. However, this has never been observed before and there was no notable looseness in the clamps when the couplings were disassembled at the end of the experiment. The cause of the change in the amount, and pattern, of instantaneous asperity contacts remains unknown, but ultimately it is not of great consequence as it did not affect the overarching decrease in asperity contact. As a final point, it is interesting that the disturbance to the nature of the asperity contact did not affect the $FAE_{RMS\ 150\text{-}300}$ to the same extent as it did the contact voltage. It can be seen in Figure 5-11 that there is a small step change in the AE from test stage 12 to 13 but it is much more subtle than that of the contact voltage. This highlights a difference in sensitivity of the two measurement types which is not surprising as they are measuring fundamentally different characteristics of asperity contacts (one electrical and one mechanical).

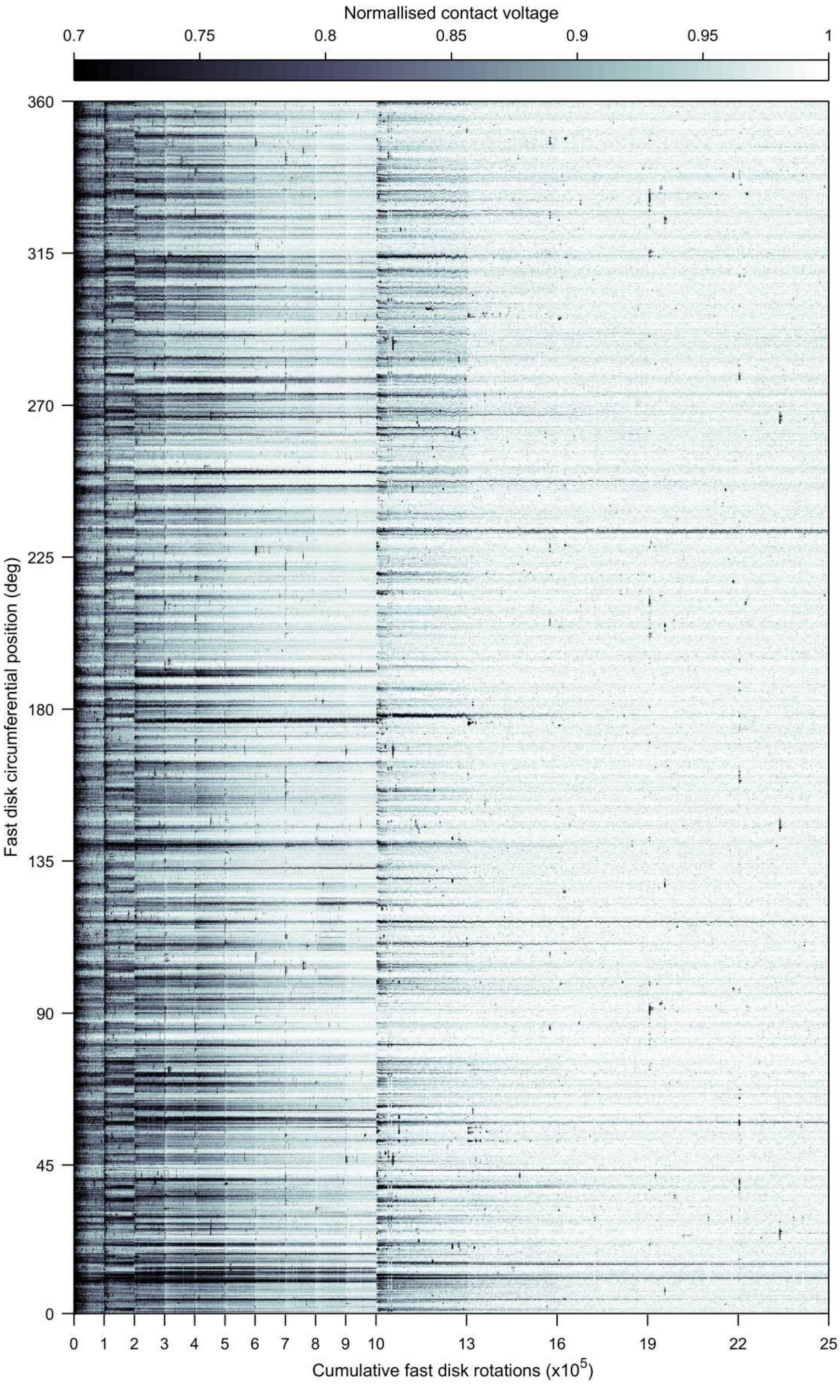


Figure 5-12 The instantaneous contact voltage vs. cumulative fast disk rotations for the entire micro-pitting experiment. The tick marks on the rotations axis delineate the different test stages. (A 30 kHz low pass filter was used for the instantaneous contact voltage).

5.6.1 The relationship between AE and

This section examines the relationship between the $FAE_{RMS\ 150\text{-}300}$ and the (RMS roughness) during the micro-pitting experiment, Figure 5-13 shows, on the same plot, both of these vs. the cumulative fast disk rotations. The values were calculated using the five central profiles from both disks and a fit of these was also shown. The fit is a rational function of the form:

——— (15)

The ordinates have been scaled so that the first and last values of the $FAE_{RMS\ 150\text{-}300}$ and the fit are vertically aligned on the plot. Only the steady-temperature $FAE_{RMS\ 150\text{-}300}$ measurements are of interest here and so the first two minutes of every test stage have been cropped to exclude the periods of rapid temperature rise.

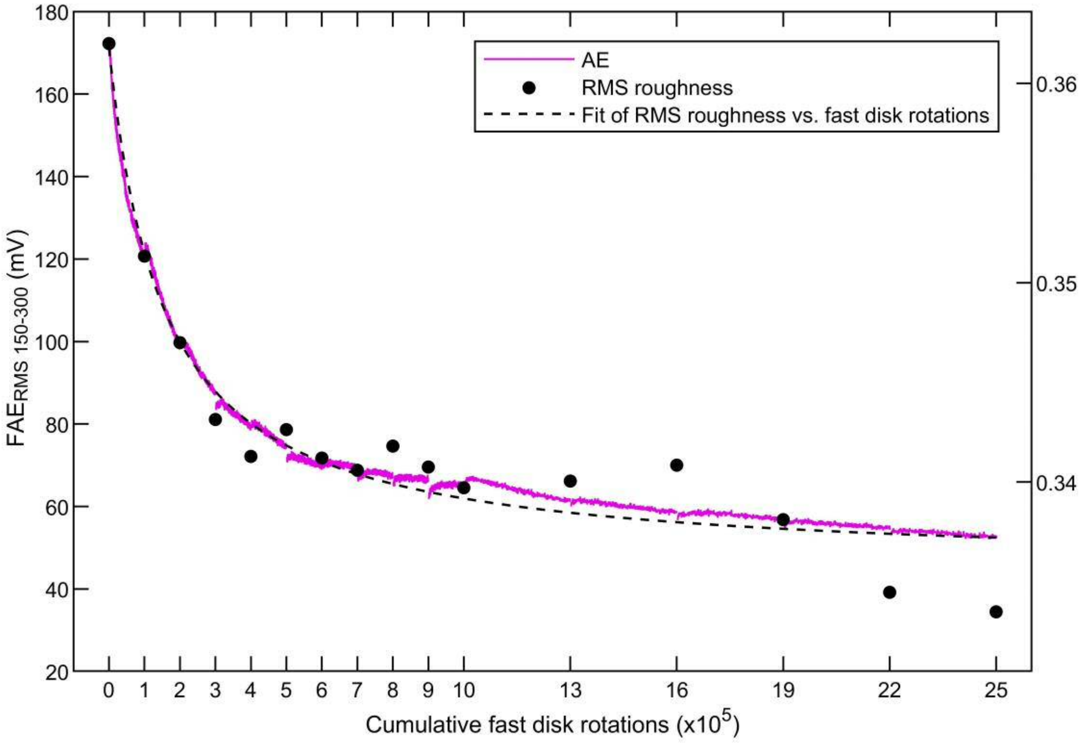


Figure 5-13 Comparison of the change in $FAE_{RMS\ 150\text{-}300}$ and over the entire micro-pitting test. The was calculated using the 5 central profiles from both disks. The $FAE_{RMS\ 150\text{-}300}$ measurements taken during the initial two minutes of each test stage have been excluded.

Figure 5-13 shows that broadly the $FAE_{RMS\ 150\text{-}300}$ and had a good correlation over the experiment, i.e. they were directly proportional. However, it can be seen that there is considerable inconsistency, i.e. spread, in the samples and thus there is some uncertainty in just how well the $FAE_{RMS\ 150\text{-}300}$ represents the change in roughness in the later stages of the experiment. It is noteworthy that the uncertainly results from inconsistency in the measurements and not the

$\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$. In other words, in this experiment, the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ seems to represent the change in roughness better, or with less uncertainty, than the surface measurements themselves. It is likely that this is due to a difference in sample size. The AE measurements averaged the effects of changing roughness over the entire contact width and specific contact topography. In contrast the surface measurements only averaged a very limited portion of the contact surface and were also susceptible to noise caused by axial alignment error. Had it been practical to make 3D surface measurements of the entire contact paths from both disks, either directly or by using replica methods, then it is expected the would have been much more consistent. The results of Figure 5-13 suggest that AE measurements could feasibly be used to investigate the real-time changes in the of a surface, but this would only be the case in situations where the change affected the amount of asperity contact, i.e. only within the mixed lubrication regime. It would also require reference points, e.g. the starting and ending roughness, as was used here.

5.6.2 Spectral analysis of the AE

The AE results presented in this chapter have considered a single parameter, the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$, which represents the AE observed in the 150 to 300 kHz frequency range. In this section the wider spectral content is briefly considered and compared with that from AE measurements on the bearing housing. Figure 5-14 shows the 10 kHz binned spectral distribution vs. cumulative fast disk rotations as measured by a) the rotating sensor on the fast disk, and b) the sensor on the housing of fast shaft bearing. The amplitude (colour) scales of the two plots are not identical but are from zero to the maximum measured by each sensor individually, this was 1.6 mV for the rotating sensor, and 0.05 mV for the bearing housing sensor.

For the rotating sensor, the 150 – 300 kHz band shows the largest scale of change which supports the use of the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$ parameter. Compared to the spectral content of the AE from the variable film thickness experiment (Chapter 4), the spectral content here is considerably less complex and there is virtually no contradiction in the response of different frequency bands. This suggests that the complexities of Chapter 4 were due to the effects of large temperature changes and the associated large reductions in film thickness, which were absent for this experiment. It can be seen that aside from the 150 – 300 kHz band, there is some significant activity in the 20 – 150 kHz band. The $\mathrm{FAE}_{\mathrm{RMS}\ 20\text{-}150}$ was examined and found to have a similar but much less precise form to that of the $\mathrm{FAE}_{\mathrm{RMS}\ 150\text{-}300}$. The lower precision may have been caused by lower frequency noise and / or reduced transducer sensitivity in this region.

The AE measurements made on the bearing housing show that the bearing contacts are very unlikely to have affected the measurements made at the test contact. The AE measured at the bearings remained relatively constant throughout the micro-pitting experiment, and in the 150 – 300 kHz band were comparatively negligible at all points in the test procedure. The relative absence of AE from the bearings is expected, as, primarily by virtue of their smoother surfaces, they were likely to have been operating in full (or near-full) film conditions at all points during the testing[1]. There is some low frequency activity in the bearing housing measurements, and although it cannot be certain that this was due to the bearing contacts, it seems likely as there was no other obvious source. It is reasonably time-invariant, which again suggests it is from the bearing contacts operating at constant speed and load.

It is worth noting that the bearing housing sensor did not show any obvious evidence of being able to detect AE from the test contact, i.e. there was no convincing asymptotic decrease in the AE amplitude at any frequency. This is not surprising given the subtlety of the changes in lubrication conditions at the test contact, and the rather torturous transmission path from this to the housing sensor, (through multiple lubricated contacts and metallic interfaces). It highlights the problem of attenuation when attempting to use a housing sensor to monitor remote contacts.

As discussed in Chapter 2 in some circumstances researchers have been able to attribute specific frequency components of AE signals to specific wear mechanisms. This is not the case here. Because the rate of micro pit formation was similar to the rate of reduction in roughness there is no *obvious* way to distinguish between AE from the pitting *process* and AE from general asperity contact. Both must have decreased similarly over the experiment. However, for several reasons it is highly probable that the AE measured here was due to general asperity contact rather than specifically crack growth and fracture associated with micro-pitting or other wear mechanisms. Firstly, the temperature related changes in AE at the start of each stage are very consistent with the expected changes in the amount of general asperity contact, rather than rate of wear. And additionally the results of Chapter 4 showed the AE to be very dependent on the amount of asperity contact despite minimal wear, and there is no reason to think this should have been different here. It should however be noted that the analysis of the AE samples has not been exhaustive and has used broad time averages to describe the AE behaviour. It is not impossible that examination at finer temporal resolutions, (perhaps using Wavelet Transforms) might reveal AE attributes related solely to wear mechanisms rather than general asperity contact. There was insufficient time to investigate this as part of this thesis.

---

[1] The calculation of the lubrication conditions for the bearing contacts should be possible, but is more complex than for the test contact and is outside the scope of this thesis.

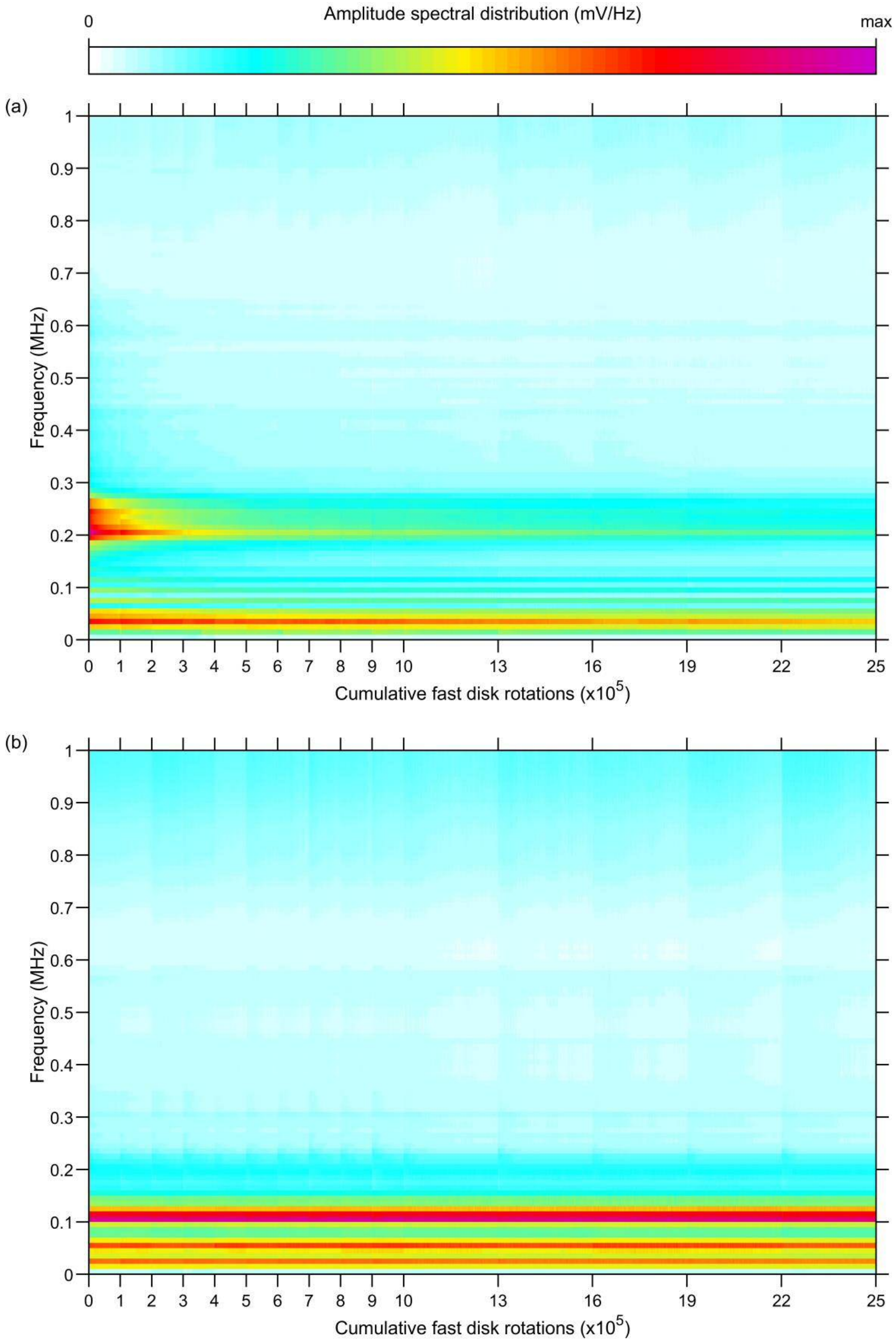


Figure 5-14 The 10 kHz binned AE spectral distribution vs. the cumulative fast disk rotations for measurements from (a) the rotating sensor and (b) the sensor attached to the bearing housing. The amplitude ranges are between 0 and the maximum measured by each sensor individually.

5.7 Review

This section presents a review of the entire experiment, it is split into two parts: the first regarding the running-in stages, and the second the micro-pitting stages.

Running-in

The behaviour of the AE during running-in, in conjunction with the surface measurements, suggests that the rate of the initial running-in is largely governed by the rate at which a new, or more aggressive, contact is made. In other words, as soon as asperities come into contact more aggressively than previously they will immediately deform, fracture or abrade to become more conformal. It has been shown before by Weeks [57] and Clarke [61], using before-and-after surface roughness measurements, that the initial running-in on the twin disk rig can be quick: within tens of seconds, but it has taken the real-time AE measurements to show just how quick (much more so than previously proven). It is also clear that AE measurements are superior to the contact voltage for the study of very aggressive lubrication conditions, such as those found in running-in. The contact voltage had an obvious measurement limit, when the contact resistance became negligible, however no corresponding measurement limit occurred for the AE.

The AE measurements during this experiment have provided some new insight into the running-in process, however the procedure could be improved to reveal even more details. Its biggest limitation was in using time-averaging for the load samples. This hindered the comparison between the AE and load during the short loading ramp. Rather than averaging the load over every fast disk rotation it would have been more beneficial to record the continuous load, and in synchronisation with the continuous AE[1]. However, it was not anticipated that time-scales shorter than a single rotation would need to be examined. Regarding the scope of the procedure: as only one running-in was measured its repeatability is as yet unknown. It would be desirable to repeat the running-in procedure on other similar disk pairs to assess this. If the AE proved repeatable, then different running-in parameters could then be tested to determine if the AE correlates with the scale of the running-in, i.e. with the volume of material deformed / removed, or the change in . This need not involve numerous disk pairs as multiple running-in stages could be induced on one disk pair using incremental increases in load or temperature, or incremental decreases in speed.

---

[1] The AE acquisition computer has a number of channels which can accept general signals. If the load were to be sampled by the AE computer rather than the rig computer then the certainty of the synchronisation would be much higher.

Micro-pitting

As was expected, some micro-pitting occurred during the procedure and this was easily verified at the end by use of 3D surface measurements. However, it was difficult to use the in-situ measurements to identify *when* the micro-pitting occurred. This was largely because 2D profilometry is not ideally suited for identification and characterisation of micro-pits. It is predicted that implementing an in-situ 3D measurement scheme would be extremely beneficial for the study of micro-pitting, as it would allow easy pit identification and would enable post-measurement axial alignment. Staying on the topic of surface measurements, there was some noise, and consequential uncertainty, in the measurements which became apparent, and significant, in the later stages of the experiment, i.e. when the rate of change became very small. It is suspected this was due to the combination of increasing axial variation due to micro-pitting, and the axial alignment error. It is probable that using a larger sample size, i.e. longer and or more profiles, would reduce the noise. Despite the limits of small sample size, and 2D only measurements it was shown that micro-pitting developed at a similar rate to the decrease in . However because the wear was not exclusively limited to micro-pitting it cannot be assumed that there was a direct correlation between them. Analysis of the surface measurements was not exhaustive and it is possible that a correlation between and micro-pitting might be established with more work, however this was not necessary for the aims of this thesis.

Importantly, all three measurements: , contact voltage and AE show that the wear, which included micro-pitting, was self-limiting, i.e. the rate of wear decreased continually. This implies that the micro-pitting had no adverse consequences on the lubrication conditions. And since the lubrication conditions actually improved over the duration of the micro-pitting experiment it should be regarded as a secondary, and much slower, phase of running-in rather than a damaging wear regime. It should be emphasised that the limited and benign form of micro-pitting seen here differs from that found on gears where it can become much more extensive and can modify the surface so that it does adversely affect the operating conditions [126]. The question is, why was a damaging micro-pitting regime not seen here? There are many possible explanations: it may be because the operating conditions were too benign, e.g. not enough sliding or load; or due to the use of disk rather gear kinematics, or it may simply have been that the test was not continued for a sufficient length of time, and that had it been, a new more damaging regime of micro-pitting would have eventually occurred. Also absent in this test were the stress-concentration effects associated with tip-to-root contact in gear teeth, and where micro-pitting commonly occurs [127].

The specifics of the micro-pitting regime were not overly important for the purpose of this thesis, what is important is that the micro-pitting experiment has shown that the AE is affected by *subtle* changes in the lubrication conditions caused by wear. Furthermore, out of the three measurements: , contact voltage and AE, the AE indicated the rate of wear with the most consistency and precision. During the start-up of each stage there were incongruities between the AE and contact voltage measurements and these need to be investigated further. It might help to repeat the micro-pitting experiment but using a single long test stage to mitigate the complications caused by repeat start-ups. This would allow for a simplified comparison between the AE and the contact voltage.

6 Conclusions

6.1 Important findings

- **Using a two disk contact can be more effective than using a gear contact to investigate the AE from asperity contacts in mixed EHL.**
  Many researchers have attributed the AE measured during gear tooth contacts to asperity contacts. But due to the many variables associated with the tooth contact cycle it has been difficult for gear tests to irrefutably demonstrate that asperity contacts are the true source of the AE. Using the twin disk rig it was easy to vary different operating parameters in isolation to provide a more controllable, and measurable variation in the lubrication conditions than is possible with real gear contacts. This, in conjunction with the ease of making surface roughness measurements on the disks, has allowed a much more robust investigation of the AE from asperity contacts in mixed EHL. So far, simulated gear contacts have been underutilised in AE investigations; it is hoped the experiments presented in this thesis may change this.

- **An oil coupled and stationary AE sensor can measure the signal from a rotating component.**
  This was demonstrated by comparing the measurements made by this sensor with those made by a more traditionally interfaced: a clamped, sensor. The quality of the measurements from the oil coupled sensor were not as good as those from the clamped sensor and this was likely due to inconsistent coupling. Improvements to the design of the oil coupling will hopefully address this.

- **In the mixed lubrication regime, the AE observed at lower frequencies (20 - 600 kHz) was predominantly caused by asperity contacts.**
  This was demonstrated by comparison of tests on smooth and rough disks: For the rough disk tests, the amount of asperity contact increased with decreasing film thickness, and the AE activity in the 20 – 600 kHz range did likewise. For the smooth disk tests the amount of asperity contact did not increase with decreasing film thickness, and neither did AE activity in the 20 – 600 kHz range.

- **It was not difficult to identify an AE parameter that had a useful relationship with the amount of asperity contact.**
  No conditioning of the AE signal using esoteric techniques was required to return a useful parameter. All that was required was a spectral analysis, using FFTs, and then a standard digital filter. The ease of analysis is attributed to asperity contacts being the dominant source of AE in the test contact, and the good signal to noise ratio achieved by positioning the sensor close to the contact. This highlights the advantages of placing a sensor at a contact rather than on a housing if at all possible.

- **An AE parameter, the $FAE_{RMS\ 150\text{-}300}$ (the RMS of AE in the 150 – 300 kHz pass band), varied in accordance with the amount of asperity contact, and did so precisely and consistently.**

  In experiments where the thickness of the lubricant film was reduced, thereby instigating a change from full-film to mixed lubrication, the $FAE_{RMS\ 150\text{-}300}$ responded to the changes in the amount of asperity contact (as verified by contact voltage measurements). In the full-film regime the $FAE_{RMS\ 150\text{-}300}$ was insensitive to changes in film thickness and in the mixed regime it was extremely sensitive.

- **An accurate mathematical model for the relationship between the $FAE_{RMS\ 150\text{-}300}$ and specific film thickness at constant speed was developed.**

  A simple, two termed, mathematical model describes the relationship between the $FAE_{RMS\ 150\text{-}300}$ and the specific film thickness at a constant speed. One term, a constant, represents a background level of AE at full film conditions, and the other, an exponent, represents the increasing amount of AE from asperity contacts in the mixed regime.

- **The speed has a direct effect on the amplitude of AE from asperity contacts and is a complicating factor in the relationship between the AE and specific film thickness.**

  As has been identified by other researchers, the speed has a direct effect on the AE amplitude independent of its effect on the film thickness and consequent effect on the amount of asperity contact. At constant amounts of asperity contact, whether considerable or negligible, the amplitude of the $FAE_{RMS\ 150\text{-}300}$ always increased in conjunction with the speed.

- **The direct effect of speed was systematic and a mathematical model for the relationship between $FAE_{RMS\ 150\text{-}300}$, specific film thickness and speed was developed.**

  The accuracy of this model was less than that of the constant speed model, but it was sufficient so that $FAE_{RMS\ 150\text{-}300}$ could be reasonably estimated for intermediary speeds not tested. This model goes someway to being sufficiently general for use in a practical lubrication monitoring scenario. With the inclusion of load as a variable it would cover the most common operating variables in a given system.

- **There were no AE parameters that varied in accordance with the coefficient of friction.**

  This is attributed to the fact that the friction in the contact is not solely caused by asperity contacts (boundary lubrication) but also by the viscosity, i.e. resistance to shearing, of the fluid film which changes dramatically with temperature and with film thickness. So if the AE increased directly in proportion to the friction generated by asperity contacts (which has yet to be proved for mixed EHL) it would still not necessarily increase in proportion to the friction of the contact as a whole.

- **The AE measurements have proved their worth as a tribological measurement tool.**
  Once an AE parameter that measured the amount of asperity contact had been identified (the FAE$_{\mathrm{RMS}\ 150\text{-}300}$) it was used to investigate running-in and micro-pitting and was found to provide useful insights into these processes.

- **AE measurements can investigate the aggressive lubrication conditions of running-in more effectively than contact voltage measurements.**
  The contact voltage has often been used to investigate the geometric running-in of asperities at the start of new contacts. However the initially very aggressive lubrication conditions can easily result in an amount of asperity contact beyond that which the contact voltage can measure. In the running-in test presented in this thesis, it is demonstrated that the AE measurements were able to detect changes in the amount of asperity contact beyond that which the contact voltage could measure.

- **The geometric running-in of asperities is not a gradual deformation or material removal process.**
  The AE measurements have revealed that the geometric running-in of asperities in a new contact occurs at the rate at which the contact is made, i.e. the asperities run-in at the rate the load is increased. There is essentially no waiting period for the asperities to adapt to the new contact conditions.

- **In simple mixed lubrication experiments AE measurements could potentially be used to indicate the rate of wear by any associated change in surface roughness.**
  In the micro-pitting test presented in this thesis, where all operating parameters were kept constant, and the surfaces were allowed to wear naturally, it was found that the AE measurements were sensitive to very subtle changes in the amount of asperity contact caused by a gradual reduction in surface roughness. There was a good correlation between the measurements and the FAE$_{\mathrm{RMS}\ 150\text{-}300.}$ The AE measurements effectively acted as a real time measurement of changes in surface roughness and likely could do so in other mixed lubrication experiments used to investigate wear phenomenon.

- **The AE from asperity contacts in mixed lubrication likely obscures any AE from the cracking and fracture associated with micro-pitting.**
  Although many experiments on small-scale tribometer contacts have identified AE caused by wear mechanisms (e.g. deformation and fracture) there was no obvious evidence of AE caused by the formation of micro-pits. It is likely that, if there was any, the AE from asperity contacts obscured it. The analysis of the AE signals was not exhaustive and evidence of the micro-pitting process might still be hidden in the minutiae of the signals. But even if so, its subtlety compared the AE from asperity contacts, may preclude it from being useful from a damage detection perspective.

6.2 Novelty

The section provides a brief discussion on the novelty of this research. Fundamentally it is based upon the marriage of acoustic emission measurements with precise knowledge of the lubrication conditions, and for a semi-applied mixed EHL contact. Yes, there have been previous studies which have linked acoustic emission with precise lubrication conditions, but these have tended to be for smaller and more idealised contacts. And yes, there have been previous studies which have investigated the acoustic emission from real components and machines, but these have relied on significant assumptions regarding the lubrication conditions. In short, this work represents a novel, and useful amalgam of fundamental and applied research into the AE from mixed EHL contacts.

Some specifics: The conclusions that direct asperity contacts cause AE, and that speed has a direct effect on AE, are not revelatory. Previous researchers have also come to these conclusions, however this means by which they are demonstrated are more explicit in this work and the uncertainty less. Examination of the relationship between the specific film thickness and AE is also not unique however the level of detail, and rigour, at which it was explored in this thesis is. For example, no instances in literature were found where the amount of asperity contact in applied contacts was actually verified using an alternative measurement system, as was the case here using the contact voltage. Investigating changes to the lubrication conditions by varying the uniform film thickness, as was the case for the first experiment, is relatively common. However investigating changes caused by natural and progressive wear, as was the case for the second experiment, is more unusual. In the mixed regime, the existence of a good correlation between changes in the surface roughness and the acoustic emission is a new finding. And as a final example, the attempted use of an exclusively oil coupled AE sensor is novel. Although the AE from a gear has been previously measured using a static sensor, in that research it was not clear if the coupling was designed to involve any direct contact between the gear and the sensor or otherwise.

6.3 Implications

This work has potential benefits for both laboratory research and also condition monitoring of rotating components. The benefits for laboratory research are more certain and easily realised. The mechanisms of wear in realistic mixed EHL contacts are far from fully understood and doubtless these types of contact will continue to be investigated for many years to come. This research has shown that AE measurements can provide a useful indication of changes in the amount of asperity contact, and by inference, the scale and rate of wear. As a measurement tool the AE offers some significant advantages over the contact voltage, namely greater measurement range and no requirement for electrical isolation of components. However it does also have some significant disadvantages. It requires calibration, is affected by speed, and is unlikely to be able to indicate the *instantaneous* amount of asperity contact. So AE monitoring should be considered as a new weapon in the armoury of the researcher, not a like for like replacement of the contact voltage.

Research explicitly for condition monitoring tends to focus on narrow and specific applications, often comparing the differences in AE from damaged and healthy components, or adverse and normal operating conditions. In these situations the research outcomes are limited in scope. The advantage of this, more fundamental, research is its potentially greater generality. It is possible to envisage a situation in which the relationship between AE and all contact variables has been sufficiently characterised, so that it may be used to indicate damaging lubrication conditions for a wide range of real contacts. (An analogy would be, how the CDDT equation is used to predict the minimum film thickness for many different specifications of EHL contact). However there are some substantial barriers to recognising this vision.

Firstly, the number of influencing parameters. For lubrication monitoring it would perhaps be ideal if the AE was solely influenced by the amount of asperity contact, but this is not the case. As has been shown here, and by other researchers, the speed has an independent effect and the same may be true of other parameters, e.g. load, contact geometry, contact material, etc. It is likely that many parameters, like the temperature, do not independently affect the AE but this still needs to be tested. A twin disk setup like the one used here would be suitable for much of this groundwork.

Secondly, because the *observed* AE is affected both by the transmission path and the sensor characteristics, the calibration required to adapt a general relationship between AE and lubrication conditions, to a specific application, is likely to be extensive. And finally, for gear applications, the impracticalities of positioning a sensor close to the contact need to be addressed. The further development of an oil-coupled sensor used in this research may do this, or it may not. It remains to be seen if the quality of an oil-coupling can be made to match that a directly clamped sensor, or if the design requirements for doing so are any less onerous than using a slip ring to connect a clamped sensor. There are possible alternatives to both the sensor placement methods used here, such as wireless transmission, or transmission through rolling element bearings, but so far there is no evidence to show that these can produce the quality of measurements shown here.

6.4 Further work

This section gives some recommendations for further work. Firstly the analysis of the measurements made for this thesis has not been fully exhausted. For example, with the exception, of the running-in experiment, little attention has been given to the nature of rapid changes in the AE, i.e. those faster than the rotational frequency of the fast disk. It would be worth examining some of the fine detail of the individual samples, perhaps using wavelets, to confirm or refute the existence of any parameters which could relate to specific instances of wear, asperity contacts, or circumferential variation in roughness.

Secondly, although perhaps not an exciting prospect, the repeatability, or uniqueness, of the results needs to be established. For example, does repeating the experiments on different disk pairs, i.e. different *specific* contact topography, effect the results, and does changing the type of AE sensor?

Looking further into the future it would be wise to prioritise testing the contact parameters which are often variables in real contacts, such as load and the slide roll ratio. Both of these can be tested without further modification to the twin disk rig however the load is somewhat complex case. Changing the load, also changes the width of the contact, and so careful attention must be given to axial variation in roughness. (In theory it might be possible to isolate the effects of area change from those of load by using different load and geometry combinations to give the same area but this could not easily be implemented on the twin disk rig and it is difficult to see the value of this). Using the twin disk rig it should be possible, to investigate the effects of different slide roll ratios at constant, or at least similar, temperatures and speeds, something that is not possible in gears. This would help answer questions concerning whether AE from contacting asperities is dependent on the amount of *relative* motion between them, or if it is due simply to the amount of deformation.

Ultimately, and only once the fundamental relationship between AE and lubrication conditions has been characterised further using the twin disk rig, it would be logical to take the same suite of measurement techniques used for this research and apply them to contact between real gears to test for any differences.

References

[1] Jones D.E., Schott G.A., editors, 1896, Miscellaneous Papers by H. Hertz. London: Macmillian.

[2] Johnson K.L., 1987, Contact Mechanics. Cambridge University Press.

[3] Williams J.A., 1994, Engineering Tribology. Oxford University Press.

[4] Stachowiak G.W., Batchelor A.W., 2005, Engineering tribology. Elsevier Butterworth-Heinemann.

[5] Johnson K.L., Cameron R., 1967, Fourth Paper: Shear Behaviour of Elastohydrodynamic Oil Films at High Rolling Contact Pressures, *Proceedings of the Institution of Mechanical Engineers*, vol. 182, pp. 307–30.

[6] Johnson K.L., Tevaarwerk J.L., 1977, Shear Behaviour of Elastohydrodynamic Oil Films, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 356, pp. 215–36.

[7] Barus C., 1893, Isothermals isopiestics and isometrics relative to viscosity, *American Journal of Science*, vol. s3-45, pp. 87–96.

[8] Roelands C.J.A., 1966, Correlational aspects of the viscosity-temperature-pressure relationship of lubricating oils. University of Groningen.

[9] Houpert L., 1985, New Results of Traction Force Calculations in Elastohydrodynamic Contacts, *Journal of Tribology*, vol. 107, pp. 241-5.

[10] Gohar R., Rahnejat H., 2012, Fundamentals of Tribology, Imperial College Press.

[11] Reynolds O., 1886, On the Theory of Lubrication and Its Application to Mr Beauchamp Tower's Experiments, Including an Experimental Determination of the Viscosity of Olive Oil, *Philosophical Transactions of the Royal Society of London*, vol. 177, pp. 157–234.

[12] Grubin A.N., 1949, Contact Stresses in Toothed Gears and Worm Gears, Moscow.

[13] Dowson D., Higginson G.R., 1959, A Numerical Solution to the Elasto-Hydrodynamic Problem, *Journal of Mechanical Engineering Science*, vol. 1, pp. 6–15.

[14] Dowson D., Higginson G.R., Whitaker A.V., 1962, Elasto-Hydrodynamic Lubrication: A Survey of Isothermal Solutions, *Journal of Mechanical Engineering Science*, vol. 4, pp. 121–6.

[15] Crook A.W., 1958, The Lubrication of Rollers, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 250, pp. 387–409.

[16] Dyson A., Naylor H., Wilson A.R., 1965, Paper 10: The Measurement of Oil-Film Thickness in Elastohydrodynamic Contacts, *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, vol. 180, pp. 119–34.

[17] Kirk M.T., 1962, Hydrodynamic Lubrication of 'Perspex', *Nature*, vol. 194, pp. 965–6.

[18] Cameron A., Gohar R., 1966, Theoretical and Experimental Studies of the Oil Film in Lubricated Point Contact, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 291, pp. 520–36.

[19] Gohar R., 1971, Oil Film Thickness and Rolling Friction in Elastohydrodynamic Point Contact, *Journal of Lubrication Technology*, vol. 93, pp. 371-9.

[20] Jalali-Vahid D., Rahnejat H., et al., 1998, Comparison between Experiments and Numerical Solutions for Isothermal Elastohydrodynamic Point Contacts, *Journal of Physics D: Applied Physics*, vol. 31, pp. 2725–32.

[21] Hamrock B.J., Dowson D., 1976, Isothermal Elastohydrodynamic Lubrication of Point Contacts: Part 1—Theoretical Formulation,. *Journal of Lubrication Technology*, vol. 98, pp. 223-8.

[22] Hamrock B.J., Dowson D., 1976, Isothermal Elastohydrodynamic Lubrication of Point Contacts: Part II—Ellipticity Parameter Results,. *Journal of Lubrication Technology*, vol. 98, pp. 375-81.

[23] Hamrock B.J., Dowson D., 1977, Isothermal Elastohydrodynamic Lubrication of Point Contacts: Part III—Fully Flooded Results, *Journal of Lubrication Technology*, vol. 99, pp. 264-75.

[24] Chittenden R.., Dowson D., et al., 1985, A Theoretical Analysis of the Isothermal Elastohydrodynamic Lubrication of Concentrated Contacts I Direction of lubricant entrainment conincident with the major axis of the Hertzian contact ellipse, *Proceedings of the Royal Society of London*, vol. 397, pp. 245-69.

[25] Chittenden R.., Dowson D., et al., 1985, A Theroretical Analysis of the Isothermal Elastohydrodynamic Lubrication of Concentrated Contacts II General case, with lubricant entrainment along either principal axis of the Hertzian contact ellipse or at some intermediate angle, *Proceedings of the Royal Society of London A*, vol. 397, pp. 271–94.

[26] Alicona, 2018, Focus-Variation, [online] Available at: https://www.alicona.com/en/focus-variation/ [Accessed 29th June 2018].

[27] British Standards Institute, 2012, BS EN ISO 16610-21:2012 Geometrical product specifications (GPS) - Filtration. Part 21: Linear profile filters: Gaussian filters.

[28] Tallian T.E., 1967, On Competing Failure Modes in Rolling Contact, *ASLE Transactions*, vol. 10, pp. 418–39.

[29] Halme J., Andersson P., 2010, Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics – state of the art, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 224, pp. 377–93.

[30] Jacobson B., 2003, The Stribeck memorial lecture, *Tribology International*, vol. 36, pp. 781–9.

[31] Hamel M., Addali A., Mba D., 2014, Investigation of the influence of oil film thickness on helical gear defect detection using Acoustic Emission, *Applied Acoustics*, vol. 79, pp. 42–6.

[32] Tzeng S.T., Saibel E., 1967, Surface Roughness Effect on Slider Bearing Lubrication, *ASLE Transactions*, vol. 10, pp. 334–48.

[33] Christensen H., 1969, Stochastic Models for Hydrodynamic Lubrication of Rough Surfaces, *Proceedings of the Institution of Mechanical Engineers*, vol. 184, pp. 1013–26.

[34] Chow L.S.H., Cheng H.S., 1976, The Effect of Surface Roughness on the Average Film Thickness Between Lubricated Rollers, *Journal of Lubrication Technology*, vol. 98, pp. 117-24.

[35] Goglia P.R., Conry T.F., et al., 1984, The Effects of Surface Irregularities on the Elastohydrodynamic Lubrication of Sliding Line Contacts Part I—Single Irregularities, *Journal of Tribology*, vol. 106, pp. 104-12.

[36] Goglia P.R., Cusano C., Conry T.F., 1984, The Effects of Surface Irregularities on the Elastohydrodynamic Lubrication of Sliding Line Contacts Part II—Wavy Surfaces, *Journal of Tribology*, vol. 106, pp. 113-9.

[37] Karami G., Evans H.P., Snidle R.W., 1987, Elastohydrodynamic Lubrication of Circumferentially Finished Rollers having Sinusoidal Roughness, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 201, pp. 29–36.

[38] Lubrecht A.A., Ten Napel W.E., Bosma R., 1988, The Influence of Longitudinal and Transverse Roughness on the Elastohydrodynamic Lubrication of Circular Contacts, *Journal of Tribology*, vol. 110, pp. 421-6.

[39] Kweh C.C., Evans H.P., Snidle R.W., 1989, Micro-Elastohydrodynamic Lubrication of an Elliptical Contact With Transverse and Three-Dimensional Sinusoidal Roughness, *Journal of Tribology*, vol. 111, pp. 577-84.

[40] Kweh C.C., Patching M.J., et al., 1992, Simulation of Elastohydrodynamic Contacts Between Rough Surfaces, *Journal of Tribology*, vol. 114, pp. 412-9.

[41] Venner C.H., ten Napel W.E., 1992, Surface Roughness Effects in an EHL Line Contact, *Journal of Tribology*, vol. 114, pp. 616-22.

[42] Chang L., Cusano C., Conry T.F., 1989, Effects of Lubricant Rheology and Kinematic Conditions on Micro-Elastohydrodynamic Lubrication, *Journal of Tribology*, vol. 111, pp. 344-51.

[43] Chang L., Webster M.N., Jackson A., 1993, On the Pressure Rippling and Roughness Deformation in Elastohydrodynamic Lubrication of Rough Surfaces, *Journal of Tribology*, vol. 115, pp. 439-44.

[44] Chang L., Webster M.N., 1991, A Study of Elastohydrodynamic Lubrication of Rough Surfaces, *Journal of Tribology*, vol. 113, pp. 110-5.

[45] Venner C.H., Lubrecht A.A., 1994, Transient Analysis of Surface Features in an EHL Line Contact in the Case of Sliding, *Journal of Tribology*, vol. 116, pp. 186-93.

[46] Ai X., Cheng H.S., 1994, A Transient EHL Analysis for Line Contacts With Measured Surface Roughness Using Multigrid Technique, *Journal of Tribology*, vol. 116, pp. 549-56.

[47] Zhu D., Ai X., 1997, Point Contact EHL Based on Optically Measured Three-Dimensional Rough Surfaces, *Journal of Tribology*, vol. 119, pp. 375-84.

[48] Xu G., Sadeghi F., 1996, Thermal EHL Analysis of Circular Contacts With Measured Surface Roughness, *Journal of Tribology*, vol. 118, pp. 473-82.

[49] Zhu D., Hu Y.Z., 2001, A Computer Program Package for the Prediction of EHL and Mixed Lubrication Characteristics, Friction, Subsurface Stresses and Flash Temperatures Based on Measured 3-D Surface Roughness, *Tribology Transactions*, vol. 44, pp. 383–90.

[50] Venner C.H., 2005, EHL film thickness computations at low speeds: Risk of artificial trends as a result of poor accuracy and implications for mixed lubrication modelling, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 219, pp. 285–90.

[51] Zhu D., 2007, On some aspects of numerical solutions of thin-film and mixed elastohydrodynamic lubrication, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 221, pp. 561–79.

[52] Evans H.P., Hughes T.G., 2000, Evaluation of deflection in semi-infinite bodies by a differential method, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 214, pp. 563–84.

[53] Hughes T.G., Elcoate C.D., Evans H.P., 2000, Coupled solution of the elastohydrodynamic line contact problem using a differential deflection method, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 214, pp. 585–98.

[54] Elcoate C.D., Evans H.P., et al., 2001, Transient elastohydrodynamic analysis of rough surfaces using a novel coupled differential deflection method, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 215, pp. 319–37.

[55] Holmes M.J.A., Evans H.P., et al., 2003, Transient Elastohydrodynamic Point Contact Analysis using a New Coupled Differential Deflection Method Part 2: Results, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 217, pp. 289–304.

[56] Holmes M.J.A., Evans H.P., Snidle R.W., 2005, Analysis of Mixed Lubrication Effects in Simulated Gear Tooth Contacts, *Journal of Tribology*, vol. 127, pp. 61–9.

[57] Weeks I., 2015, An Experimental Investigation into the Mixed Lubrication of Steel Surfaces, PhD Thesis, Cardiff University.

[58] Hansen J., Björling M., et al., 2018, Performance and mechanisms of silicate tribofilm in heavily loaded rolling/sliding non-conformal contacts, *Tribology International*, vol. 123, pp. 130–41.

[59] Spikes H., 2015, Friction Modifier Additives, *Tribology Letters*, vol. 60.

[60] Gosvami N.N., Bares J.A., et al., 2015, Tribology Mechanisms of antiwear tribofilm growth revealed in situ by single-asperity sliding contacts, *Science*, vol. 348, pp. 102–6.

[61] Clarke A., Weeks I.J.J., et al., 2016, Running-in and micropitting behaviour of steel surfaces under mixed lubrication conditions, *Tribology International*, vol. 101, pp. 59–68.

[62] Olver A. V., 2005, The Mechanism of Rolling Contact Fatigue: An Update, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 219, pp. 313–30.

[63] Lainé E., Olver A.V., Beveridge T.A., 2008, Effect of lubricants on micropitting and wear, *Tribology International*, vol. 41, pp. 1049–55.

[64] Evans H.P., Snidle R.W., et al., 2012, Analysis of Micro-Elastohydrodynamic Lubrication and Prediction of Surface Fatigue Damage in Micropitting Tests on Helical Gears,. *Journal of Tribology*, vol. 135.

[65] Grosse C.U., Ohtsu M., editors, 2008, Acoustic Emission Testing, Springer Berlin Heidelberg.

[66] Wadley H.N.G., Technical Report, n.d., Acoustic Emission: Nature’s Ultrasound, National Bureau Of Standards, Gaithersburg USA.

[67] Unnthorsson R., Runarsson T.P., Jonsson M.T., 2008, Acoustic emission based fatigue failure criterion for CFRP, *International Journal of Fatigue*, vol. 30, pp. 11–20.

[68] Devleker K., 2016, Understanding Wavelets, Part 1: What Are Wavelets, MathWorks, [online] Available at: https://uk.mathworks.com/videos/understanding-wavelets-part-1-what-are-wavelets-121279.html [Accessed 28th March 2019].

[69] Asamene K., Sundaresan M., 2012, Analysis of experimentally generated friction related acoustic emission signals, *Wear*, vol. 296, pp. 607–18.

[70] Hase A., Mishina H., Wada M., 2012, Correlation between features of acoustic emission signals and mechanical wear mechanisms, *Wear*, vol. 292–293, pp. 144–50.

[71] Hsu N.N., Breckenridge F.R., 1981, Characterization and calibration of acoustic emission sensors, *Materials Evaluation*, vol. 39, pp. 60-8.

[72] Wada M., Mizuno M., 1989, Study on friction and wear utilizing acoustic emission. Relation between friction and wear mode and acoustic emission signals, *Journal of the Japan Society for Precision Engineering*, vol. 55, pp. 673–8.

[73] Yang L., Zhou Y.C., 2006, Wavelet analysis of acoustic emission signals from thermal barrier coatings, *Transactions of Nonferrous Metals Society of China (English Edition)*, vol. 16, pp. 270–5.

[74] Chang H., Han E.H., et al., 2009, Acoustic emission study of fatigue crack closure of physical short and long cracks for aluminum alloy LY12CZ, *International Journal of Fatigue*, vol. 31, pp. 403–7.

[75] Ramadan S., Gaillet L., et al., 2008, Detection of stress corrosion cracking of high-strength steel used in prestressed concrete structures by acoustic emission technique, *Applied Surface Science*, vol. 254, pp. 2255–61.

[76] Chung K.H., Oh J.K., et al., 2004, Particle monitoring method using acoustic emission signal for analysis of slider/disk/particle interaction, *Tribology International*, vol. 37, pp. 849–57.

[77] Ferrer C., Salas F., et al., 2010, Discrete acoustic emission waves during stick-slip friction between steel samples, *Tribology International*, vol. 43, pp. 1–6.

[78] Benabdallah H.S., Aguilar D.A., 2008, Acoustic Emission and its relationship with friction and wear for sliding contact, *Tribology Transactions*, vol. 51, pp. 738–47.

[79] Price E.D., Lees A.W., Friswell M.I., 2005, Detection of severe sliding and pitting fatigue wear regimes through the use of broadband acoustic emission, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 219, pp. 85–98.

[80] Sun J., Wood R.J.K., et al., 2005, Wear monitoring of bearing steel using electrostatic and acoustic emission techniques, *Wear*, vol. 259, pp. 1482–9.

[81] Wang L., Wood R.J.K., 2007, The influence of contact conditions on surface reaction layers formed between steel surfaces lubricated by an aviation oil, *Solar Energy*, vol. 81, pp. 1482–91.

[82] Wang L., Wood R.J.K., 2009, Acoustic emissions from lubricated hybrid contacts, *Tribology International*, vol. 42, pp. 1629–37.

[83] Nagata M., Fujita M., et al., 2012, Evaluation of tribological properties of bearing materials for marine diesel engines utilizing acoustic emission technique, *Tribology International*, vol. 46, pp. 183–9.

[84] Hase A., Mishina H., Wada M., 2013, Microscopic study on the relationship between AE signal and wear amount, *Wear*, vol. 308, pp. 142–7.

[85] Tian P., Tian Y., et al., 2015, A correlation analysis method for analyzing tribological states using acoustic emission, frictional coefficient, and contact resistance signals, *Friction*, vol. 3, pp. 36–46.

[86] Saeidi F., Shevchik S.A., Wasmer K., 2016, Automatic detection of scuffing using acoustic emission,. *Tribology International*, vol. 94, pp. 112–7.

[87] Moshkovich A., Perfilyev V., et al., 2017, Study of the transition from EHL to BL regions under friction of Ag and Ni I Analysis of acoustic emission, *Tribology International*, vol. 113, pp. 189-196.

[88] Tan C.K., Mba D., 2005, Experimentally Established Correlation Between Acoustic Emission Activity, Load, Speed, and Asperity Contact of Spur Gears Under Partial Elastohydrodynamic Lubrication, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 219, pp. 401–9.

[89] Hamel M., Addali a., Mba D., 2013, Monitoring oil film regimes with acoustic emission, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 228, pp. 223–31.

[90] Raja Hamzah R.I., Al-Balushi K.R., Mba D., 2008, Observations of Acoustic Emission Under Conditions of Varying Specific Film Thickness for Meshing Spur and Helical Gears, *Journal of Tribology*, vol. 130.

[91] Raja Hamzah R.I.I., Mba D., 2009, The influence of operating condition on acoustic emission (AE) generation during meshing of helical and spur gear, *Tribology International*, vol. 42, pp. 3–14.

[92] Al-Dossary S., Hamzah R.I.R., Mba D., 2009, Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing, *Applied Acoustics*, vol. 70, pp. 58–81.

[93] Mba D., Tan C.K., 2004, The source of Acoustic Emission during meshing of spur gears, *Proceedings of the 26th European Conference on Acoustic Emission Testing*, Berlin, vol. 46, pp. 469–74.

[94] Elforjani M., Mba D., 2008, Detecting the onset, propagation and location of non-artificial defects in a slow rotating thrust bearing with acoustic emission, *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 50, pp. 264–8.

[95] Tan C.K., Mba D., 2005, Correlation between acoustic emission activity and asperity contact during meshing of spur gears under partial elastohydrodynamic lubrication, *Tribology Letters*, vol. 20, pp. 63–7.

[96] Eftekharnejad B., Mba D., 2009, Seeded fault detection on helical gears with acoustic emission, *Applied Acoustics*, vol. 70, pp. 547–55.

[97] Tan C., Irving P., Mba D., 2007, A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears, *Mechanical Systems and Signal Processing*, vol. 21, pp. 208–33.

[98] Toutountzakis T., Mba D., 2003, Observations of acoustic emission activity during gear defect diagnosis, *NDT & E International*, vol. 36, pp. 471–7.

[99] Couturier J., Mba D., 2008, Operational Bearing Parameters and Acoustic Emission Generation, *Journal of Vibration and Acoustics*, vol. 130.

[100] Mirhadizadeh S.A., Moncholi E.P., Mba D., 2010, Influence of operational variables in a hydrodynamic bearing on the generation of acoustic emission, *Tribology International*, vol. 43, pp. 1760–7.

[101] Toutountzakis T., Tan C.K., Mba D., 2005, Application of acoustic emission to seeded gear fault detection, *NDT and E International*, vol. 38, pp. 27–36.

[102] Raja Hamzah R.I., Mba D., 2007, Acoustic Emission and Specific Film Thickness for Operating Spur Gears, *Journal of Tribology*, vol. 129, pp. 860-7.

[103] Mirhadizadeh S.A., Mba D., 2009, Observations of acoustic emission in a hydrodynamic bearing, *Journal of Quality in Maintenance Engineering*, vol. 15, pp. 193–201.

[104] Douglas R.M., Steel J.A.J., Reuben R.L., 2006, A study of the tribological behaviour of piston ring/cylinder liner interaction in diesel engines using acoustic emission, *Tribology International,* vol. 39, pp. 1634–42.

[105] Loutas T.H., Kalaitzoglou J., et al., 2008, A Novel Approach for Continuous Acoustic Emission Monitoring on Rotating Machinery Without the Use of Slip Ring, *Journal of Vibration and Acoustics*, vol. 130.

[106] Singh a., Houser D.R., Vijayakar S., 1999, Detecting Gear Tooth Breakage Using Acoustic Emission: a Feasibility and Sensor Placement Study, *Journal of Mechanical Design*, vol. 121, pp. 587.

[107] Pullin R., Clarke A., et al., 2010, Detection of Cracking in Gear Teeth Using Acoustic Emission, *Applied Mechanics and Materials*, vol. 24–25, 2010, p. 45–50.

[108] Pullin R., Clarke A., et al., 2012, Identification of the Onset of Cracking in Gear Teeth Using Acoustic Emission, *Journal of Physics: Conference Series*, vol. 382.

[109] Scheer C., Reimche W., Bach F., 2007, Early fault detection at gear units by acoustic emission and wavelet analysis, *Journal of Acoustic Emission*, vol. 25, pp. 331–40.

[110] Elforjani M., Mba D., 2010, Accelerated natural fault diagnosis in slow speed bearings with Acoustic Emission, *Engineering Fracture Mechanics*, vol. 77, pp. 112–27.

[111] Niknam S.A., Songmene V., Au Y.H.J., 2013, The use of acoustic emission information to distinguish between dry and lubricated rolling element bearings in low-speed rotating machines, *The International Journal of Advanced Manufacturing Technology*, vol. 69, pp. 2679–89.

[112] Vicuña C.M., 2014, Effects of operating conditions on the Acoustic Emissions (AE) from planetary gearboxes, *Applied Acoustics*, vol. 77, pp. 150–8.

[113] Qu Y., He D., et al., 2014, Gearbox tooth cut fault diagnostics using acoustic emission and vibration sensors a comparative study, *Sensors*, vol. 14, pp. 1372–93.

[114] Wirtz S.F., Beganovic N., et al., 2016, Frequency-based damage detection of spur gear using wavelet analysis, *8th European Workshop On Structural Health Monitoring*, pp. 5–8.

[115] Novoa A.B., Vicuña C.M., 2016, New aspects concerning the generation of acoustic emissions in spur gears, the influence of operating conditions and gear defects in planetary gearboxes, *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 58, pp. 18–27.

[116] Strömbergsson D., Marklund P., et al., 2017, Acoustic emission monitoring of a mechanochemical surface finishing process, *Tribology International*, vol. 112, pp. 129–36.

[117] Merritt H.E., 2006, Gear tooth contact phenomena, *ARCHIVE: Proceedings of the Institution of Mechanical Engineers 1847-1982 (Vols 1-196)*, vol. 176, pp. 141–63.

[118] Alanou M.P., 2006, Study of parameters influencing surface distress of gears. PhD Thesis, Cardiff University.

[119] Clarke A., Weeks I.J.J., et al., 2016, An investigation into mixed lubrication conditions using electrical contact resistance techniques, *Tribology International*, vol. 93, pp. 709–16.

[120] Davies C.N., 2005, Effects of non-newtonian rheology on the line contact elastohydrodynamic lubrication problem. PhD Thesis, Cardiff University, 2005.

[121] Defence Procurement Agency, 2006, Defence Standard 91-74, Lubricating Oil, Steam Turbine and Gear, Extreme Pressure Joint Service Designation: OEP-80.

[122] REM Surface Engineering, 2018, Surface Finishing Process | Isotropic Superfinishing Process | ISF, [online] Available at: https://www.remchem.com/services/isf-process/ [Accessed 29th June 2018].

[123] Ingle V.K., Proakis J.G., 2012, Essentials of digital signal processing using Matlab, 3rd ed, Cengage Learning.

[124] MathWorks, n.d., Smoothing Splines, [online] Available at: https://uk.mathworks.com/help/curvefit/smoothing-splines.html [Accessed 28th Feb 2019].

[125] Seber G.A.F., Wild C.J., 1989, Nonlinear Regression. Hoboken NJ USA, John Wiley & Sons Inc.

[126] Oila A., Bull S.J., 2005, Assessment of the factors influencing micropitting in rolling/sliding contacts, *Wear*, vol. 258, pp. 1510–24.

[127] Jamali H.U., Sharif K.J., et al., 2015, The Transient Effects of Profile Modification on Elastohydrodynamic Oil Films in Helical Gears, *Tribology Transactions*, vol. 58, pp. 119–30.

[128] Lee K., Cheng H.S., 1973, Effect of surface asperity on elastohydrodynamic lubrication. Technical Report, Northwestern University, Evanston, Illinois.