Plastic Deformation and Residual Stress of Surface Roughness Asperities in Elastohydrodynamic Contact with Consideration of Surface Fatigue

by

Maasi F. AL-Mayali

Thesis submitted in candidature for the degree of
Doctor of Philosophy at Cardiff University

Institute of Mechanics and Advanced Materials
Cardiff School of Engineering
Cardiff University

July, 2017

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or another award.

Signed …… (Maasi AL-Mayali) Date ……

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed …… (Maasi AL-Mayali) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University Policy on the Use of Third Party Editors by Research Degree Student. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ……. (Maasi AL-Mayali) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (Maasi AL-Mayali) Date ……

Summary

Micropitting is a contact initiated fatigue failure on the scale of surface roughness features that occurs in lubricated concentrated contacts which are subject to combined sliding and rolling motion. It is a significant problem in hardened and ground gears which, due to the surface roughness, operate in a micro-elastohydrodynamic lubrication (micro-EHL) regime, causing cyclic loads as the roughness features interact.

To gain a much clearer understanding of the failure mechanism that is associated with rough surface elastohydrodynamic lubrication (EHL), a full theoretical model of the lubrication of gear contacts under rough surface and micro-EHL conditions is presented in this thesis. In addition, the study offers some important insights in to the influence of residual stresses on the fatigue life of rough surfaces. This significant residual stress, resulting from plastic deformation of rough surface asperities during the initial running-in process, has been instrumental in our understanding of the micropitting phenomenon.

The Abaqus FEA package has been used to perform a full elastic-plastic contact analysis of real rough surfaces using profiles taken from the surfaces of unrun test disks, which are used in micropitting tests. The analysis provides a detailed view of the plastic deformation, and the magnitude and distribution of the residual stress fields at the asperity level. The residual geometry and stress field obtained over a range of applied loads are then used to introduce the residual stress in elastic fatigue calculations based on Micro-EHL simulations. Fatigue damage and fatigue life is then obtained at the scale of the surface roughness asperities by using multiaxial and variable amplitude fatigue models based on a critical plane approach. The results obtained allow the effects of the residual stress due to running-in to be quantified. The analysis method is also applied to test disk experiments to compare the predicted fatigue life with the observed onset of micropitting.

Acknowledgements

First of all, I would like to praise and thank God for helping me to complete this thesis.

I would like to express my gratitude and thank my three supervisors: Prof. H.P. Evans, Dr. K.J. Sharif and Dr Alastair Clarke for their encouragement, thoughtful advice and support throughout my study.

I would particularly like to thank Prof. H.P. Evans for providing guidance and invaluable experience to develop my skills as a researcher and also for dealing with me as a colleague more than a student.

The work presented in this thesis would not have been possible without the financial support of the Iraqi Ministry of Higher Education and Scientific Research and University of Al-Qadisiyah.

I have also had the pleasure to share my office with a colourful group of students: Dr Sergey Khaustov, Dr. Nabeel Almuramady and Dr Ovidiu Savencu. Each of whom has been a pleasure to work alongside and interact with on a daily basis.

Thanks are also due to my friends in the Tribology research group, in particular: Dr Shaya AL-Shahrany for almost daily discussions about our projects and Abaqus software. Thanks are also due to Dr. Ingram Weeks for introducing me to the Profilometer.

My parents have given me love and support throughout my life, and without them I would not be where I am today. I am eternally grateful for everything they have given me, and am forever indebted to them.

Last but not least, I would like to thank my wife who has been supportive and understanding, and put up with me during the highs and lows of this project.

Contents

Declaration **II**

Summary **III**

Acknowledgements **IV**

Contents **V**

Chapter 1 Introduction and Literature Review **10**

1.1 Introduction 10
1.2 A brief history of tribology 10
1.3 Development of elastohydrodynamic lubrication (EHL) 11
1.4 Real rough surface in lubricated contacts 23
1.5 Lubrication regimes 28
1.5.1 Boundary lubrication 29
1.5.2 Mixed lubrication 30
1.5.3 Hydrodynamic lubrication 30
1.6 Running-in 31
1.7 Gear modes of failure 35
1.7.1 Pitting 35
1.7.2 Micropitting 36
1.7.3 Scuffing 40
1.8 Research objectives 42
1.9 Thesis organisation 42

Chapter 2 Modelling techniques suitable for residual stress with Abaqus **44**

2.1 Introduction 44

2.1.1 Numerical analysis of contact problems in Abaqus 44

2.1.2 Creating the model parts 46

2.1.3 Meshing element type and their effect 47

2.1.4 Managing mesh characteristics 49

2.1.5 Adaptive meshing and distortion control 52

2.1.6 Penetration of master surface to slave surface in contact analysis 53

2.1.7 Strain hardening 55

2.2 Repeated load contact modelling of real rough surfaces 58

2.3 Conclusion 66

Chapter 3 Extracting roughness profile from experimental tests and modelling by Abaqus/ CAE 6.12 **67**

3.1 Introduction 67

3.1.1 Rough surface topography. 68

3.1.2 Surface texture measurements 69

3.1.3 Filtering the raw profile 71

3.1.4 Description and extracting profiles of experimental work. 73

3.2 Creating the finite element model 75

3.2.1 Rough roller model construction method 75

3.2.2 Importing profiles into Abaqus/CAE and creating a two-dimensional model 77

3.2.3 Python and the Abaqus scripting interface 77

3.3 Determination of residual profile from Experiment and comparison with Abaqus finite element contact analysis 81

3.3.1 Introduction 81

3.3.2 Realignment of experimental profiles in the trace direction 83

3.3.3 Connection lines technique to relocate the deepest valley features vertically 84

3.3.4 Abaqus and experiment profiles alignment 88

3.4 Modelling details of finite element analysis package used for the contacting problems. 91

3.4.1 Introduction 91

3.4.2 The Abaqus CAE software package / 6.12V 92

3.4.3 Pre-processing stage 92

3.4.4 Simulation with Abaqus/standard 92

3.4.5 Post- processing (visualizing the finite element analysis result) 92

3.5 Abaqus /CAE working environment /6.12 V 93

3.6 Conclusion 97

Chapter 4 Basic theories of fatigue: A review **98**

4.1 Introduction 98

4.2 Principles of fatigue 99

4.2.1 Fatigue life models 99

4.2.2 Fatigue stress cycle 101

4.2.3 The fatigue life of materials 106

4.2.4 Determination of fatigue properties 109

4.3 Mean stress effects on fatigue life 111
4.4 Multiaxial fatigue approaches 113
4.4.1 Stress-based approach 113
4.4.2 Strain-based approach 118
4.5 Fatigue from variable amplitude loading 119
4.5.1 Cycle counting for irregular histories 120
4.6 Cyclic shakedown limits 122
4.7 Conclusion 124
Chapter 5 Fatigue analysis – applying multiaxial fatigue criteria based on critical plane approach **125**
5.1 Introduction 125
5.2 EHL line contact fatigue analysis 126
5.2.1 Introduction 126
5.2.2 Micro-Elastohydrodynamic theory 127
5.3 Fatigue and damage theory based on critical plane approach 152
5.4 Interpolation of residual stress 156
5.5 The Process of applying the critical plane model 162
5.6 Numerical results for critical plane models without and with residual stress 164
5.7 Conclusion 182
Chapter 6 Fatigue analysis – applying variable amplitude multiaxial fatigue criteria **183**
6.1 Introduction 183
6.2 Fatigue models based on critical plane approach 184
6.3 Fatigue damage evaluation procedures. 185
6.4 Weibull cumulative density function (*CDF*) 188
6.5 Cracking mechanisms in multiaxial fatigue 188

6.6 Fatigue analysis results and discussion 190

6.6.1 Damage distributions with and without residual stress 191

6.6.2 Results for a micropitting endurance test 209

6.7 Conclusion 215

Chapter 7 conclusions and future work **216**

7.1 Conclusions 216

7.2 Future work 220

References **222**

Chapter 1 Introduction and Literature Review

1.1 Introduction

This chapter first introduces a general overview of the field of tribology, the specialised area of EHL and reports some applications related to modes of surface failure. It will also review the literature of tribology relevant to the research. Finally, the aims of the current work and the organisation of this thesis are outlined at the end of this chapter.

1.2 A brief history of tribology

Tribology is the science and technology of two interacting surfaces in relative motion, and of related subjects and practices (Bhushan, 1998). This word derives from the Greek verb tribos, which means rubbing. Although the topic of tribology was first reported in 1966, in practice the study of tribology has existed since the beginning of recorded history. Dowson, (1998) published a comprehensive historical review of tribology from the beginning of humanity when friction was used for the generation of fire by percussion of flint stones and the rubbing of wood-on-wood. An interesting vision of the technological development of tribology relates to the oldest known wheel, which was found in an archaeological excavation in Mesopotamia and dates to around 2600 BC (Dowson, 1998). The word 'Mesopotamia' means the land between the rivers (Tigris and Euphrates) and it is located in modern Iraq. Figure1.1 shows the oldest known wheel in a copper model of a Mesopotamian chariot from Tell Agrab, Iraq (2600 BC). A tomb in Egypt that was dated several thousand years BC gives further evidence of the use of the technology of tribology, in this case the Egyptians used animal-fat as a lubricant in their chariot wheel bearings. The Greek and Roman period brought new technologies and materials, as well as scientific research into tribological matters. In Roman civilisation, the use of rolling elements and bronze to reduce

friction became a common practice. During and after the Roman Empire, military engineers devised both war machinery and methods of fortification using tribological technology.

Figure 1-1 A copper model of a Mesopotamian chariot from Tell Agrab, Iraq (2600 BC) (copyright for the image, University of Chicago).

Many important improvements occurred during the 1500s, particularly in the use of improved bearing materials. Early developments in the technology of tribology started in Scotland, before moving to Canada and the United States in the 1850s (Dowson, 1998). At the beginning of the nineteenth century, the Industrial Revolution saw rapid and impressive development of the machinery of production. This enormous industrial growth led to the demand for a better understanding of tribology. Finally, modern tribology has seen areas such as bio-tribology and nano-tribology emerge and the subject continues to play a significant role in manufacturing.

1.3 Development of elastohydrodynamic lubrication (EHL)

EHL is a type of lubrication that occurs between the non-conforming contacting surfaces that are encountered in heavily loaded machine elements (Dowson and Higginson, 1966; Bhushan, 2000; Hamrock et al., 2004). Typical examples of EHL can be found in the gear tooth contacts of large reduction gear boxes, rolling element bearings, cams and tappets, and many other heavily loaded nonconformal contacts. Nonconforming surfaces generate a

concentrated contact over a very small contact area and result in very high contact pressures, which can be several GPa in magnitude. Such extreme operating conditions may play a significant role in the breakdown of the EHL mechanism. The research in this thesis offers some important insights into the effects of residual stresses on fatigue life predictions and the damage caused to the components in these lubricated contacts. Under EHL operating conditions, the surfaces are separated by an oil film whose function is to reduce wear, reduce pitting, remove heat and prevent corrosion. The film thickness in EHL is thinner (typically 0.5–5 μm) than that in conventional hydrodynamic lubrication (Stachowiak and Batchelor, 2013). Despite this very thin layer of lubricant film its very high viscosity due to the lubricant pressure-viscosity behaviour allows it to be sufficient to prevent metal to metal contact between the interacting bodies. The pressure distribution in an EHL application is closely related to that found in a dry contact. Hertz (1881) theory determines the size of the elastic contact area for a specified non conforming contact giving the contact stress and contact area in terms of the elastic properties and load. The deformation of two equal cylindrical lenses under an applied load is shown in Figure 1.2. The problem was treated as an elastic non-conformal contact, which means that the contact happens between bodies that have surfaces with very different shapes. This model only considers elastic deformation and it neglects the effects of the asperity features of the contacting parts. Hertz theory still provides the basis for studies on non-conforming contacts. For smooth surfaces, the Hertzian theory gives a maximum subsurface shear stress with a value of about a third of the maximum Hertzian pressure. This is the position at which plastic deformation will first occur as the load is increased so that in a Hertzian contact the plastic zone is subsurface and contained by elastic material that surrounds it. This occurs at a depth well beneath the surface, of the order of the contact dimension. By introducing an inverse solution of the Reynolds equation, which gives detailed results of the pressure and film thickness, Dowson and Higginson in 1959 found a

full EHL numerical solution for the line contact problem, and they later published the famous book on EHL in 1966.

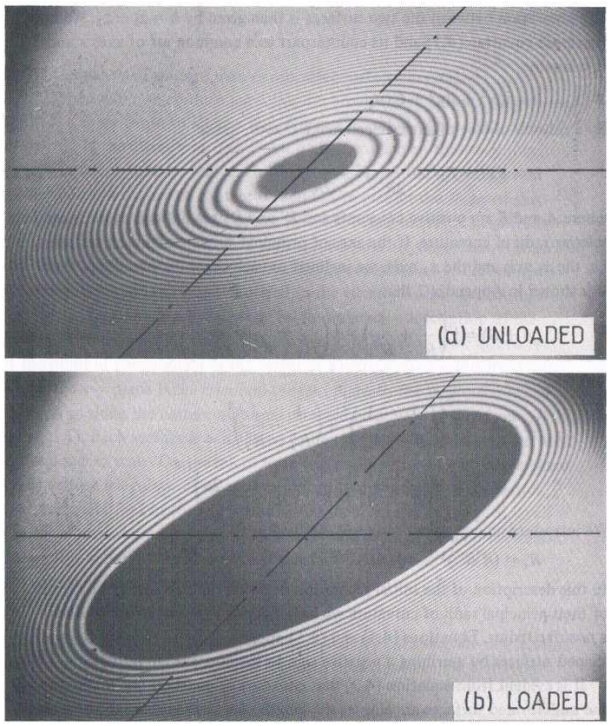


Figure 1-2 Deformations of the glass lenses under an applied load a) unloaded, b) loaded (Johnson, 1985).

Over the last three decades, many researchers have tried to predict, model and explain the elastic and elastic-plastic behaviour of material in order to get a better understanding of the deformation of bodies that touch each other at one or more contacting points. Komvopoulos et al., (1993) developed a model by using the Abaqus FEM program to understand how a changing load can cause a deformation that varies between the fully elastic and fully plastic regimes. Up to four load cycles of a load of up to three hundred times the yield load were performed in order to assess contact pressure, surface and subsurface stresses, as well as plastic zone initiation. In addition, the influence of strain hardening and residual stresses on deformation characteristics were investigated. The evolution of the contact pressure was found to depend on the strain hardening characteristics and accumulated plastic deformation

at a specific load, rather than the elastic modulus. Furthermore, Komvopoulos et al., (1993) found that, as with a smooth surface, in Hertzian contacts, the plastic zone is initially only found in the subsurface but it grows and penetrates to the surface as the load increases. As the load increases even further, the area of maximum plastic strain moves radially outwards and toward the surface, just inside the edge of the maximum contact radius (Komvopoulos et al., 1993).

As part of the wider investigations into single point analysis, a study was carried out by Kogut and Etsion, (2002) to investigate the contact of a deformable sphere against a rigid flat. The finite element method was employed to analyse this contact by using the ANSYS 5.7 package. Asymmetry was adopted so that the hemisphere was modelled as a quarter circle while the flat took the form of a line. The material was assumed to be elastic-perfectly plastic with identical behaviour under tension and compression. The Von Mises yielding criterion was used to analyse local transition from elastic to plastic deformation. The maximum interference applied was 110 times that for first yield (the critical interference), as shown in Figure 1.3. The outcomes from this study show that completely elastic contact models are not always sufficient to understand what happened during contact analysis (Kogut and Etsion, 2002).

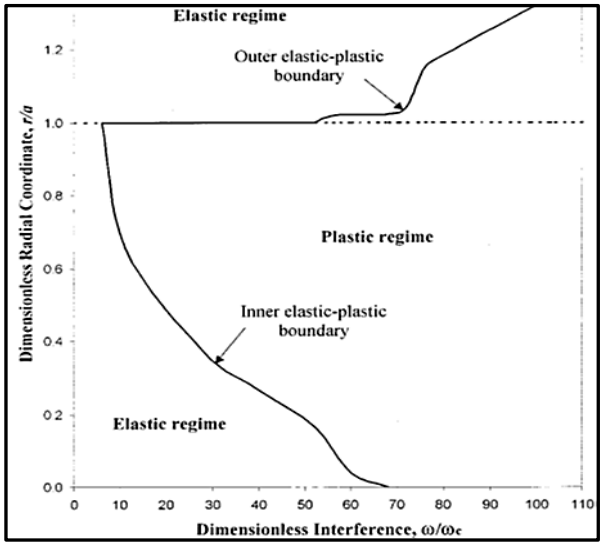


Figure 1-3 Elastic-plastic boundaries on the sphere surface (Kogut and Etsion, 2002).

In figure 1.3 the symbols, a, r, ω, and ω_c , respectively, are defined as: the radius of contact area, the radial coordinate, the interference and the critical interference.

The earliest and simplest research into the micro contact model was carried out by Greenwood and Williamson, (1966). This model adopted a contact approximation for curved elastic bodies. The authors used a new approach to measuring surface topography which is significant in modelling contact problems. This model assumed that asperities had a given statistical height distribution (Gaussian distribution) and a uniform radius of curvature for all asperities. Their model assumed that the deformation of each asperity is not affected by the deformation of the others. The Hertzian approach was used in this study to model a deformable half-space and a deformable sphere in order to calculate contacting area, load and pressure distributions (Greenwood and Williamson, 1966). A criticism of GW model is the use of the same radius of curvature for all contacting asperities while in the actual case the solution required a radius of curvature for each asperity (Jackson and Streator, 2006).The GW model has been adopted by many researchers because it provides an effective rough contact model and introduced a plasticity index that determines the critical load at which the deformation changes from elastic to plastic mode.

In the following the plasticity index ψ which was defined by Greenwood and Williamson will be defined as; $\psi = \frac{E'}{H}\sqrt{\frac{\sigma}{\beta}} \leq 0.6$ for elastic behaviour.

Where E' is the equivalent elastic modulus of the two contacting surfaces, $H$is the indentation hardness of the softer of the two surfaces, σ is the standard deviation of the asperity hight distribution and β is the typical radius of an asperity. It has also been gradually employed and improved by researchers to investigate other contacting issues involving rough surface features.

The first analytical model on the contact of rough curved surfaces was accomplished by Greenwood and Tripp, (1967) who employed the Hertzian theory of elastic contact between spheres by considering one of the spheres to be rough, so that direct contact occurs, as in practice, at a number of discrete microcontacts. It was found that the Hertzian results are valid at sufficiently high loads.

One of the first models attempting to specify the transitory state between elastic and plastic behaviour was proposed by Chang et al. (1987). This model is commonly referred to as the CEB elastic-plastic model, which is based on work by Tabor (1951). The main characteristic of the CEB model is the volume conservation of an asperity control volume during plastic deformation. The contact area of a plastically deformed asperity at any given interference was calculated at the point where material yield first occurs. The GW elastic model and the CEB elastic-plastic model give close results for the true contact area at both very low and very high values of the plasticity index.

The GW model has been employed by Beheshti and Khonsari, (2012) to examine dry rough line-contact features, such as pressure distribution and area of contact, as shown in Figure 1.4. The model involves a simultaneous solution of the asperity interaction with the elastic

bulk deformation of the surface using the Newton–Raphson technique. The results show that the important factors in the analysis of the rough surface are the pressure distribution and total load. In addition, the elastic–plastic microcontact models give a lower maximum normal pressure distribution and a greater real contact area compared to the predictions of the GW model.

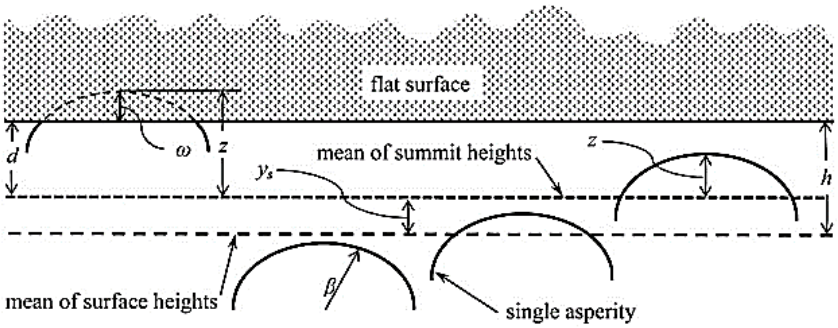


Figure 1-4 Contact of a rough surface with an ideally smooth flat surface (Beheshti and Khonsari, 2012).

Gelinck and Schipper, (1999) extended the model of Greenwood and Tripp, (1967) to the rough line contact problem. This model provides convenient empirical formulas for the contacting parameters, such as maximum pressure and contact area. Good agreement was found between the two models for a variety of parameters, including contact dimensions and pressures.The surface roughness was found to be significantly affected by the deformation of a line contact problem. Jackson and Streator, (2006) extended the GW model to examine the contact of rough surfaces by incorporating the effect of asperity deformations at multiple scales into a simple framework for modelling the contact between nominally flat rough surfaces. This model provided an estimation of the real contact area as a function of contact load and is compared to the GW model. The fast Fourier transform (FFT) was used to calculate the real asperity density and asperity radius, which are considered to be the most important parameters for describing the surface topography.The results show a greater non-dimensional contact area as a function of the non-dimensional surface load than the GW

model. This resulted in a somewhat different trend between load and contact area, as shown in Figure 1.5. The most interesting aspect of this graph is that the GW curve reflects a near linear relationship between area and load, with a slope of 0.969 in the log–log scale. However, the application of the multi-scale model show almost perfect linearity with slopes of 1.00 in the log–log domain. Note that a slope of unity reflects a direct proportionality between load and area.

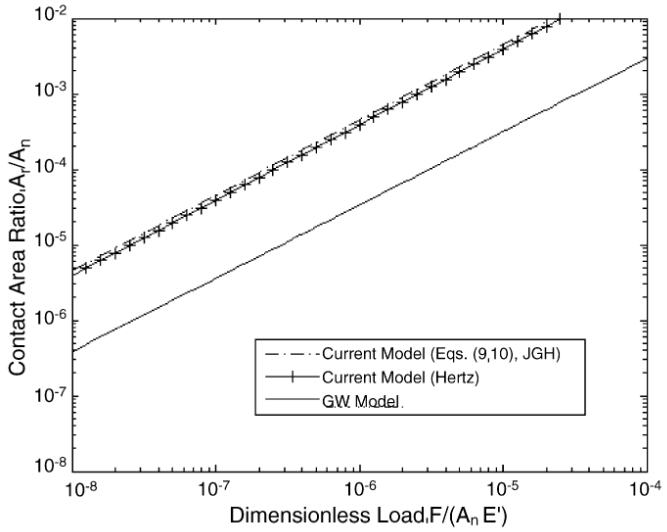


Figure 1-5 Comparison of elastic rough surface contact model (Jackson and Streator, 2006).

Supplementary research has been conducted by Jackson and Green, (2005). This study examined asperity contact and the von Mises criterion was used to define the yielding of the material. Finite element methods were used to model the case of an elastic-perfectly plastic sphere in frictionless contact with a rigid flat. The contact region was meshed by 100 contact elements. The meshed contact area was also managed to ensure that at least 30 contact elements were in contact for each applied interference (i.e., maximum contact radius error of 3.3%). These are, in essence, very stiff springs attached between surface nodes and they only activate when penetration into the rigid flat is detected. It is important to assign a large value of stiffness for these contact elements so that negligible penetration occurs between

the surfaces. However, using high stiffness can result in convergence problems. The results agree closely with Hertz theory at small interference, as shown in Figure 1.6.

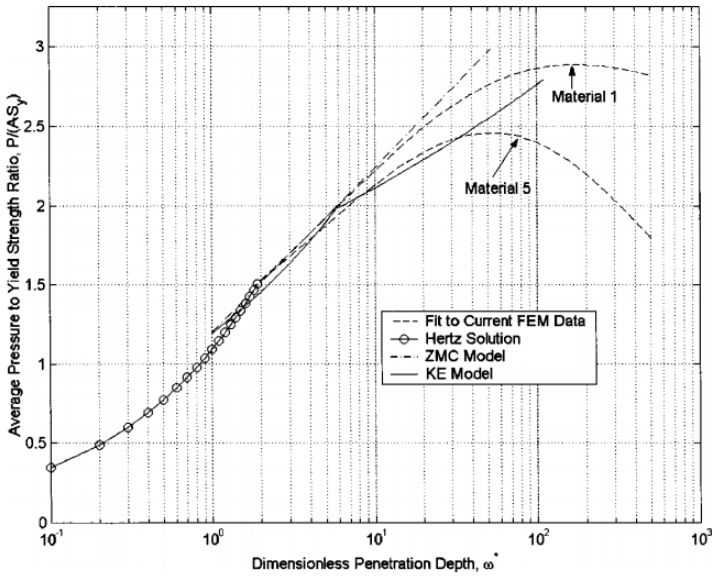


Figure 1-6 Predicted average pressure to yield strength ratio for various models (Jackson and Green, 2005).

Further investigations were conducted on an elastic–plastic sphere and a rigid flat. For example, Kadin et al., (2006) found that the majority of plastic deformation takes place in the first loading, while most of the secondary plastic flow may occur at the beginning of the first unloading. Furthermore, Kadin et al., (2006) examined a range of materials and concluded that this behaviour has significant effects on the results. Yielding in the first unloading cycle was highly affected by the changing mechanical properties, such as the Poisson's ratio and strain hardening. However, there are certain drawbacks associated with the changing material properties concerning the relation between the Poisson's ratio and high values of interference. This behaviour can be seen clearly when increasing the Poisson's ratio, which leads to increasing interference, as shown in Figure 1.7. The reason for this behaviour is still not clear to researchers.

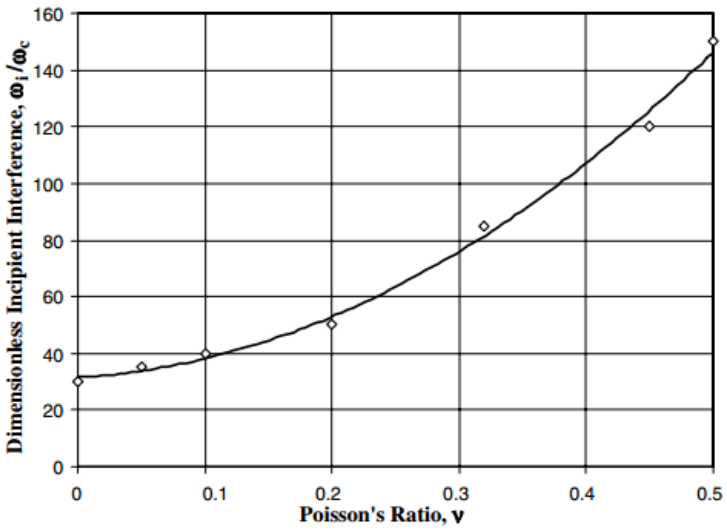


Figure 1-7 Effect of the Poisson's ratio on the interference (Kadin et al., 2006).

Similar related research to the study of rigid contact problems has been accomplished by examining a rigid body in contact with rough surfaces. This kind of theoretical model has been improved by Peng et al., (2013). In this model, a finite element method has been used to establish the elastic–plastic contact model of rough surfaces. This study investigated behaviour of a single asperity with a rigid body, as shown in Figure 1.8. The effects of material properties and surface topography were also examined. This study ignored the strain hardening effects. The elastic contact zone and contacting load differed from Hertz theory by about 3.2%. As previously described, the yield stress has significant effects on single asperity contact behaviour. Increasing material yield stress will minimise the contact area and contact load.

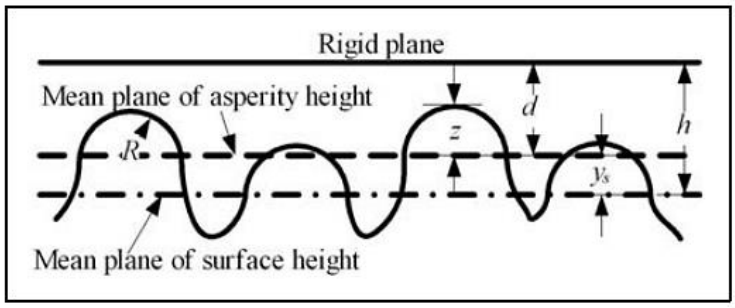


Figure 1-8 Schematic of the contact of the statistical rough surface (Peng et al., 2013).

Early studies of real rough surfaces were based on the assumption of dry rough contact; see for example,Webster and Sayles, (1986); Lubrecht and Ioannides, (1991); Snidle and Evans, (1994). Each one of these studies used numerical methods to investigate the contact behaviour of real rough surfaces. The contact pressures at micro-asperity contacts were found to be much higher than the nominal Hertzian pressure, as shown in Figure 1.9. This method has the advantage of simplicity and it can cope with both continuous and disconnected contacts.

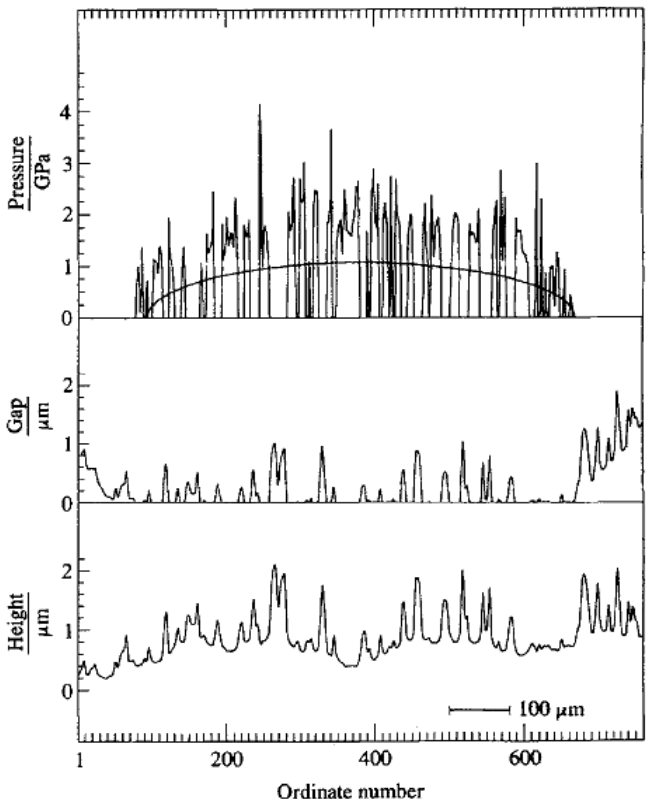


Figure 1-9 Results of simulated contact between a run-in, ground elastic surface and a rigid flat (Snidle and Evans, 1994).

Another group studying rough surface contacts looked more specifically at the evolution of the true contact area under increasing squeezing (Yastrebov et al., 2015). This model enabled the researchers to compare the numerical results both with asperity-based models at light pressures and with Persson's contact model (2002). for the entire range of pressures. In addition, this model has the advantage of simplicity and speed, and it can be used to assess

the effect of roughness parameters under increasing squeezing pressure from zero up to full contact.

As has been previously discussed, no surface is perfectly smooth. The incorporation of surface roughness into the dry contact problem has proven to be challenging and, therefore this was an active area of research before the lubricated contact problem was introduced. Further progress in the analysis of dry contacts incorporated with surface roughness was made in a series of investigations where the plastic behaviour of real rough surfaces was investigated (see for example, Jamari et al., 2007; Jamari and Schipper, 2006a; Jamari and Schipper, 2007a; 2007b; 2008; Almuramady and Borodich, 2016 and 2017). They described several experiments to confirm their prediction models and to evaluate different contact surfaces, including elastic, elastic-plastic and fully plastic surfaces. This was carried out by examining contact parameters, such as contact area, load and pressure. The rough surface was measured using an optical interference microscope technique before and after application of the contact load. Good agreement was found to occur between the Jamari and Schipper's, (2008) proposed model and experimental work. The repeated stationary contact load of rough surfaces was also considered (Jamari and Schipper, 2008). The results of the latter model showed that the tips of the most prominent asperities are plastically deformed during the initial loading cycles, as can be seen in Figure 1.10.

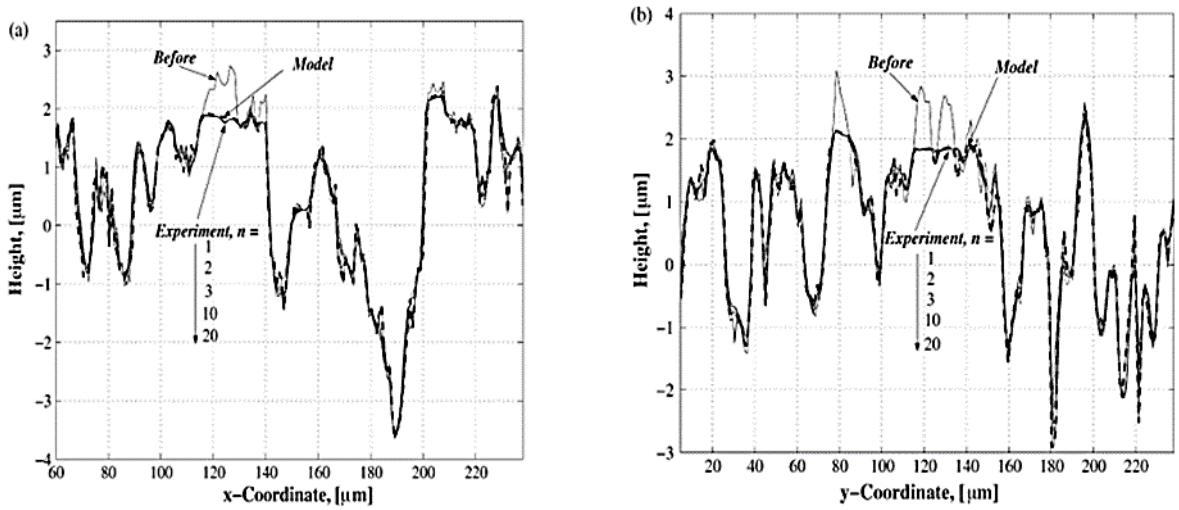


Figure 1-10 Profile of the matched and stitched isotropic surface: a) x- profile at y = 120 μm and b) y - profile at x = 129 μm (Jamari and Schipper 2008).

1.4 Real rough surface in lubricated contacts

The previous sections introduced the incorporation of surface roughness into the dry contact problem. In reality, many mechanical systems contain lubricated rolling/sliding contacts in some shape or form, especially when they operate in heavily loaded contacts. Consequently, to study the roughness effects in lubricated contacts, transient roughness behaviour has to be investigated inorder to examine the development of pressure and film thickness within the EHL contact analysis. It is believed that the transient effects that occur in gears due to the roughness of the surface which move relative to the contact and have features greater than the predicted oil film, cause interaction of surface asperities generating pressures that are well in excess of the dry Hertzian pressure. Consequently, the majority of power transmission components tend to operate in a regime that can be termed as “mixed” or “micro” EHL, in which there is a significant interaction of roughness asperities on the two contacting surfaces. Accordingly, the mixed EHL regime plays a vital role in the occurrence of surface failures of gears, such as pitting, micropitting and various other types of surface distress in which the asperities on the two surfaces come into contact (Johnson, 1989). Therefore, it will become increasingly essential to consider the nature and influence of such

contact mechanisms and their potential effects on the condition of the surface and on the life of contacting components.

Over recent years, a number of different solution approaches have been introduced in an attempt to understand the effect of incorporating a real surface into EHL models. Some of the main methods that have enabled rapid progress in this field are described in this section. The early works that modelled the incorporation of surface roughness into the EHL contact were based upon the work of Lee and Cheng, (1973); Ai and Cheng, (1994) and Jiang et al., (1999).

Lee and Cheng's, (1973) work sheds some light on the effect of surface roughness on EHL, where a single, one-dimensional asperity was run against a smooth contact surface for the conditions of pure rolling and moderate slide roll ratios. Their results showed that there was a considerable amount of direct contact occurring between the two surfaces, where the inlet pressure increases from the ambient value to a very high pressure in a short distance. This was followed by Lubrecht et al., (1986), who developed the multigrid, or multilevel, method. This method is a powerful technique that overcomes the speed difficulties encountered using single mesh level iterative solution techniques. It was used to investigate the effect of the roughness features at moderate and high loads, finding that the deformation is larger for transverse roughness comparing to the longitudinal roughness model.

Venner, (1991), using Lubrecht et al's. (1986) multilevel method, also investigated two-dimensional roughness features by developing solvers for both the EHL line and point contact problem. Point contact modelling of the EHL problem was achieved using optically measured three-dimensional rough surface profiles as input data by Zhu and Ai, (1997). It was found that, for the circular contact case simulation, surface roughness features and orientation do not have a significant effect on the average film thickness but they do significantly affect the contact pressure and asperity deformation. However, Zhu and Ai's,

(1997) analysis was done under conditions of relatively thick films and they did not consider mixed lubrication conditions.

As previously stated, the trend to include real surfaces means that the film thicknesses have tended to become very small and the scale of surface roughness features can be of the order of, or greater than, the predicted film thickness, and so metal to metal contact has occurred. This situation in which surface roughness significantly affects the performance of the contact leads to the regime of mixed lubrication.

Zhu and Hu, (2001a; 2001b); Hu and Zhu, (2000); and Masjedi and Khonsari, (2015) developed numerical models to obtain a solution to the micro-elastohydrodynamic lubrication (micro-EHL) problem under heavily loaded conditions. The results indicate that the surface topography and orientation may noticeably influence the asperity contact area and load sharing in the mixed lubrication regime. Hu and Zhu, (2000) presented a new numerical model for lubricated contact that is simple and robust, capable of handling three-dimensional measured surfaces that are rough and moving at different rolling and sliding velocities. It was found that when compared to the smooth surface solution, the same isotropic surface roughness could make the lubrication film thicker under the pure rolling conditions but made it thinner under the simple sliding condition. Evans and Hughes, (2000) developed the differential deflection technique, which reformulates the deflection equation in a differential form. The Laplacian of the deflection is obtained for a semi-infinite body subjected to a pressure loading. The method is applicable to both the line and point contact EHL problems. It was shown that the effect of pressure in a differential form is extremely localised compared to the direct approach. The advantage of this formulation is that it is not necessary to consider all pressure influences actively in the fully coupled problem. The results showed that the introduction of roughness produced large pressure spikes that are much greater than those seen in smooth surface solutions.

Elcoate et al. (2001), developed a fully coupled method using real roughness from gear tests and dealt with situations where asperities are separated by oil films that are ten or more times smaller than the size of the roughness features. The novel coupled approach actually solves for the two key variables—that is, pressure and film thickness—simultaneously within the solution scheme using both of the fundamental EHL equations.

Some novel approaches in tackling the micro-EHL problem were developed by Tao et al., (2003) and Holmes et al., (2003a, 2003b). They performed numerical analysis using real rough surfaces taken from the gear surface at multiple load stages and this analysis involves metal to metal contact within a consistent mass conserving numerical approach. These measurements showed that with increased running time and increased applied load, the roughness tips of the surface reduced and a significant amount of plastic deformation occurs during the first load stage of both gears. The asperity features become almost uniformly flattened, leaving rounded areas. Holmes et al., (2005) also developed a mixed lubrication model using a transient analysis technique where direct contact of asperities was found to occur between surfaces with various loading and running histories. Their results highlight the fact that the increased sliding velocities resulted in reduced levels of contact, as shown in Figure1.11. However, levels of contact were found to increase towards the transverse edges of the contact ellipse. It was thought that this higher occurrence of contact was due to transverse leakage of lubricant occurring at the transverse edges of the Hertzian contact, which causes a reduction in the lubricant film thickness. The contour plots show the results for the point contact EHL problem using the coupled method over a range of slide roll ratios at the same mean entrainment speed. The contact count, denoted by Q, and the slide roll ratio, denoted by ξ.

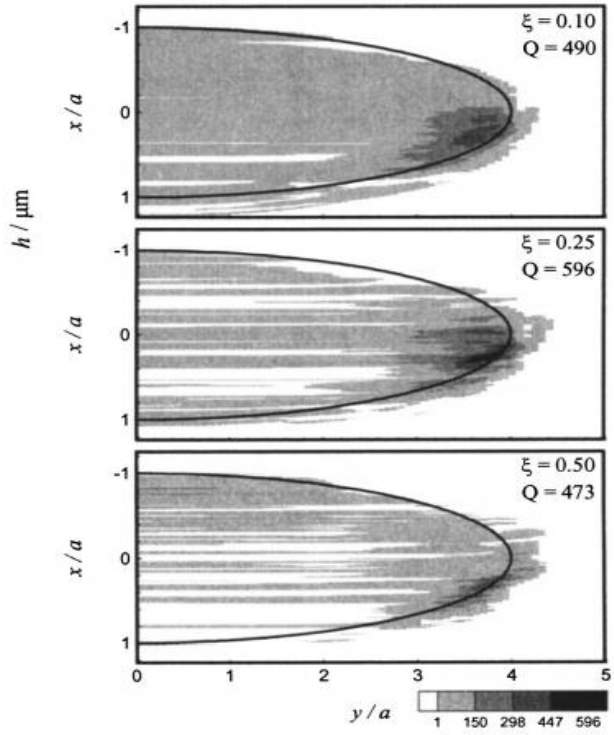


Figure 1-11 Contours of contact count rate Q/ms for the transient analysis (the heavy curve indicates a Hertzian dry contact area) (Holmes et al. 2005).

The solution method used by Holmes et al., (2005) has been extended by Weeks, (2015) to consider measured three-dimensional rough surfaces and calculated contact rates. It was found that the contact rates were skewed towards the transverse edges of the contact zone where similar results were obtained for transverse extruded surfaces (Holmes et al., 2005), as shown in Figure 1.12.

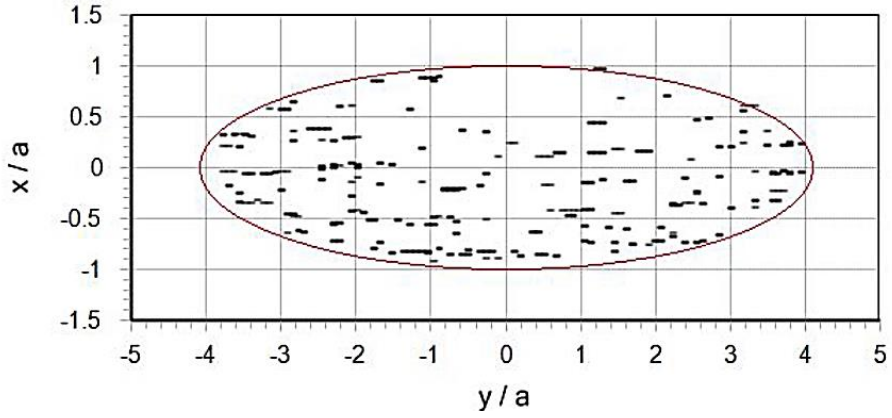


Figure 1-12 contact plot for fully rough solution (Weeks, 2015).

A recent study of real rough surfaces has been done by Evans et al., (2013), who employed the coupled solution method in a simulation of gear tooth contacts. The authors particularly highlight the effect of including real rough surfaces in the micro-EHL analysis. It was found that surface roughness may significantly affect the lubrication performance, film thickness and fatigue damage accumulation.

More recently, significant investigations have been undertaken into the mixed lubrication problem including roughness affects (see, for example, Morales et al., 2015; Dong et al., 2016; Pu et al., 2016; and Feng et al., 2017).

1.5 Lubrication regimes

Lubrication may be defined as any means that is capable of controlling friction and wear of interacting surfaces in relative motion under load. Gases, liquids, and solids have successfully been used as lubricants. Considering the nature of motion between moving or sliding surfaces, there are different types of mechanisms by which lubrication is achieved (Robinson etal., 2016). They are classified as: boundary, mixed and hydrodynamic lubrication, as shown in Figure 1.13.

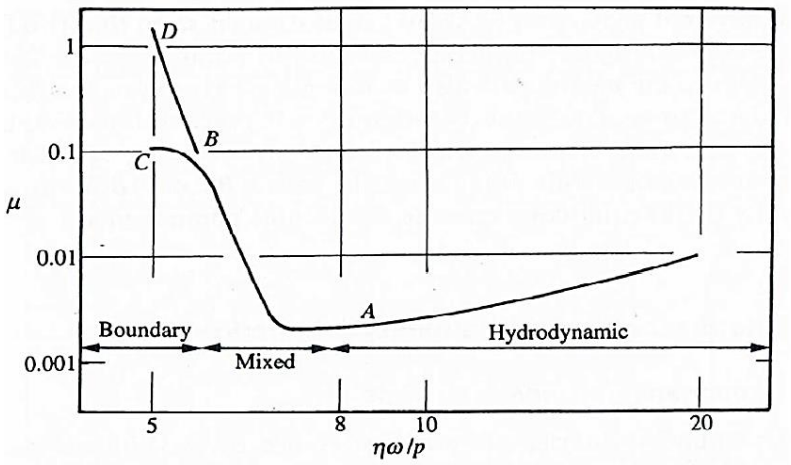


Figure 1-13 The Stribeck diagram for a journal bearing. η is the lubricant viscosity , ω the rate of rotation, and p the nominal bearing pressure (Williams, 1994).

1.5.1 Boundary lubrication

This condition can occur in a starved contact. Boundary lubrication is a condition in which the solid surfaces are so close together that the contact occurs at a number of isolated surface asperity features. The load is mainly carried by the surface in contact (asperities), so full contact (material- material contact) occurs in the equivalent Hertzian contact area. As the load increases, speed decreases or the fluid viscosity decreases in the Stribeck curve, and the coefficient of friction can increase sharply and approach high levels (about 0.1 or much higher). In this regime, the friction behaviour is controlled by an extremely thin film of lubricant and the boundary film can be attached to the surface of one or both of the contacting bodies. Effective boundary lubricant molecules are long chain molecules with a chemically active end group that will attach to the bearing surface, as shown in Figure 1.14.

Under this regime, the failure in the lubricant layer which is attached to the surface occurs by adhesive and chemical wear. Boundary lubricants form an easily sheared film on the bearing surfaces, thereby minimising adhesive wear and chemical wear (Bhushan, 2013).

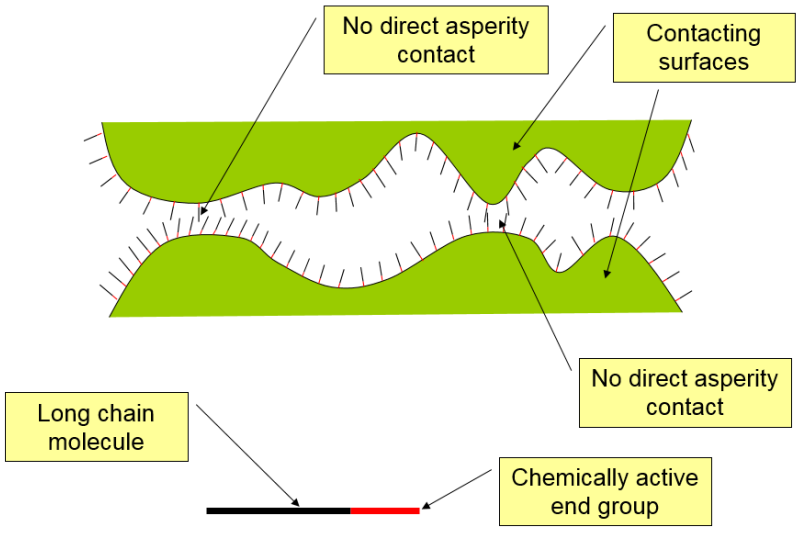


Figure 1-14 Boundary lubrication films.

1.5.2 Mixed lubrication

The transition between the hydrodynamic/elastohydrodynamic and boundary lubrication regimes is known as a mixed lubrication, in which the contact load is shared between the contacting asperities and the hydrodynamic pressure in the lubricant film. In this regime, the lubricant film is present (Evans et al. 2009) but is insufficient to separate the surfaces effectively, resulting in asperity direct contact. Mixed lubrication provides much smaller film thicknesses compared to hydrodynamic lubrication or EHL. Reduced lubricant film coincides with increased contacting load and contact pressure if the other parameters are kept constant. This characteristic is considered the basic reason for the significance of mixed lubrication. This lubrication regime is illustrated schematically in Figure 1.15.

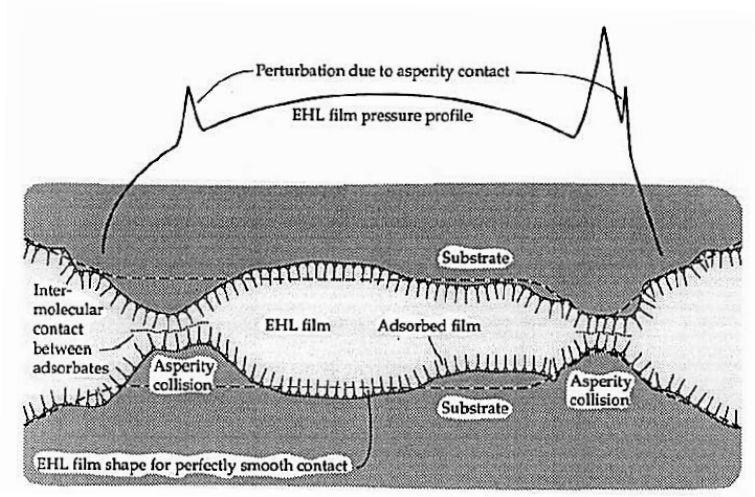


Figure 1-15 Model of mixed lubrication (Stachowiak and Batchelor, 2013).

1.5.3 Hydrodynamic lubrication

Hydrodynamic (HD) lubrication is sometimes called fluid-film or thick-film lubrication. In this type of lubrication regime, a thin layer of fluid is pulled through the contact because of viscous entrainment and is then compressed between the contacting surfaces, creating a sufficient (hydrodynamic) pressure to support the load without any external pumping. HD lubrication is considered as the ideal lubricated contact system because the lubricating films

are normally many times thicker (typically 5–500 μm) than the height of the irregularities on the bearing surface and metal direct contacts do not occur. Most commonly, contact is found in hydrodynamic lubrication regime where the pressurisation of the lubricant film results when contacting components with a convergent interface geometry slide (or roll) over a deposit of lubricant on their surface (Sherrington, 2009).

1.6 Running-in

When freshly machined surfaces are first loaded together, they have an initial settling period, where the asperity peaks are removed. This is called 'running-in' and is stated by Hutchings and Shipway, (1992) to be a series of processes during which the parameters of friction and wear for contacting bodies stabilise. These parameters are influenced by the rough surface features which change due to plastic deformation and mild wear. In addition, chemical changes that occur in the lubricant and tribo-films that form on the contacting surfaces also affect friction and the wear rate. The most aggressive contacting asperities are loaded beyond the elastic limit which results in surface topography modification due to a combination of mild surface wear and plastic deformation (Teer and Arnell 1975a; Johnson 1995; Akbarzadeh and Khonsari 2011). Thus, the components that are running against one another at lower loads and sometimes lower speeds results in reduced friction between the contacting components (Ostvik and Christen, 1968). Bishop and Snidle, (1982) performed experimental tests that investigated the running-in process as well as the scuffing limit for circumferentially ground steel disks. Their experiments showed that the harder disk was virtually unaffected by the running-in process, but roughness features of the less hard disk are modified considerably at the same load. The most striking result to emerge from this work is that the running-in process is rapid. Clarke et al., (2016), using in-situ profilometry, were able to quantify the range of surface heights present on transverse ground hardened steel surfaces under mixed lubrication conditions. They also found that the running-in is a

rapid process, where the tips of the most prominent asperities are plastically deformed during the initial loading cycles; this is shown schematically in Figure 1.16. Thus, the resulting finish has a significantly skewed height distribution, with valley features retained and asperity tips that are less prominent and more rounded.

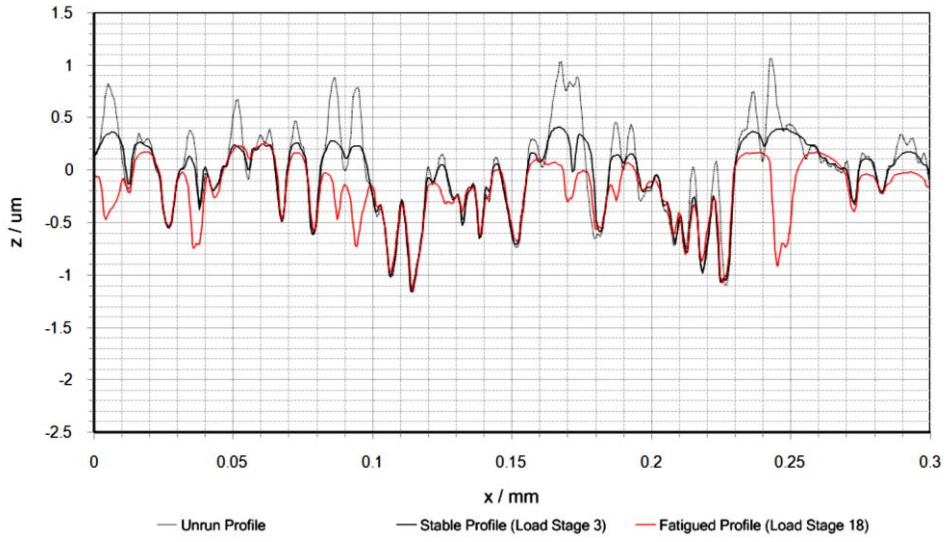


Figure 1-16 The running-in process, showing the changes in the surface topography due to plastic deformation (Clarke et al., 2016).

It has become clear in these works that the shape of asperity tips is an important factor in their running in and subsequent hydrodynamic performance.

Examples of the importance of considering running in when commissioning new surfaces can be seen in the work of Sherrington and Hayhurst, (2001). Experiments carried out monitoring both the wear particle debris density and friction coefficients made in dry sliding pin-on-disk tests. A period of running-in was applied for the pin/disk interface to ensure that the contact between the test pin and the disk was one of full contact before measurement of friction coefficients and debris improvement occur. Results showed that during running-in process the friction between the pin and disk did not have a steady value, rather it varied unexpectedly thorough a range of sliding conditions.

Blau, (2006) claimed that the initial period of the running-in process could change the subsurface microstructure of the contacting components due to plastic deformation, causing a degree of work-hardening. However, the majority of studies have concentrated on the geometric modifications in the asperities structure because this has the most immediate implications for the hydrodynamic performance of the interface.

The interest in running-in for the current project is the effect of this phenomenon on residual stresses at, or near, the surface due to the plastic deformation. This could be a significant factor in fatigue damage estimation and, thereby, affect the fatigue life of the contacting components of real rough surfaces in a micro-EHL regime.

Early experimental investigations have highlighted the importance of considering running-in of lubricated surfaces because during this process the surface profiles changed significantly due to a smoothing of the asperities by plastic deformation (Wang et al., 2000; Nogueira et al., 2002; Zhu et al., 2007; Evans et al., 2013).

The geometric modification in asperity shape during the running-in process is most commonly defined by the use of the arithmetic average roughness parameter. The average roughness (Ra) is used as a descriptor of the profile height deviations from the mean line. Barber et al., (1987) investigated experimental testing aimed at examining the running-in process as well as scuffing, using a number of contacting surfaces, with a range of Ra values attained using honing and polishing techniques. It was found that when subjected to loading the surface modification data show a significant reduction following the first 10 minutes of loading, where all materials reached a condition of nominally steady-state wear quite rapidly. The modification of the Ra value is shown in Figure 1.17. It is worth noting that the time of 10 minutes required to reach steady state will reflect the time taken to reach steady state temperature in the test. This does not contradict the observation that running in is a very

rapid process. However, in the wear test there will be a changing oil film thickness as the test component temperature changes.

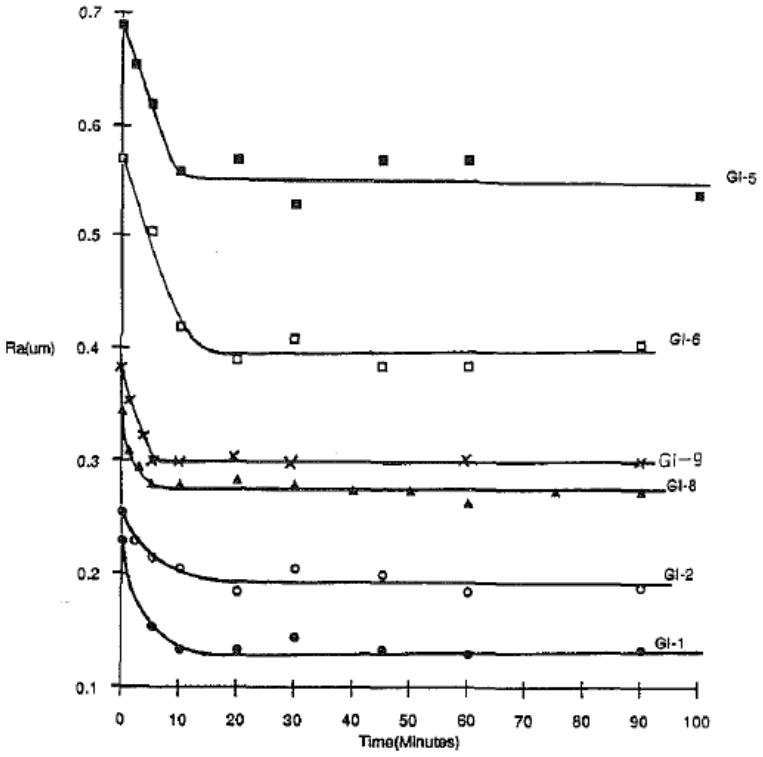


Figure 1-17 Average roughness parameter on cylinder liners versus time for a variety of different grey iron surface finishes (Barber et al., 1987).

More recently, the variation in the average roughness has been demonstrated by the work of Horng et al., (2002) when performing experimental disk on block tests, using a range of contacting loads and surface finishes, with little reference to plastic deformation during contacting bodies. A recent investigation into micropitting failure due to running-in was introduced by AL-Mayali et al., (2016), who incorporated the plastic deformation that takes place during the initial stage of running-in of gears into a full EHL solver and fatigue simulations by using a commercial finite element software (ABAQUS). Upon inspection of their results, it was found that the residual rough surface deformation only affects the asperity peaks and does not extend to the valley features. Bosman et al., (2011) made further progress in the analysis of the running-in process, where an elastic-plastic contact solver and a friction

model were used in the analysis of the running-in of lubricated concentrated point contacts. However, this study's model is based upon the approach that the plastic deformations are small and, therefore, the results of the elastic-plastic model are unrealistic for large plastic strains.

1.7 Gear modes of failure

The classical failure mode of surfaces in rolling contact components include plastic deformation, contact fatigue, pitting, micropitting and scuffing. The complex failure processes in rolling contact fatigue make it difficult to separate individual failure modes. This work will concentrate on a set of prominent failure modes of gears caused by rolling contact fatigue, which operate in the micro-EHL or mixed lubrication regime.

Contact fatigue is a surface damage process and can be defined as the mechanism of crack propagation caused by the cyclic stress field within the rolling-contact bodies, which eventually leads to subsurface plastic strain building up with increasing cycles, until a fatigue is eventually generated. The major rolling contact fatigue modes are described in the following subsections.

1.7.1 Pitting

Pitting is the development of pits or shallow cavities on the surfaces of the gears (Chue and Chung, 2000). It is well established that pitting of gear teeth is a surface fatigue phenomenon where the material undergoes cyclic contact stress. Fatigue failure is created at local stress raisers within the surfaces of components; for example, impurities, inclusions, micro-cracks or dislocations in the crystal structure.

Under cyclic loading, the local stress exceeds the yield stress of the material, which then becomes work hardened. If the loading continues, the material will fracture and this will

cause a crack that will itself act as a further stress raiser. This process will continue as the crack propagates, which has a tendency to increase rapidly. This causes the material to rupture, resulting in pitting of the surface. If pitting continues to increase, then the surfaces become modified, which increases vibration and can lead to complete failure of the gear teeth. However, pitting is increasingly being observed in hardened steel gears, occurring on the scale of the roughness asperities. Pitting is on a much larger scale than micropitting, with pits measured in microns and often associated with plastic deformation due to the running in process. An example of this is given in Figure 1.18.

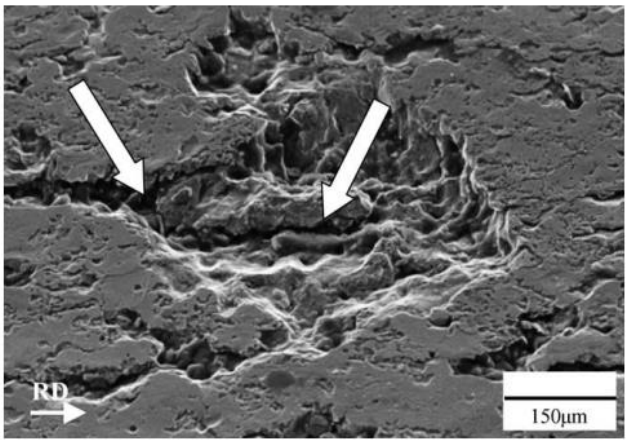


Figure 1-18 Micrograph of pitting failure on gear tooth surface (Aslantas and Tasgetiren, 2004).

1.7.2 Micropitting

Micropitting (as shown in Figure. 1.19) is a type of surface fatigue that is associated with roughness effects under mixed lubrication conditions (Evans et al., 2013). Pitting failures occur on the working faces of gear teeth, but can also occur in rolling element bearings. This type of fatigue leads to destructive wear, which can take place within the first few hours of running. There has previously been a tendency to regard micropitting as a secondary wear problem and more attention has been concentrated on pitting, which occurs due to contact fatigue at the Hertzian scale. Micropitting is on a much smaller scale than general pitting,

with pits measured in microns and normally characterised by local plastic deformation due to the running-in process. This pitting will cause the creation of small pits, typically a few microns in dimension: 10–30 μm in diameter and 5–10 μm deep (Evans et al., 2013). Micropitting is more severe in the dedendum of the teeth, where it gives a frosted or matted appearance in comparison to regions without damage.

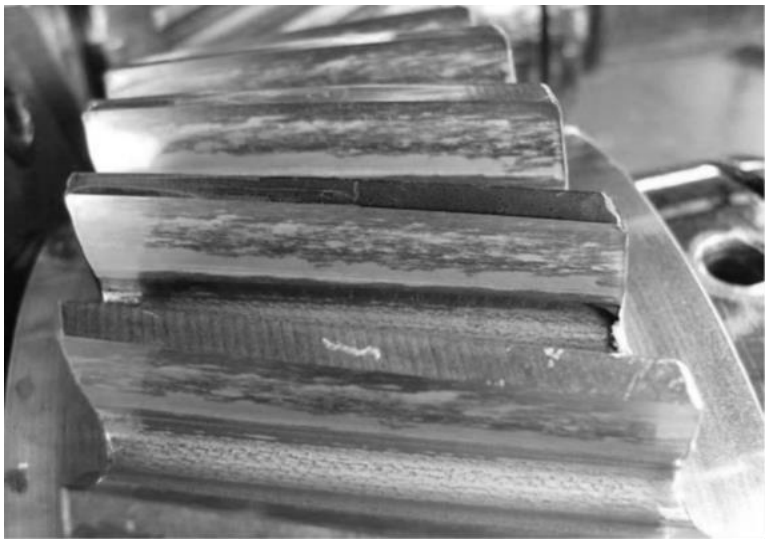


Figure 1-19 Micropitting predominantly in the dedendum region of a helical test gear tooth (Evans et al. 2013).

Some recent work in tackling the micropitting problem were studies by Li and Kahraman (2013 and 2014). They developed the boundary element approach (BEM) contact models for rough surfaces under both the line contact and point contact problems. This approach is used in the stress model to fully capture the measured three-dimensional features for the rough contacting surfaces and the variations in the near surface stress concentrations. These stresses were then used in a fatigue model using a multi-axial fatigue criterion for predicting micro-pitting failures of lubricated rough contacts under combined rolling and sliding conditions. In these two studies, the nearly singular behaviour of the stress kernels was developed to improve the numerical accuracy and efficiency in the integration for the stress

computation. However, they state that the existence of the strong singularity demands significant amount of computational time in the subdivision process that follows.

An additional significant parameter in evaluating micropitting performance is the level of sliding experienced at the contact; a greater level of sliding has been connected with poorer micropitting performance where the reason for this could be due to a higher rate of pressure cycling as asperities pass one another and also due to the shear-thinning behaviour of the lubricant (Li and Kahraman, 2013). Examples of this can be seen in micropitted gears where fatigue damage has a tendency to establish towards the root of the dedendum, where sliding is considered to be greatest (Bull et al., 1999). This was quantified in the work of Qiao et al., (2008) where comparisons of calculated fatigue for different slide roll ratios were made.

Alongside the advances in simulations, experimental work into micropitting has also been conducted. Results from experimental rolling contact fatigue work have shown micropitting failure to be heavily dependent on the surface roughness effect, where it is considered the driving force behind the high subsurface stresses and micropitting initiation (Li and Kahraman, 2011).

In addition to the high subsurface stresses introduced by the surface roughness, it is also known that crack initiation and propagation are linked to microstructural changes occurring in areas near to the material surface at the asperity level (Oila and Bull, 2005).

Oila and Bull (2005) claimed that a general feature seen in micropitted failure is the presence of small circular or semi-circular regions with diameter in the order of microns or tens of microns which have experienced material changes due to significant plastic deformation. Results from both experimentation and simulation have strongly associated micropitting failure with the surface topography (Evans et al., 2013). They found that micropits are typically 10–30 μm in diameter and up to 10 μm deep. In the current research, the calculated

micropits due to plastic deformation are shallower and about 5 to 10 μm deep, but these dimensions are those measured with a profilometer stylus rather than using (AFM) evaluations, which is compatible with Evans et al., (2013). Cracks occur at the tips of the asperities and extend at a shallow angle (typically 10–30°) to the surface (Li and Kahraman, 2011). These observations are confirmed by results obtained in the current work. The current research results shows that the positive maximum residual principal stress vectors act at angles that are tangential to the surface or at shallow angles of up to 35° to the surface. They are found at the outer edge of the asperity land and may be instrumental in forming the pits.

Brandao et al. (2012) developed a surface contact model for gear teeth by using a numerical simulation of an actual FZG spur gear micropitting test to assess surface cracking from a fracture mechanics perspective, and later a combined model, which took account of both wear and surface contact fatigue damage at asperity level (Brandao et al. 2015). In their paper, the numerical analysis is based on a model that takes overpressure effects due to mixed or boundary film lubrication into account, caused by the interaction of aggressive asperity features of the contacting surfaces. They then calculated sub-surface stresses, in order to determine the time-history of stresses and employed the Dang-Vang multi-axial high-cycle fatigue criterion to calculate micropitting experienced by the spur gear tooth flanks. They compared their results to a gear test, and found qualitative agreement. In general, the findings in these studies has revealed similar damage area characteristics to those reported in the current research as will be seen in fatigue damage analysis of chapters 5 and 6.

1.7.3 Scuffing

Scuffing is an adhesive failure that is associated with high speed and heavily loaded lubricated contacts, and is particularly common in gear teeth. High speeds can be expected to enhance film thickness, but they are also associated with high sliding speeds in gears. This results in high temperatures being developed, which lowers oil viscosity.

Kweh et al. (1992) used two sinusoids of different period and amplitude to simulate real roughness behaviour. The most striking result to emerge from the results of this work was that a significant degree of asperity deformation and corresponding ripple in the pressure distribution was found. In addition, it was found that the roughness features underwent less proportional deformation than the waviness feature.

Experimental work on the determination of the scuffing limits of material in a two disk-scuffing rig was conducted by Patching et al., (1994). Their scuffing tests were performed on a disk with a ground surface with roughness orientated transverse to the direction of oil entrainment. It was found that the scuffing scar occurred at the edges of the effective contact area and was visible as a darker band located towards the edge contact track. It was also found that scuffing did not appear to start at the position of highest contact pressure, or highest transient or 'flash' temperature generation, but was located at the relatively lightly loaded edge. This phenomenon forms the basis for a physical model of scuffing, as described by Evans and Snidle (1993).

In another study, Holmes et al., (2005) found that the primary cause of scuffing failure in gear tooth contacts was the breakdown of the EHL film as a result of direct contact between the contacting surfaces, as can be seen in Figure 1.20. Holmes et al. (2005) also suggest that this failure mode is associated with roughness effects and film thinning in gears at the scale of surface asperities features. In general, the failure approach of scuffing is based on the failure that occurs due to intermittent contact through the lubricating oil film because of

asperity contact, poor entrainment or disruption of the oil film. Furthermore, it has been established that whether or not rolling contact gives rise to scuffing can also be related to the local increase in temperature in the adjacent metal during gear contacts (Stolarski et al., 2002). This increase in temperature reduces the oil viscosity, which in turn decreases the oil film thickness and promotes further asperity contact. The beginning of scuffing implies a situation that is both thermally and mechanically unstable, and its development tends to 'snowball' progressively.

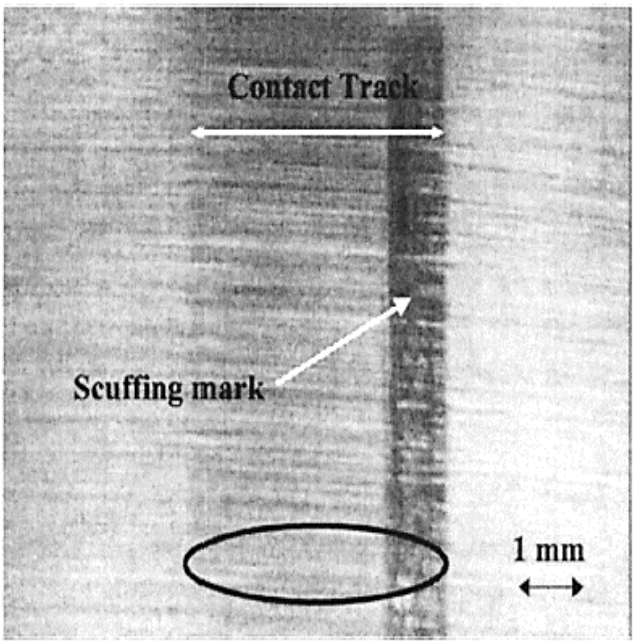


Figure 1-20 A photograph of a test disk taken from the experiential showing a track subject to scuffing damage, the Hertzian contact ellipse for the operating load at which scuffing occurred is also shown (Holmes et al. 2005).

1.8 Research objectives

The general objective of this thesis is to model mixed EHL line contact and to estimate the fatigue life at the asperity level for the actual test programme that was carried out in a micropitting test. The principal aims of this thesis are to investigate the possible consequences of the surface contact fatigue of residual stress fields resulting from plastic deformation of surface asperities, and to compare the predicted fatigue life with the observed onset of micropitting in the test.

The key elements of the work in this study which is concerned with the fundamental aspects of computing fatigue at the asperity contact level include: a mixed transient EHL model to determine the instantaneous surface contact pressure and lubricant film thickness at each point in the computing mesh, a stress field prediction model to calculate elastic stress for line contact problem (2*D* plane strain) and a shear strain model that is used to predict the number of loading cycles, which are required for micropitting fatigue to appear on the contacting surfaces. To include the residual stress due to running-in in these analyses a FEA elastic-plastic contact model was developed and a suitable strategy for combining its results with the fatigue analysis was implemented.

1.9 Thesis organisation

This thesis has seven chapters. A brief description of the remaining chapters follows.

Chapter 2 is devoted to examining the abilities of the ABAQUS FEA software system to model the contact of a rough surface. This incorporates the real rough surfaces that are used in the micropitting tests for full elastic-plastic contact analysis. The results of this analysis will determine the detailed nature of deformation, and the magnitude and distribution of the residual stress field at the asperity level. These residual stresses may have a significant effect on the fatigue life of the real rough surfaces. In addition, this chapter describes various key

aspects for generating a predictive finite element model that is capable of capturing a very large strain due to the asperity residual deformation.

Chapter 3 describes the techniques employed and the results obtained in the simulation of dry elastic/plastic contact of rough surfaces using the profile taken from the surfaces of a twin disk rig. These results focus on the asperity residual deformation and stress field, which remain following unloading of the model. The amount of plastic residual deformation that the asperities have experienced in the experimental work is also compared and correlated with the finite element analysis.

Chapter 4 introduces an overview of the concepts and theories of fatigue, which are developed for rolling contact fatigue. It will also investigate the capability of some well-known multiaxial models in fatigue life prediction. These models are evaluated numerically in the area near the rough surfaces to capture the possible high-stress gradients due to asperity contact. In the current study, multiaxial fatigue criteria based on the critical plane approach are employed to perform fatigue analysis for the EHL line contact with rough surfaces.

Chapters 5 and **6** deal with the numerical evaluation of contact fatigue performance with and without residual stress using critical plane approaches and different affecting parameters. In particular, Chapter 5 applies multiaxial fatigue criteria, while Chapter 6 uses a shear strain based model that incorporates cumulative damage analysis.

Chapter 7 summarises the main conclusions of this study, and also makes some recommendations for future work.

Chapter 2 Modelling techniques suitable for residual stress with Abaqus

2.1 Introduction

This chapter describes the approach used for elastic-plastic contact modelling using the Abaqus Finite Element (FE) package. Representative profiles have been selected from the experimental work for both disks used in a two disk testing machine. A set of FEA contact analyses were carried out to select the best nominal loading for estimating the residual stress field associated with the level of asperity modification observed in experimental work. These residual stresses, which develop in the material during running-in, may have a significant effect on the fatigue life of the surfaces.

The residual stress field will be combined with the stress resulting from EHL analysis to give improved indication of damage and assess surface fatigue life. This work to examine elastic-plastic contact analysis depends on the roughness profiles, which can also be taken from gear teeth at locations that interact during the meshing cycle. The approach can therefore be extended for use in failure analyses including micropitting tests.

2.1.1 Numerical analysis of contact problems in Abaqus

Generally, contact is the study of deformable bodies that contact each other at a minimum of one contacting point during simulation process, so contact stress is transmitted between two bodies when a load is applied. Contact problems are normally based on the continuum contact (bulk components) with many nodes in contact under load. Such contact may cause local deformation and shear, but it causes little bending. On the other hand, structural contact happens when there are a few nodes in contact at one time and contact causes bending. Contact problems can be classified into two classes.

***General Contact*:** with a single interaction description, contact is forced over many or all sections of a model. This type of contact allows the user to define contact between many or

all areas of a model with a single interaction and can span many disconnected regions of a model.

***Contact Pairs*:** Two surfaces can be described only by contact pairs. This type of analysis requires more careful definition of contact where every possible contact pair interaction must be defined as shown in Figure 2.1. This is the method chosen for the current research.

Figure 2-1 contact domain a) general contact; b) contact pairs, (Abaqus 2013)

A contact problem in Abaqus CAE software package / 6.12V requires; definition and creating of model parts that may be in contact, model meshing ,definition of pairs of surfaces that interact (contact interactions), specifying property assignments of the surfaces in contact with each other, specifying contact properties including mechanical properties (friction), thermal properties, yield stress, plastic strain ,Young's modulus. The solution scheme makes use of an algorithm to control contact interaction during the simulation and this allows same factors to be specified for example, convergence tolerances associated with contact.

In many contact applications, a user can use default Abaqus settings for each of the parameters listed above to reduce convergence problems as shown in detail in the next sections.

2.1.2 Creating the model parts

This section illustrates the development of basic contact modelling utilising Abaqus. At the first stage of the contact simulation, the real unrun roughness profile is superimposed on a 38.1 mm radius circle to form a 2D deformable part that represents the contacting part of the twin disk rig. Having taken unrun roughness profiles from a twin disk rig, 1.5mm sections of the filtered unrun gear surface profile were imported to Abaqus using a Python script, to create a 2D deformable part. The final model contained four separate parts that formed the part of the disk to be analysed, the group assembly function was used to join them to create the rough roller part that is loaded in plane strain against a rigid plane. This basically loads the rough roller part against a reflection of itself in the rigid counterface so that the profile asperities are aligned with their reflected counterparts. This rough deformable body was then brought into contact with the rigid plane by applying a small displacement to the lower edge of the elastic-plastic part in a plane strain analysis. Making contact between the two parts is a necessary first step in the Abaqus process. This displacement was then removed in the following steps, to be replaced by a distributed force applied to the lower edge in the next analysis step. The rough deformable body was subjected to a range of loads using distributed force values of 500 MPa, 750 MPa, 1250MPa and 1500 MPa using an elastic-plastic analysis incorporating strain hardening . These loads were removed in the following steps, bringing the two parts out of contact so that the residual stress and deflection could be observed. Then the best nominal loading was selected for estimating the residual stress field associated with the deformed asperity features.

The residual stress field, residual profile shape and x,y coordinates of the nodes can be extracted corresponding to the specific load for further analysis to be carried out based on the results. This model has several aspects that must be sufficiently detailed to be able to capture the important phenomena and obtain a significant correlation between the experimental and the simulation results. For example, theoretical assumptions, the

experimental conditions and other important items, including mechanical properties, element types and their resolution, FE mesh, applied loading, convergence and boundary conditions. These issues must not be complex otherwise this would increase the time cost for running. The next sections will describe the steps of simulation of rough surface contact model in detail.

2.1.3 Meshing element type and their effect

Abaqus CAE, or "Complete Abaqus Environment" gives a wide range and variety of elements for different contact problems and analysis types. The Abaqus element is characterized by different parameters such as number of nodes (depends on element shape and way of interpolation), degrees of freedom per node (depends on the field variable), formulation (small- and finite-strain shells) and integration method (reduced and full integration). Selection of the element family from the Abaqus library requires some background on the theory and application of finite element analysis (FEA). Generally, for different problems and analysis methods there are some rules on the element selection and the user should take that into consideration during the construction of a model. For example, solid elements may be used for linear and nonlinear analysis including contact, plasticity, and large deformations. Also, shell and membrane elements can be used in some special applications, but require a good level of understanding about the theories of shells and membranes. In this research the element type CPE4 (4 node, quadrilateral, plane strain, 2D element) was adopted for the rough surface model simulation. This type of element is strongly recommended by Abaqus in relation to contact modelling and high stress field gradients. Quadratic elements have corner nodes and midside nodes and do not deal with the non-linear contact problem as efficiently as linear elements. From an engineering perspective, not only does the element type affect the result accuracy, but also the mesh size selection has a significant influence on the Abaqus modelling construction that will be shown in the next sections. The mesh size significantly affects the accuracy of the numerical

simulation results and because of that it is essential to choose the best meshing strategy in modelling the rough surface. Yastrebov et al., (2011) studied the effect of mesh transition away from the contact surface as shown in Figure 2.2, with this model consisting of thousands of nodes and spending hundreds of hours to complete. Bryant (2013) also used Abaqus finite element software to investigate the mesh resolution effects. Different mesh sizes were adopted in the current research and it was found that when the mesh spacing was reduced below 1 μm mesh generation became problematical, as some of resulting asperity features caused high levels of element distortion in spite of the adaptive meshing used. It was decided to accept that the surface features of interest were obtained with a profile having a 1 μm spacing and this was then the finest element size used in the FEA analysis.

Meshing a rough surface using a 0.5 μm mesh resolution caused more element distortion under load which directly affected the convergence of the solution. Therefore, a 1.0 μm mesh resolution was found to be suitable to model the rough contacts problem, whilst keeping acceptable running times as shown in Figure 2.3. In the current research a 1.0 μm resolution size mesh was used for the rough profile for all the loads to determine residual asperity deflections. These models were found to have less element distortion and reasonable computation times as well .

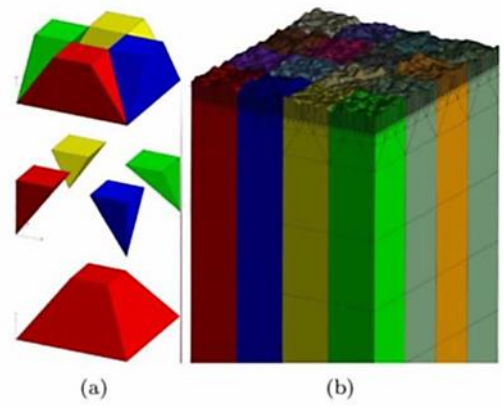


Figure 2-2 mesh transition model: a) illustration of the transition mesh from 1 to 9 elements; b) mesh of the rough surface 54 x 63 (Yastrebov et al. 2011).

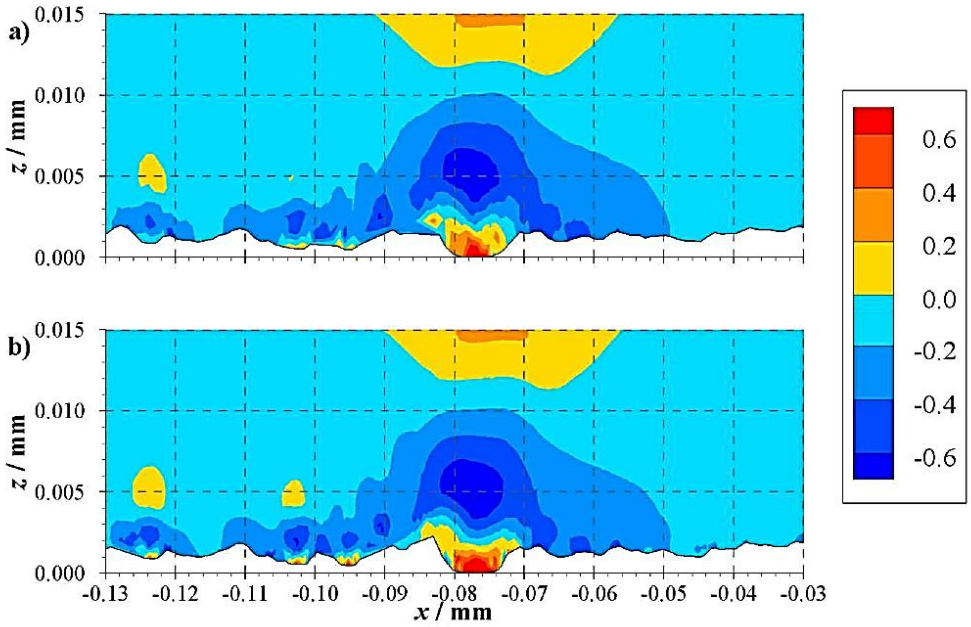


Figure 2-3 Contours of normalised residual direct stress in the x direction ; a) 1.0 µm resolution, b) 0.5 µm resolution (Bryant, 2013).

2.1.4 Managing mesh characteristics

A finite element mesh should be able to accommodate changes in element sizes from region to region. Abqaus FEM packages require a mesh to be conforming, where neighbouring elements share a whole edge or a face. The number of elements and mesh resolution at the surface is significant in terms of the accuracy of the analysis at the asperity level and the localised stresses in the surrounding material. To generate an adequate mesh and gain a high resolution at and near the contact area for the rough roller part that contains odd geometric shapes and curvature, the geometry was first divided into shapes that are more easily processed by the pre-processor tool. Then a suitable algorithm was selected to control meshing of the partitioned geometry part. The mesh is controlled by using the partition tool sets to improve mesh quality. This allowed different mesh densities to be more effectively implemented in the various regions of the model. For instance, a finer mesh was used around the contact area between the rough part and the analytical rigid part, whilst a coarser mesh was used in areas remote from it. To avoid discontinuity in results occurring at partition

edges, the transition between these different regions was kept smooth. Figure 2.4(a) shows the rough middle part mesh resolution with continuous transition in element size based on the partitioned zone. Figure 2.4(b) shows a schematic of the rough part assembled together with the surrounding three support parts. The boundary conditions used were roller supports on the transverse boundaries of the assembled model. Small vertical displacement applied to top base of rough surface to initiate contact. The pressure load was applied to the base and the assembled part was restrained by contact with a rigid plane parallel to the base as shown in the Figure 2.4(b). The rough middle part shown in Figure 2.4 had a finite element mesh that consisted of 48208 nodes and 47741 linear plane strain quadrilateral elements (CPE4). The mesh in the smallest section at the surface level is hidden in the figure because at the scale presented in the figure, the mesh size is too small to be seen. The rough roller part (middle part) was divided into five different mesh density zones where zones a - c were at respective distance of 0.01 mm, 0.03 mm and 0.05 mm from the rough surface. In addition, the last two zones were located at a depth of 0.5mm and 1.0 mm below the surface. Zone a-c had the finest mesh around the contact area and all others zones had gradually coarser meshes as the distance from the rough surface increased. This mesh density transition enabled the mesh to be conforming without the introduction of distorted elements. The other three parts that are used to restrain the transverse expansion of the central part elastically were modelled with coarser mesh sizes and were then tied to the main body using the constraints tool set.

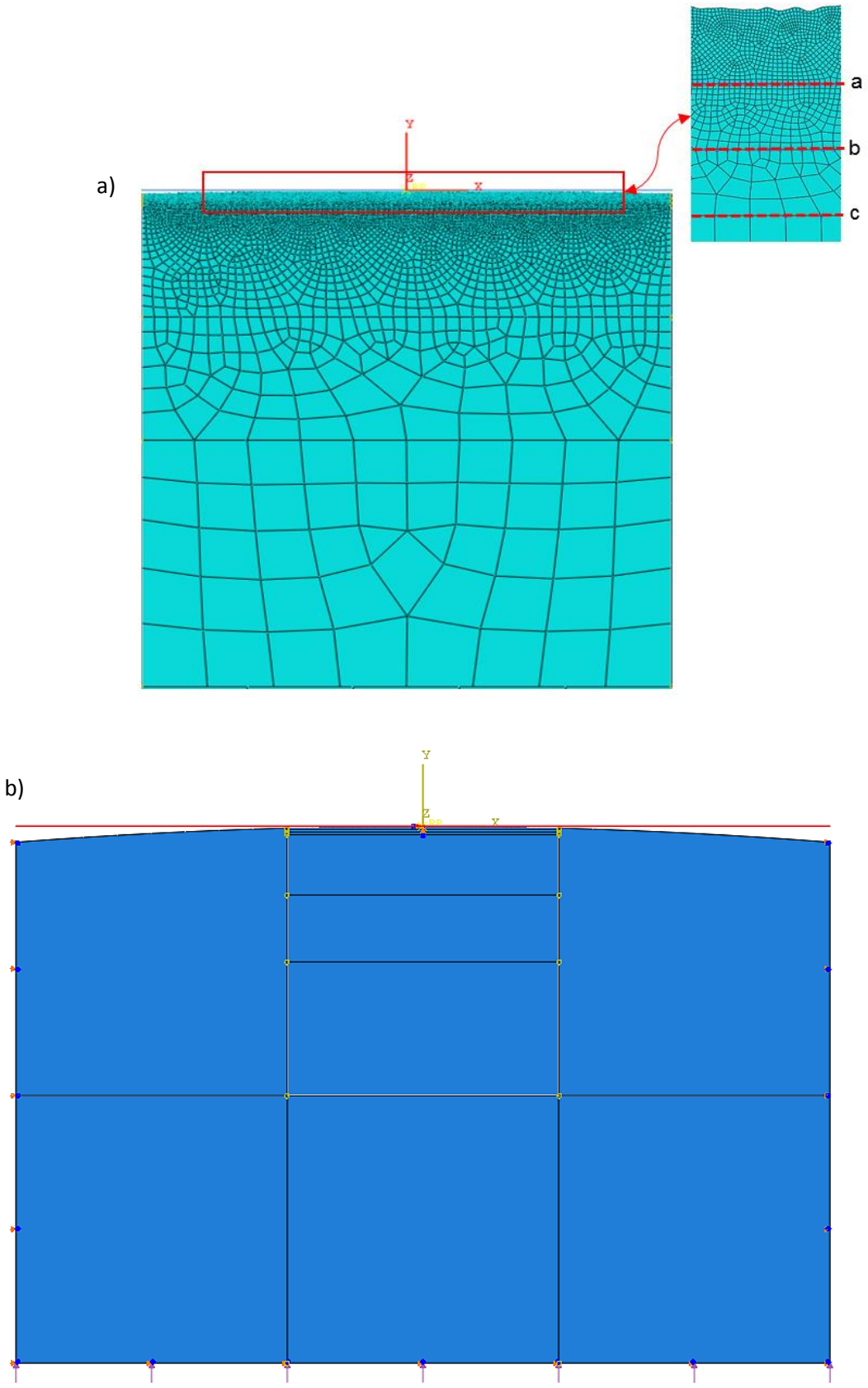


Figure 2-4 showing (a) the rough middle part mesh resolution with continuous transition in element size based on the partitioned zone, (b) a schematic of the rough part assembled together with the surrounding three support parts.

2.1.5 Adaptive meshing and distortion control

In many nonlinear contact problem simulations the material undergoes very large plastic deformations which distorts the finite element mesh and the quality of the residual stress results will deteriorate. In such a modelling situation it is essential to use the adaptive meshing technique (ALE) to periodically minimize the distortion in the mesh during the simulation. This technique is a tool set that maintains a high-quality mesh by generating a smoother mesh at regular intervals to reduce element distortion and to ensure good element aspect ratios. At the load cases corresponding to complete contact of the rough roller part it was noticed that excessive elements distortion happened when a mesh spacing of $0.5\mu m$ was adopted and so a standard 1.0 μm mesh spacing was adopted in the current research to determine the residual asperity deflections and stress fields. The mesh topology is maintained the same, the total number of elements and nodes and the connectivity of the mesh do not change or alter i.e., elements are not destroyed or created when the ALE is activated. Once the user chooses the region of the model that will use adaptive meshing, the algorithm is automatically performed. Figure 2.5 shows an example from Abaqus, (2010a) on how adaptive meshing minimizes the distortion that develops in the surface mesh of a rough roller part. It was found that increasing both the number of sweeps and ALE frequency gives the most satisfactory results. The frequency can be defined and specifies how often the mesh smoothing taken place. On the other hand, the number of sweeps controlls the number of mesh sweepings the ALE adaptive mesh undergoes during the simulation process.

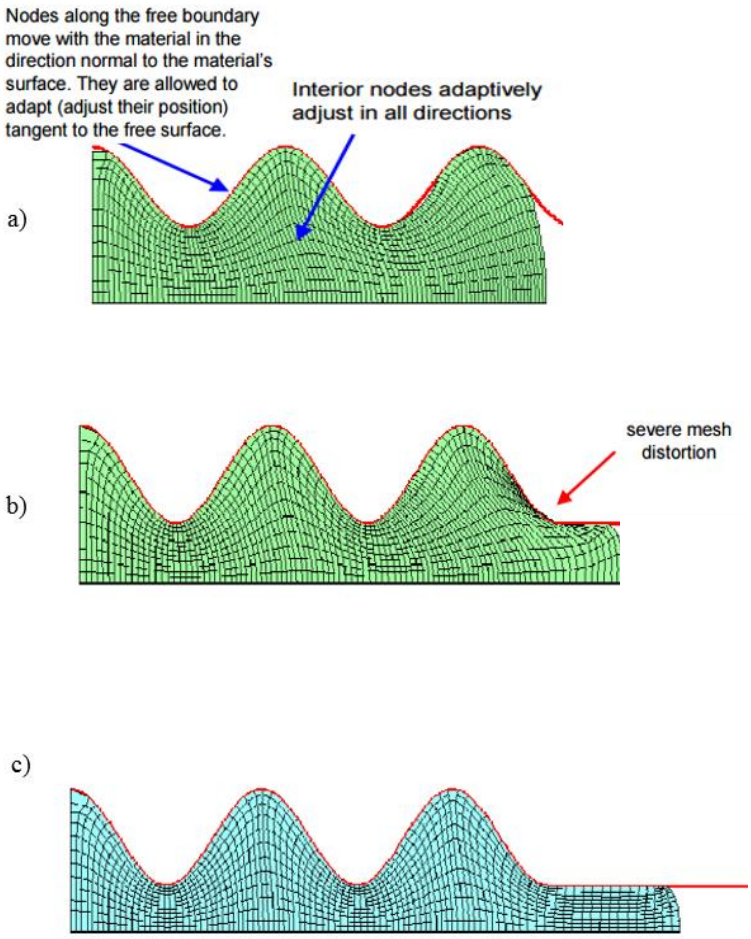


Figure 2-5 Axisymmetric forging problem with adaptive meshing, a) mesh at start of analysis, b) without adaptive meshing, c) with adaptive meshing (Abaqus 2010a).

2.1.6 Penetration of master surface to slave surface in contact analysis

The general contact algorithm in Abaqus/Standard is based upon the interactions and constraint characteristics applied to master and slave surfaces. The user must define a master and a slave surface for each contact pair. Moreover, it is recommended that analytical rigid surfaces and rigid element-based surfaces must always be master surfaces. Master surfaces should be more coarsely meshed, while the slave rough part surface should be more finely meshed and has lower stiffness. The master-slave contact algorithm places no restrictions on the master analytical rigid surface; it may penetrate the slave surface between slave nodes, as illustrated in Figure 2.6. The contact interaction and penetration definition between the surfaces can be specified using one of two approaches; node-to-surface discretization and

surface-to-surface discretization. A node-to-surface approach is based on each single slave node interacting with a group of master surface nodes. Thus, the slave nodes are assumed not to penetrate the master surface, while the nodes of the master surface can penetrate into the slave surface. Bryant, (2013) found that the node-to-surface discretization method provides less accurate residual stress results than the surface-to-surface discretization as shown in Figure 2.7, because the shape of both master and slave surfaces are not considered in the contact formulation. Therefore, to avoid such problems the surface-to-surface discretization method was chosen for the current research. Additionally, in this method the features of both master and slave surfaces are considered and that minimises the localized penetrations, as well as providing more accurate results for both the residual stresses and residual deflection results.

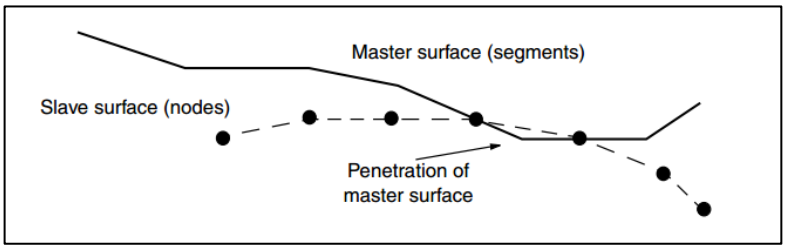


Figure 2-6 Definition of master and slave surfaces.

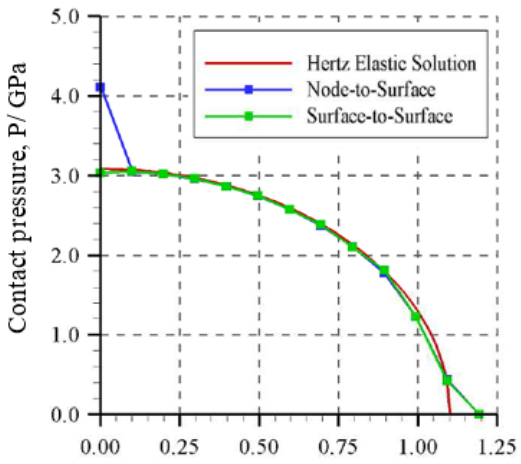


Figure 2-7 Contact pressure plots for surface-to-surface and node-to-surface contact discretization method (Bryant, 2013).

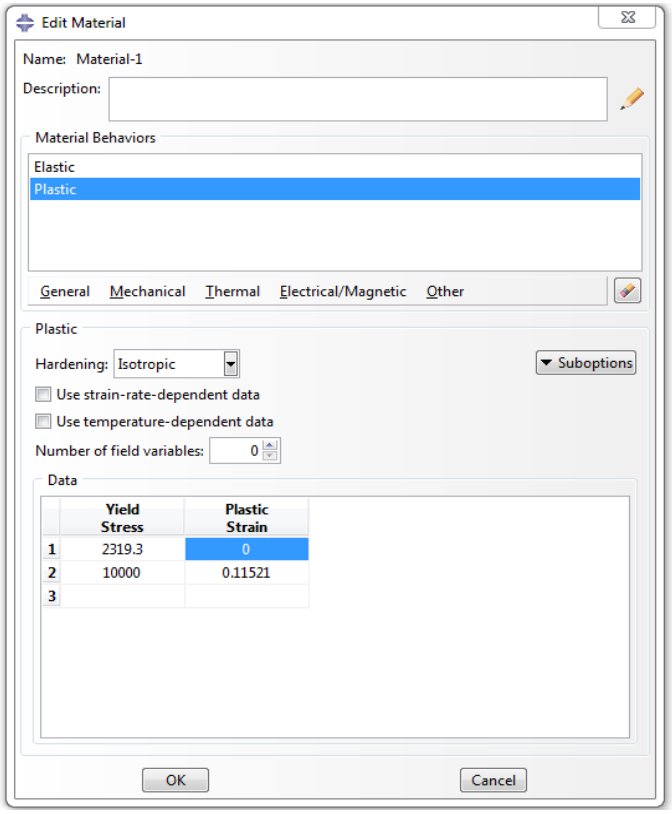
2.1.7 Strain hardening

To specify the elastic-perfectly plastic or elastic-plastic behaviour, Abaqus requires yield stress and plastic strains to be defined in terms of true strain and the true stress instead of nominal stress and nominal strain. If plasticity behaviour has not been specified for the Abaqus contact analysis, the stress/strain relationship will behave linearly. The elastic-plastic calculation of stress and strain distributions at low strains are based on linear elasticity. The onset of strain hardening is attributed to plastic behaviour and happens at a stress level regarded as the first yield stress (Batdorf and Budiansky, (1949) ; Cook et al. (1989); Bryant et al., (2012) for example). Any subsequent increase in stress with increased strain occurs according to linear strain hardening. In this approximation, the tangent modulus, ET, characterises the stress-strain relationship post-yield.

This modulus quantifies the level of ‘‘hardening” or ‘‘softening” of the model material that normally happens when it begins to yield, Shankar and Mayuram, (2008). In the model described in section 2.1.2, the elastic-plastic behaviour was defined as elastic perfectly plastic as the initial attempt of finite element simulation for the contact of rough part pressed against the rigid body. However, that analysis aborted and experienced convergence difficulties in obtaining solutions when the model reached a particular deformation. This is related to the stiffness degradation problem of model material associated with elastic-perfectly plastic behaviour. So, the concept of a linear strain hardening behaviour will be adopted for the current research to accommodate the material behaviour and to avoid the numerical convergence problems that occur with the elastic-perfectly plastic approach. In addition, this is helpful in terms of controlling the penetration and contact interaction during the simulation. A value of 50 GPa was used for the tangent modulus (E_T) in the current research, giving E_T / E ratio of 0.25. However, Kogut and Etsion, (2002) found that the majority of practical materials have $E_T / E \leq 0.05$, the strain hardening behaviour was chosen to be a relatively large value in the current simulation to evaluate its effects. This

gives a clear picture for a judgment of the suitability of the assumption of elastic- plastic behaviour. The strain hardening behaviour can be defined in Abaqus by using the property module, and is specified in the edit materials menu as shown in Figure 2.8 a.

a)



b)

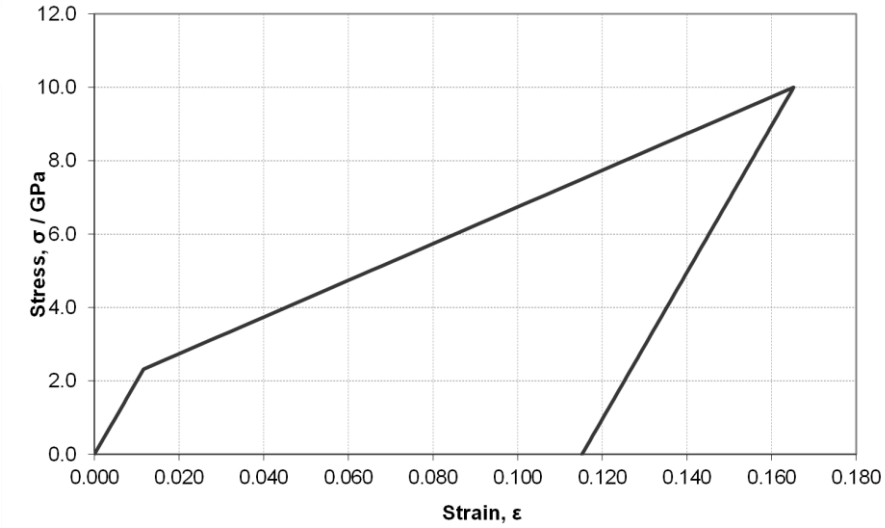


Figure 2-8 Stress-strain relationships; a) Abaqus/CAE 6.13 "Edit Materials" menu for linear strain hardening , b) linear strain hardening (E_T / E = 0.25).

The stress strain behaviour is specified by a table that gives a piecewise linear stress-strain curve. The example of Figure 2.8 specifies a strain of 0.11521 that will be achieved when the von Mises stress is 10.0 GPa in the right and left columns, respectively in the edit materials menu. This high true von Mises stress value of 10 GPa is chosen in order to ensure the required plastic behaviour is monitored throughout the material without any extension of elastic-perfectly plastic behaviour.

The mathematical calculation of initial strain input is calculated by using the following relationship.

$$\varepsilon = \sigma/E \tag{2.1}$$

However, the elastic- plastic approach uses the true stress and strain, instead of the nominal stress and strain, as the latter does not take into account the instantaneous change of cross-sectional area as the model material deforms plastically. The true stress (is the applied load divided by the actual cross-sectional area) and strain takes this factor into account. They can be expressed as follows.

$$\sigma_{true} = \sigma_{nom}\,(1 + \varepsilon_{nom}) \tag{2.2}$$

$$\varepsilon_{true} = ln\,(1 + \varepsilon_{nom}) \tag{2.3}$$

The true strain must then be converted into the true plastic strain in terms of recoverable strain (true stress/E) and true strain as shown in equation below.

$$\varepsilon_{true}^{pl} = \varepsilon_{true} - \frac{\sigma_{true}}{E} \tag{2.4}$$

Abaqus expects the true strain and stress data to be entered in the properties table in the right and left columns, respectively. Abaqus interpolates linearly between the data points provided.

2.2 Repeated load contact modelling of real rough surfaces

One of the methods used to reduce, improve the surface topography and stabilise the rates of scuffing and friction for contacting elements is commonly named ‘running in’. It is normally performed at lower loads and lower velocities, resulting in reduced friction between the contacting components and improvement in the surfaces under contact occurs rapidly.

During this research the author had the benefit of interaction with colleagues who were investigating the running-in process experimentally. The first of these studies is reported in Weeks (2015) and in Clarke et al. (2016) and Clarke et al. (2016).

Weeks used transverse ground disks in the elliptic contact illustrated in Figure 1.20. His tests were focused on the running-in process including the effect of load and speed. Tests were carried out with short running times and the disks were measured in the circumferential direction in the as-manufactured form and then after each load stage. These loads stages are referred to as LS1, LS2, ……etc and full details are given in Weeks (2015).

When a pair of disks had been run-in to a stable condition they were then used in a test sequence where the running speed was increased to a maximum in a sequence of 3 minute stages. It was then ramped down through the same stages to complete the test. A test of this sort gave a sequence of operating conditions that ran in a steady state with a speed dependent temperature.

Weeks was studying contact resistance and its dependence on steady state operating conditions. The second experimental project is ongoing and will be reported in due course,

Hutt (2018). Hutt has extended the study to include measurement of Acoustic Emission. As part of his study he conducted an endurance micropitting test where run-in surface were run for total of 1,000000 cycles in a serious of steps with surface profile measurement after each step.

The current author was able to assist Weeks in his tests and to conduct a number of the profile measurements. These profiles were used in the current research together with measurements taken by Hutt of the endurance test.

During the pre-run phenomenon the highest peaks of the rough surfaces are removed or become flattened, and the asperity features will deform plastically until the bearing area is sufficiently large to support the applied load. However, the change in heights begins to level off in the following stages of running and the difference in plastic deformation between final stages is relatively insignificant. The FEA modelling of the multiple loadings contact of the rough surfaces involved in the research inorder to approximate the running-in phenomena in gears. In the current work two axially ground disks were used for the running-in process over a period of time. It was concluded that running-in happened in, literally, a small number of the rotations of rough surface disks. Subsequent running over a long interval, did not produce further noticeable plastic deformation in the surfaces. Figure 2.9 shows a series of relocated profiles aligned by using the deep valley features for the unrun profile, as well as load stage 1 and 2 taken from the slow surface of an axially-finished disk at the same load. The largest geometric changes in the surface topography of the aggressive asperities can be noticed when comparing the unrun profile with that measured after the first load stage. Differences in the asperity shape following the first stage of running and the next stages are observed to be negligible and the changing of surface heights begins to level off. These findings strongly support the hypothesis that the running in process with rough contacting surfaces is not only a general flattening of asperities, but also an accommodation phenomena

where the plastic deformation of asperities is found by the interaction with corresponding asperities on the other contacting surfaces.

This point of view of rapid surface profile modifications is further supported by the results of elastic, contact solutions (micro-EHL) for fast and slow rough surface disks in rolling/sliding contact, as will be considered more fully in the next chapters. The results of the micro-EHL modelling shows that when fresh unrun ground gear tooth surfaces are brought together at normal operating loads of the order of 1.4 GPa nominal Hertzian maximum pressure, the maximum contact pressures at asperity contacts are far in excess of the corresponding smooth Hertzian values. This is caused during a transient simulation due to interaction of aggressive asperity features. Figure 2.10 illustrates a typical result gained from modelling of elastic-plastic contact using the ABAQUS FEA software system of the test rig disks used for Figure 2.9. The deepest valley feature for the pre and post running profiles shown at x = 3.85 mm in Figure 2.9 appears at x = 0.13 mm in Figure 2.10. The residual deflections of the roughness profile at different load stages of the same load in Figure 2.10 are basically coincident at the scale of the figures. An elastic/plastic model is used for the simulation of repeated loadings of the same roughness profiles and the applied load is such that the plastic deformation of the asperities is of the same order as that seen from comparison of the measured disk in the running-in process. Figure 2.10 shows a 160 μm length of the as manufactured roughness profile with two multiple loadings stages and Figure 2.11 shows a shorter (50 μm) length of the same roughness profile at a larger scale which shows more details. The two residual roughness profiles are almost identical and subsequent applications of the same load resulted in insignificant additional plastic deformation at the asperities due to the application of the second load. This is because the majority of the residual plastic deformation occurred in the first loading event and this

numerical behaviour is confirmed by the observation and evaluation of running-in experiments in Figure 2.9.

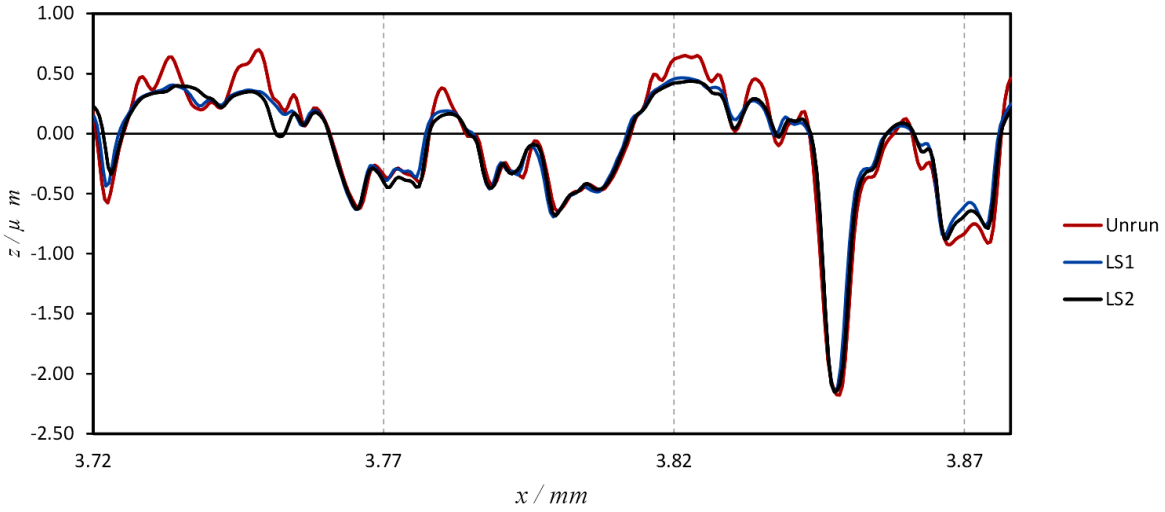


Figure 2-9 a series of re-located profiles taken from the slow disk showing the effect of running-in on the development of plastic deformation in the surface, each load stage is 27 seconds in duration; LS1 is load stage1 and LS2 is load stage 2.

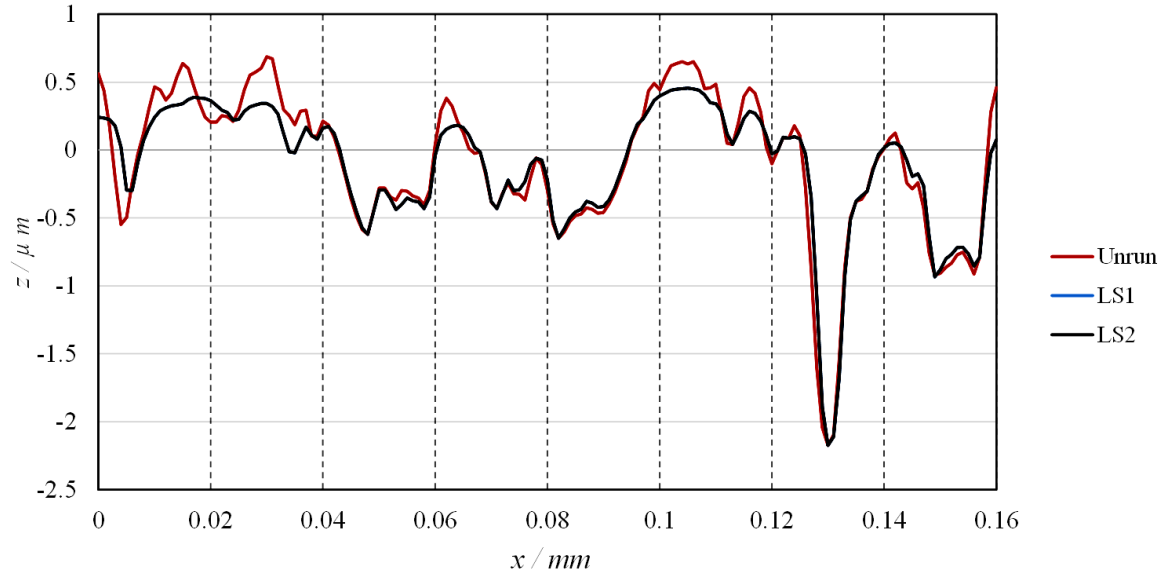


Figure 2-10 Elastic/Plastic contact simulation: Unrun roughness profile (red) with residual deflection after first load application (blue) and second load application (black).

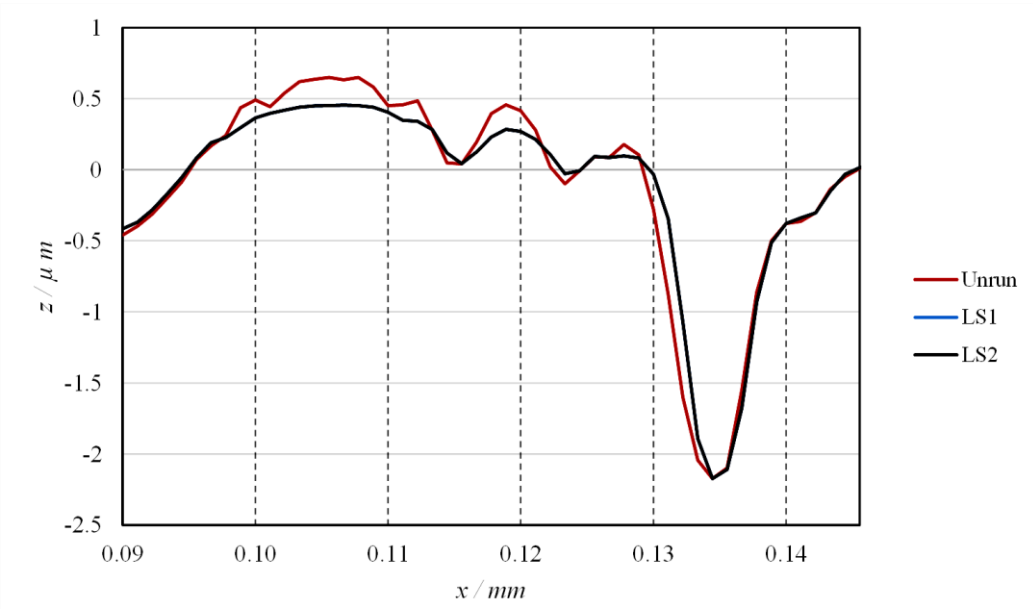


Figure 2-11 two rough asperities in greater detail.

Figure 2.12 shows the contact pressure distribution obtained from elastic/plastic contact simulation when the first and second load are applied. It is clear that there is an insignificant difference in the asperity contact pressures due to the second application of the same load. Small differences in the contact pressure occur in some heavily loaded asperities as can be seen, between x = 0.22 mm and x = 0.23 mm; x = 0.23 mm and x = 0.236 mm. It was found that the magnitude of contact pressures for some asperities is slightly in excess the limit of approximately $3\sigma_y$ given by Williams, (1994). This happens because some elements undergo considerable distortions that lead to an increase in the frequency and magnitude of contact pressures distribution at the prominent surface asperities.

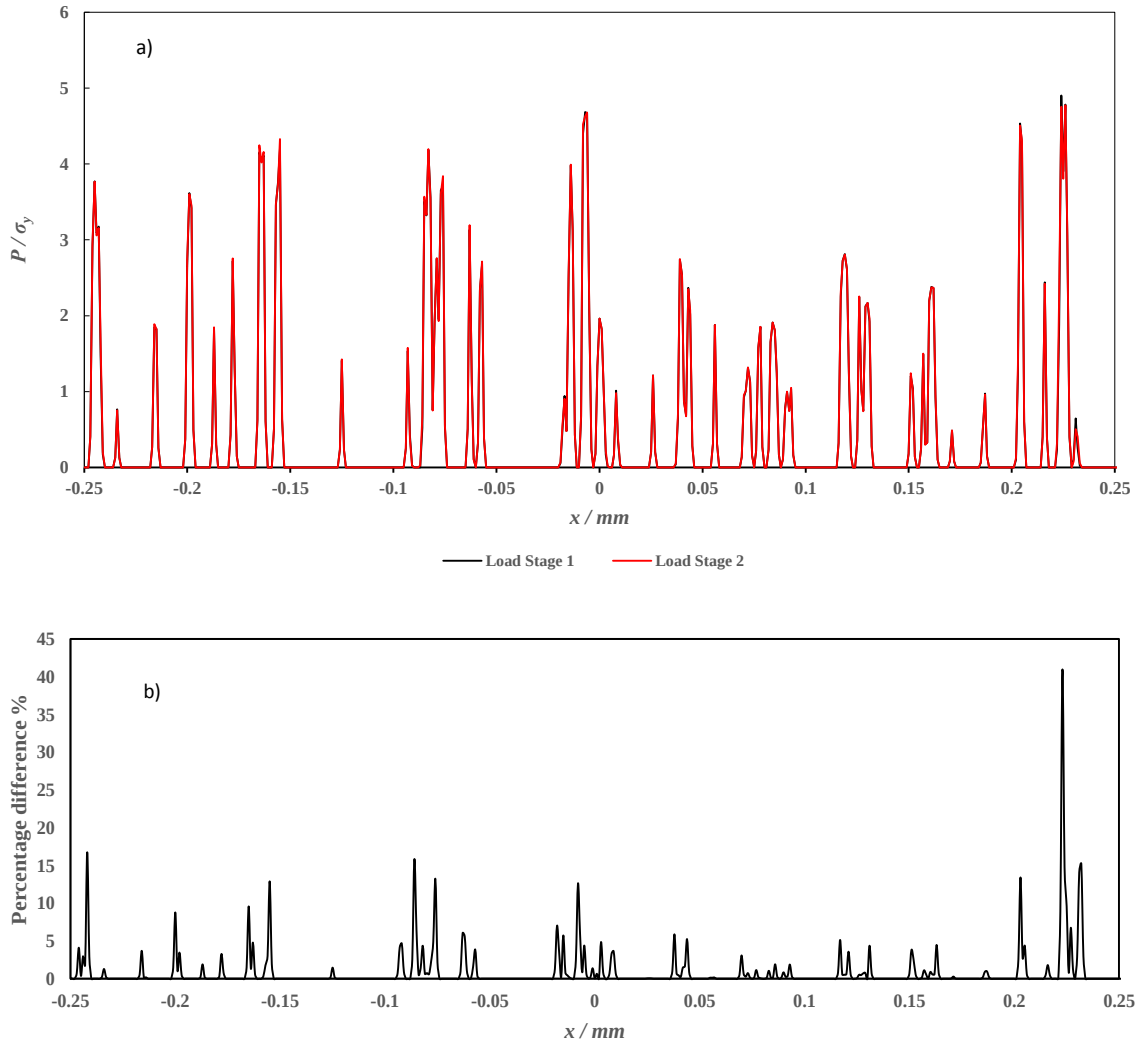


Figure 2-12: a) FEA contact pressure distributions obtained for the load stage 1 and 2; b) Percentage difference.

Figure 2.13 shows the stresses in a short section of rough surface resulting from elastic/plastic contact during loading and subsequent removal of the load. In this figure the detailed residual stress distribution and deflection shape which results at the complete contact load step is shown for the different asperities in the section. The complete contact load step is the case where the load causes the contact area to extend over the whole of the rough surface profile of the middle part of the model as illustrated in section 3.3.1. In addition, the positive principal stress vectors and the maximum principal stress at the removed load (residual) step are also described in Figure 2.13. The objective of this figure is to show the effect of (tensile) maximum principal stress and residual stress at the asperity

level on the rough surface. These are argued to cause crack initiation and fatigue failures as described by Olver, (2005). The von-Mises stress distribution is shown in the Figure 2.13d assuming the yield stress value is 1.6 GPa and the plastic deformation is indicated by the red contour area. It is noticeable from this plot that the residual deflection has a corresponding residual stress associated with it at significant asperity features. It seems possible from Figure 2.13b that the positive tensile principal stress at the surface and beneath the surface can potentially cause fatigue crack initiation at the scale of the surface asperity features.

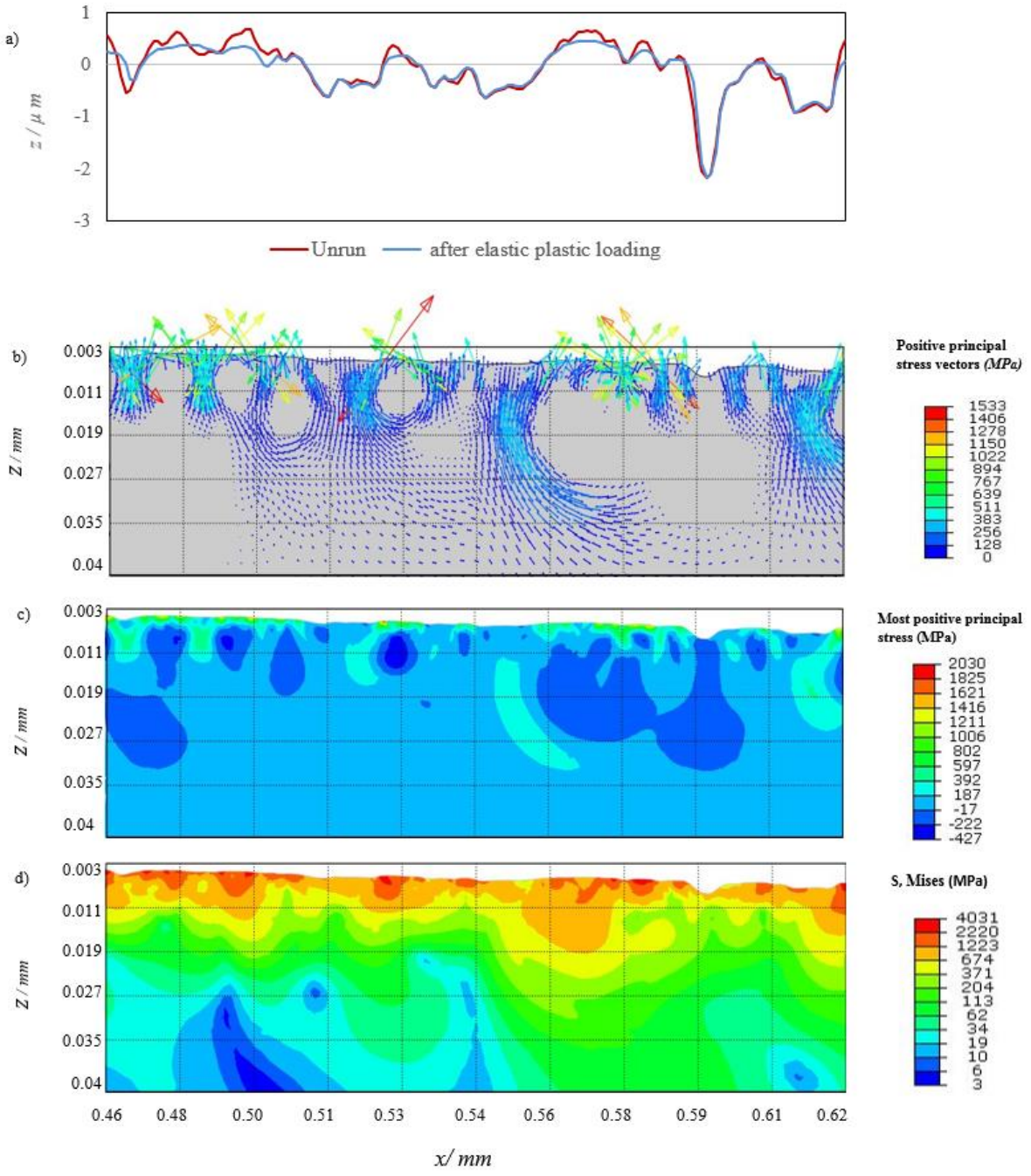


Figure 2-13 elastic/plastic contact simulation: a) Unrun (red) and after elastic plastic loading (blue); b) Positive principal residual stress vectors; c) Most positive principal stress and d) Von Mises stress.

2.3 Conclusion

This chapter describes the approach used for modelling elastic-plastic contact with strain hardening using the ABAQUS FEA software system to create the rough roller part that was then loaded in plane strain against a rigid plane. Representative surface profiles were selected from the experimental work for both unrun disks and a set of FEA contact analyses was carried out to select the best nominal loading for estimating the residual stress field associated with the level of asperity modification observed in experimental work. The residual stresses which developed in the material may have a significant effect on the fatigue life of the surfaces as will be considered in the next chapters. The residual stress field will combine with the stress resulting from EHL loading to give an improved indication of damage and to assess surface fatigue life and will be described in the fatigue assessment chapters. Simulation results of multiple loadings at the same load for the same surface show that the majority of the residual plastic deformation at an asperity level occurred in the first loading event and applications of the same load in the following stages (two stages) resulted in an insignificant amount of further plastic deformation.

Chapter 3 Extracting roughness profile from experimental tests and modelling by Abaqus/ CAE 6.12

3.1 Introduction

Generally, all surfaces in engineering applications are considered rough surfaces. Such surfaces that come into contact, will have some contact at the asperities to support the normal load. In addition, elastic or plastic deformation of smooth and rough surfaces will happen when the two bodies touch each other at the tips of one or more asperities. It has been suggested that such asperities have significant effects on the behaviour of engineering systems, having influence on stress fields , friction, real contact area, Micro-pitting and failure.

It has been claimed by many scientists (see for example, Lee and Cheng's, (1973); Zhu and Ai, (1997); Tao et al., (2003) that asperity contact leads to adhesive contacts resulting from an interaction in the contact area. Furthermore, at the start of contact between rough surfaces, contact normally happens at a few asperities in order to support the normal force. By increasing the normal force, the average distance between the two rough surfaces become smaller and the area of contact increases gradually to support the increase in load by additional contact force. Such higher load may lead to a significant deformation located in the contacting zone of asperities with associated stresses at the surface and subsurface developing to resist the applied forces. This deformation can be elastic or elastic-plastic and depends on many parameters such as, the normal force, shear force, contacting area of the rough bodies and their mechanical properties. These stresses may be considered to be in the elastic phase before the yield point is reached. When the stresses exceed this limit the rough body will deform plastically. Increasing the interaction of asperities will result in larger surface and subsurface stresses. Using Abaqus, a model can be constructed which incorporates elastic and plastic asperity deformation. The technique used in this research is

a plane strain problem. This is appropriate for the case in which the dimension of the problems in one direction is very large as compared with the other two directions.

This chapter describes the modelling of rough surfaces to provide a prediction of residual deformation as a result of changes in the surface asperity features due to plastic deformation. Profiles from real surfaces were used to analyse the behaviour of the measured surface roughness under a load. This part of the study shows the initial phases of modifications that the manufactured surface experience. In general, it seems that from the surface roughness measurements the surface roughness topography experienced a significant degree of residual plastic deformation at asperity tips

3.1.1 Rough surface topography.

Twin disk test rigs are widely used to study lubricated contacts and a rig of this form was used in this study. The disks normally have rough surface characteristics. Such surfaces have irregular asperities that are different in height and width. A typical solid surface includes a range of spatial frequencies as shown schematically in Figure 3.1. The high frequency or short wavelength components are referred to as roughness or primary texture with micro geometric imperfections, which are caused by the machining process such as, cutting, grinding, spark and erosion. The medium-frequency or medium-wavelength component is known as waviness or secondary texture. It is normally caused by vibration, workpiece deflection or undesirable machine tool effects which are always present in manufacturing solid surfaces where the guiding or the cutting machine is not perfect. Finally, in many cases the original profile may also contain significant form errors such as a curve or radius which results from departure from the intended shape of the surfaces.

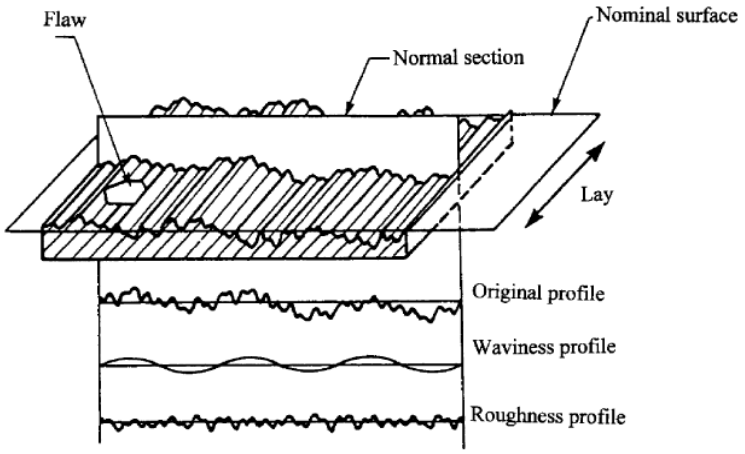


Figure 3-1 Surface topography including (Roughness, Waviness and Form), (Raja.2002).

3.1.2 Surface texture measurements

The most significant step in analysing the contact problem is the surface roughness measurement. This measurement can be accomplished in one of two common techniques which are available today for measuring and analysing surface roughness. Firstly, optical methods are used to provide an optimal picture for the surface topography from which very fine details may be identified. This techniques provide a roughness with a quantitative average representation. Non-contact surface roughness measurements are preferred in some areas because some materials are soft, fragile, and chemically sensitive in direct stylus contact, or a roughness measurement in more than one direction is needed in some engineering applications. Secondly, a stylus profilometer contact method which measures the variation in stylus position when the stylus and surface are in physical contact and the stylus is drawn over the surface. The most important feature of this technique is that it is simple to use, no difficulty arises during the calibration process and it gives an excellent quantitative measurement of surface topography. Moreover, this method gives a direct representation of the surface in terms of a surface profile. Also, the stylus technique is often beneficial in unclean environments where non direct contact methods may result in

measuring surface debris instead of the surface characteristics themselves. However, the surface profile may be distorted to some degree by the stylus tip geometry (Bhushan, 2013), as shown in Figure 3.2. The most common surface profilometers are the Talysurf range of instruments provided by Taylor Hobson. Profilometers are used to measure the roughness, wear and damage of mechanical parts which are in service. The device consists of a diamond-tipped stylus which is moved gradually over the surface and the signal from a special high resolution displacement transducer which is included in the stylus location system measures the amount by which the stylus is raised or lowered at any given point. This is recorded and digitised as asperities and valleys. The work reported in this thesis uses surface roughness measured with a stylus profilometer and the surface roughness profiles before and after running the test disks were obtained. The test disks are ground in the transverse direction. Circumferential roughness profiles were taken by a portable profilometer mounted on the test rig as shown in Figure 3.3.The shape and size of the stylus are normally important elements that affect the data acquisition from the operating surfaces. The stylus tip used has a radius of approximately 2μm and is made from diamond.

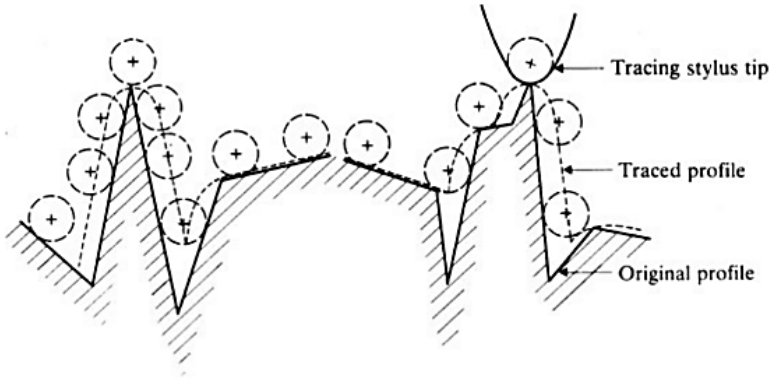


Figure 3-2 Distortion of profile due to stylus tip (Bhushan, 2013).

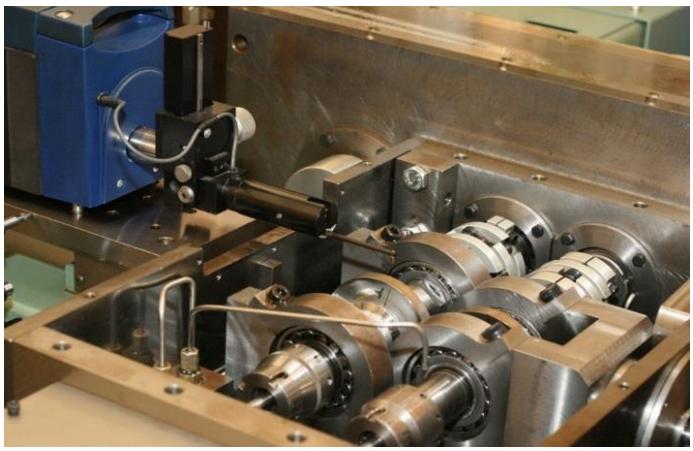


Figure 3-3 Photograph of the profilometer in the back position

3.1.3 Filtering the raw profile

The raw profile as shown in Figure 3.4 was obtained from the profilometer and is representation of the surface topography including profile characteristics other than those that are smaller than the tip radius of the stylus. Some of them may be undesirable after measuring the surface roughness. It is therefore essential to decide which profile should be considered in any assessment of roughness. Analysing the raw profile and removing the unwanted longer wavelength features that include waviness and form, requires high pass filtering of the raw data. The process used to describe the filter phase is called the "cut-off length", thus a filter having a cut-off length values of 0.25mm will normally suppress any irregularities which have a wavelength greater than 0.25mm. The cut-off to be adopted depends on the type of surface being assessed and the size of the contact area. The filtering process creates the mean line of the profile which is removed so that the roughness remains. Filtering is carried out digitally and can include more one type. The type of filtering which is currently used in this research is a Gaussian filter with cut-off length 0.25 mm, which normally has 50% transmission at the cut-off length. The aim of this process is to remove waviness elements and the form of the disks, leaving the roughness profile for evaluation.

Gadelmawla et al., (2002) state that surface roughness parameters play a major role in surface geometry definition and helps to provide a clear picture of the surface topography. The arithmetic average height parameter, Ra, is considered the most significant parameter that is used to describe the deviation of roughness from the mean line as shown in Figure 3.5. The arithmetic average height parameter can be defined in the form of $R_a = \sum|y_i| / N_i$,

where y_i represents the peak heights which are measured from the mean line and N_i total number points of the profile. The unrun roughness profile has an average roughness (Ra) between 0.3 and 0.4 µm.

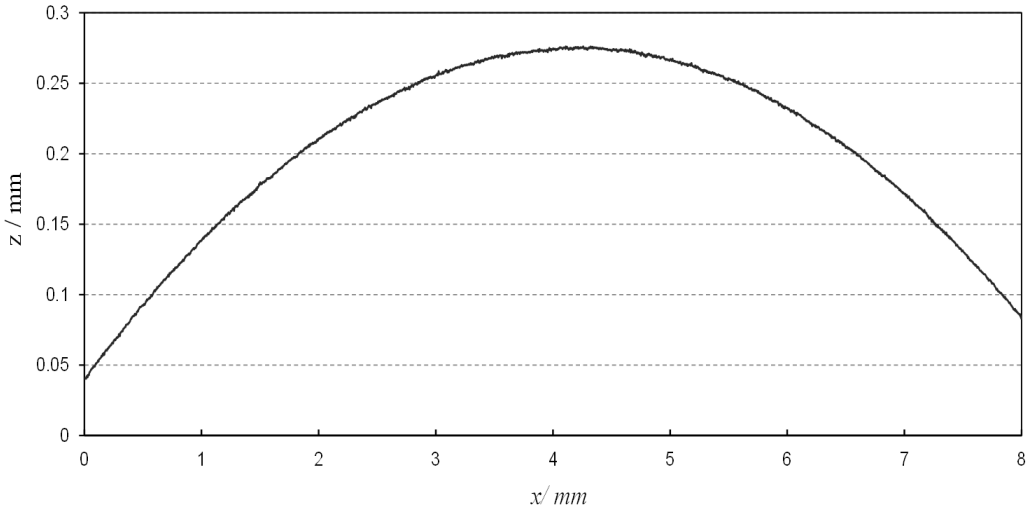


Figure 3-4 Measuring raw profile including waviness and form.

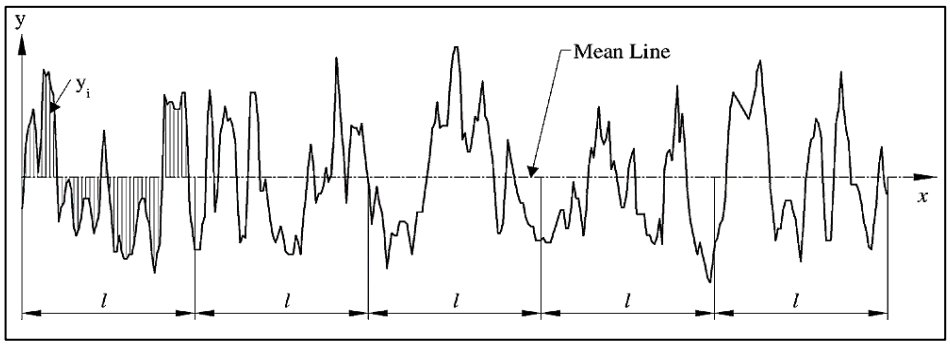


Figure 3-5 Definition of the arithmetic average height (Ra), (Gadelmawla et al., 2002).

3.1.4 Description and extracting profiles of experimental work.

To investigate the asperity contact behaviour it is essential to develop analyses of data from experimental work in order to make a comparison with numerical models by using Abaqus FEM and monitoring the evolution of the surface roughness topography. Thus, Finite Element contact analysis is used to investigate the plastic deformation occurring at surface asperity features in order to quantify the level of residual stresses developed in the material. Twin disk rig roughness profiles were acquired by using a Talysurf surface profilometer prior to and following each loading case during a running-in experiment. The operating conditions assumed in the experimental work is shown in table (1) in chapter 5. Profiles were extracted in the axial direction at four circumferential positions on the disks, where relocation was achieved by engraved lines on the face sides of the disks. These marked locations were lined up with a reference mark in taking measurements before and after running in the test rig, so the approximate circumferential position was known. Such adjustment only provides nominal relocation in the circumferential direction. Precise relocation of profiles taken at different loading stages of the experiments was accomplished by using alignment of the unique deep valley features of the profiles in the trace direction. For initiating a set of profile measurements, the Talysurf unit was set in terms of, run-up length 0.3mm, data length 8mm, spacing 0.5μm, measurement speed 0.25mm/s and a Gaussian filter with a cut off 0.25mm was applied to produce the roughness profile. Profile data from the running in experiment was acquired from the fast and slow test disks for the run and unrun disk at different load stages. These acquired profiles (unrun /as manufactured, run stage2, run stage 18) are shown in Figure 3.6.

The profiles measured at LS2 (load stage 2) have run-in and are stable after three short 30 second loading periods. LS18 (load stage 18) was measure after a number of longer running in periods.

This graph provides a clear picture for comparisons of profiles and shows that a significant amount of residual plastic deformation has occurred at some asperities. It confirms that adjusting the profiles to nominally realign deep valley features in the z direction gives an appropriate and acceptable result for comparisons purposes.

This graph also shows a significant change in the surface features and most asperities in the run profile do experience large plastic deformation at the asperity level .This takes place very rapidly in first stages of loading. In addition, the large asperity peaks have been reduced in amplitude and become rounded as the tests progress with increasing the running time, resulting in a more negative skew in the surface height distribution in comparison with the un-run surface profile. However, most of surface asperities achieved a stable topography following the initial loading stages. Jamari and Schipper, (2007b) discussed the process of changing the asperities topography and they concluded that the plasticity may be initiated either at the surface (asperities) or in the bulk depending on the contact condition.

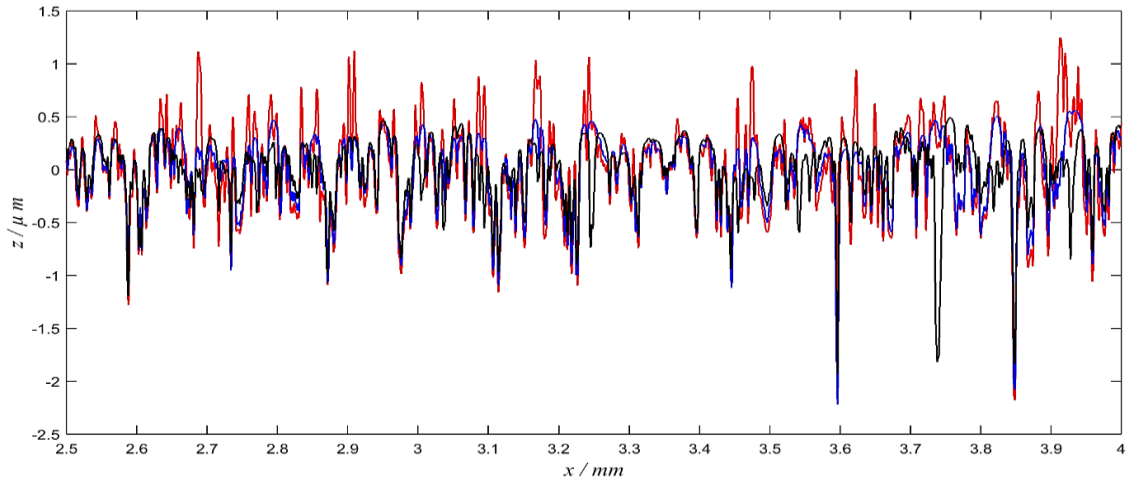


Figure 3-6 Surface roughness profiles taken from ground disk for unrun /as manufactured (red), run/stage 2 (blue), run/ stage 18 (black).

3.2 Creating the finite element model

To make an analysis using a finite element model of real rough surface contact by using the Abaqus program, the unrun (freshly manufactured disk) surface profile is superimposed onto a profile of a smooth roller of radius 38.1 mm. This particular radius was chosen as it equals the actual radius of the test disks to be used later for a comparison between real life measured deformations and those obtained using finite element simulation techniques. These steps were carried out using the procedures described in the following sections.

3.2.1 Rough roller model construction method

The real unrun rough profile is superimposed on a 38 mm radius circle to form a 2*D* deformable part that models, the contacting part of the twin disk rig as shown in Figure 3.7 a, b and c . The un-run rough roller profiles were imported to Abaqus using a Python script that was used subsequently to create a finite element 2*D* plane strain model. The rough profile was centred and aligned, so that the *x*- axis zero was located at the centre of the profile as illustrated in Figure 3.7 a. This profile was then used as input data for developing a macro to form a 2*D* deformable part as described in the next section.

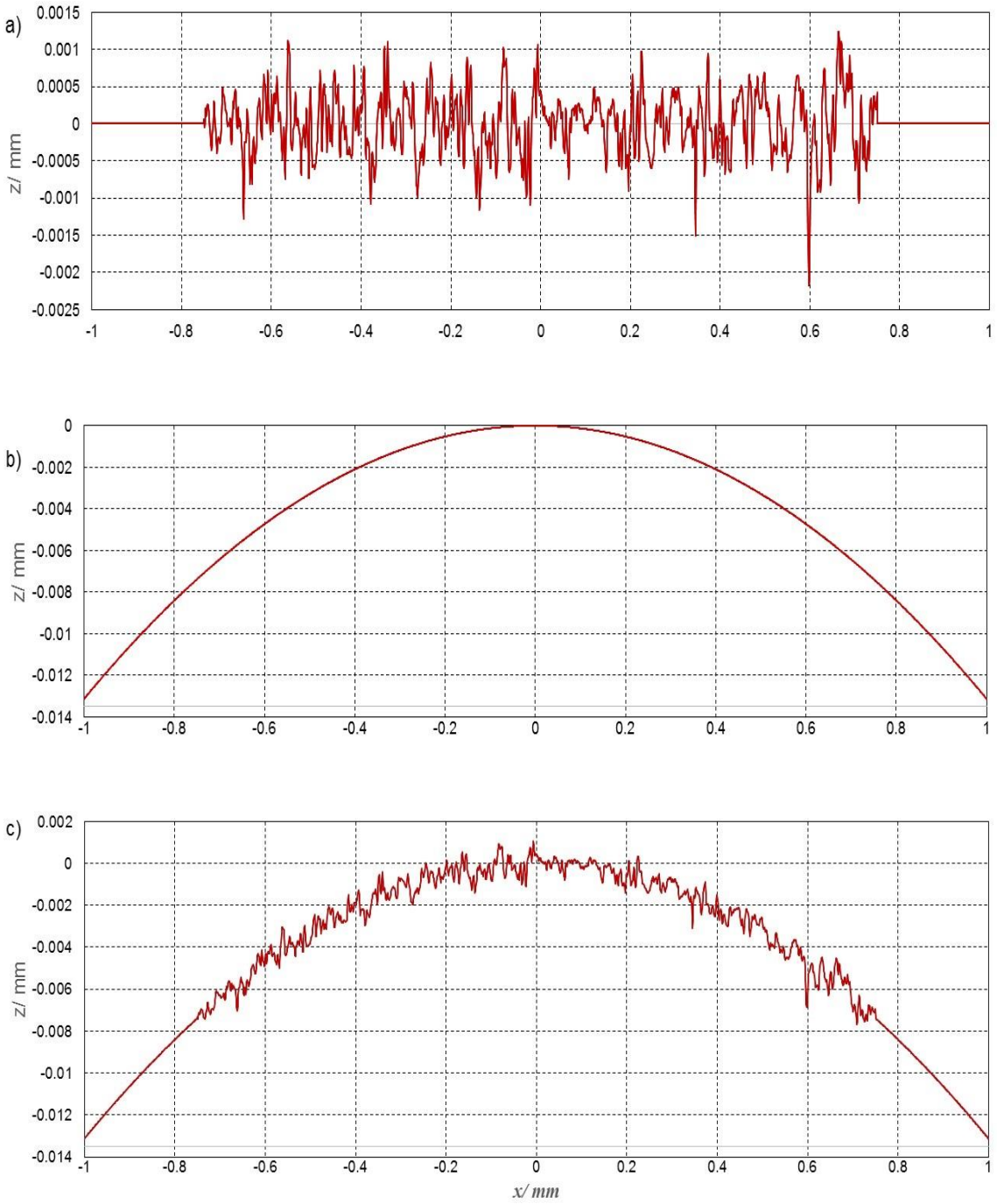


Figure 3-7 Rough roller model construction method; a) Unrun surface roughness profile, b)smooth roller profile, c) rough roller profile.

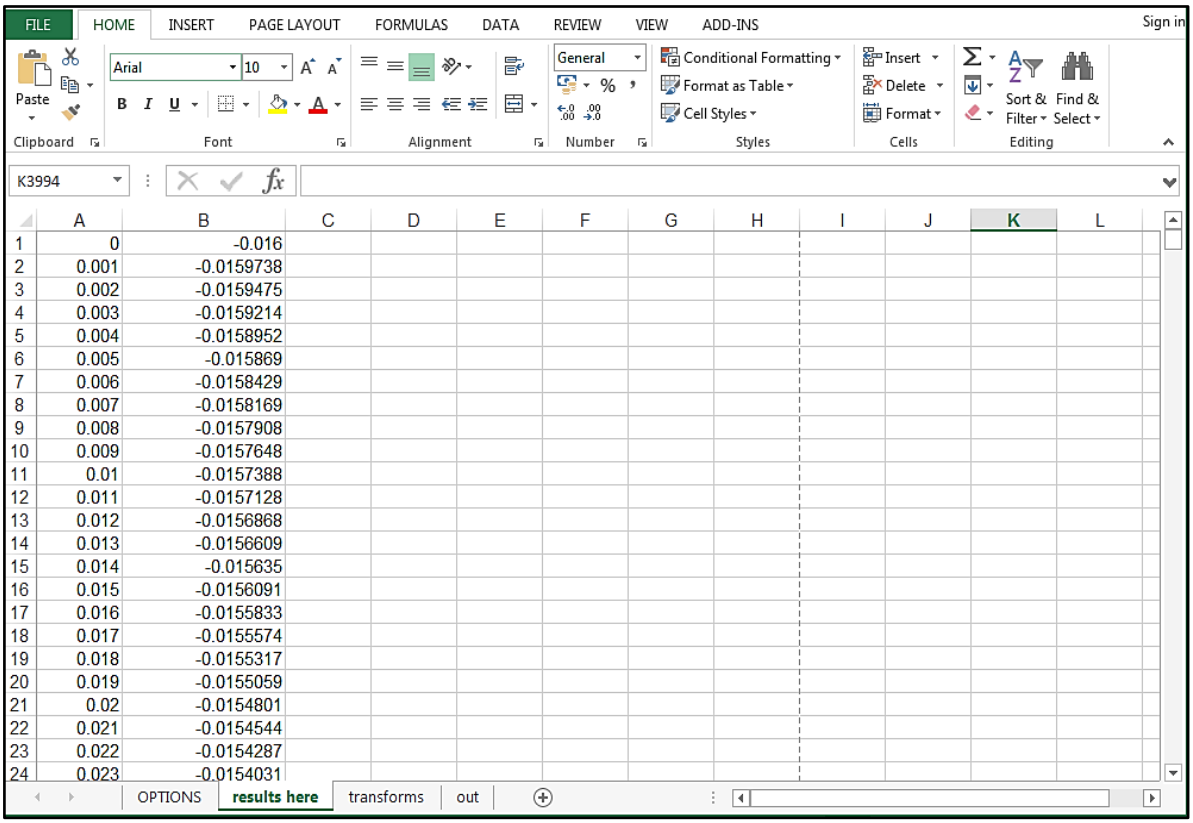
3.2.2 Importing profiles into Abaqus/CAE and creating a two-dimensional model

To simulate the contact of real rough surfaces, measured rough surface profiles must be imported to Abaqus. The un-run rough roller profile was imported to Abaqus using a Python script, and used to build 2*D* deformable part. To provide a significant powerful way to solve this problem, a macro was used in Microsoft Excel to create a Python script that would create a finite element 2*D* plane strain model in Abaqus. If the modifications are required, the user can then implement them on the part as necessary e.g. (creating sections, defining materials, etc.…) and proceed with creating a FE model in Abaqus/CAE in the same way as with any other part.

3.2.3 Python and the Abaqus scripting interface

Previous work by researchers at Cardiff tribology group has developed a python script file by using an Excel workbook containing embedded macros to write the Abaqus python script file to create a part having the measured roughness. In the current study, this script has been employed to solve the problem of micro-EHL. This workbook gives the user the ability to specify the rough surface profiles data, and the model part dimensions, while the base python file includes the required code that is essential for every part creation. The excel spreadsheet contain four sheets: “options”, “results here”, “transform” and “out”. Figure 3.8 shows part of the “results” sheet which contains the required roughness profile for the part. The details of the part to be created in Abaqus are entered in sheet options shown in Figure 3.9. The cells that must be completed by the user in that sheet can be defined as results from, excess depth and are named and highlighted, e.g. “Excess depth” and “resolution”. The x coordinate of the start and end of the rough profile required in Abaqus must be specified in C22 and E22 respectively. The transform sheet contains the Python scripts to transform the specified rough surface profile and part dimensions into the required coordinates for Abaqus. To create the model part in Abaqus running a script is required which is achieved by clicking “run

script" from the start up screen or running a script from the command line interface. The 2D deformable rough roller part is then created and ready for further steps in the analysis.



	A	B
1	0	-0.016
2	0.001	-0.0159738
3	0.002	-0.0159475
4	0.003	-0.0159214
5	0.004	-0.0158952
6	0.005	-0.015869
7	0.006	-0.0158429
8	0.007	-0.0158169
9	0.008	-0.0157908
10	0.009	-0.0157648
11	0.01	-0.0157388
12	0.011	-0.0157128
13	0.012	-0.0156868
14	0.013	-0.0156609
15	0.014	-0.015635
16	0.015	-0.0156091
17	0.016	-0.0155833
18	0.017	-0.0155574
19	0.018	-0.0155317
20	0.019	-0.0155059
21	0.02	-0.0154801
22	0.021	-0.0154544
23	0.022	-0.0154287
24	0.023	-0.0154031

Figure 3-8 Roughness profile sheet in the Excel workbook, containing embedded macros and rough surface profile.

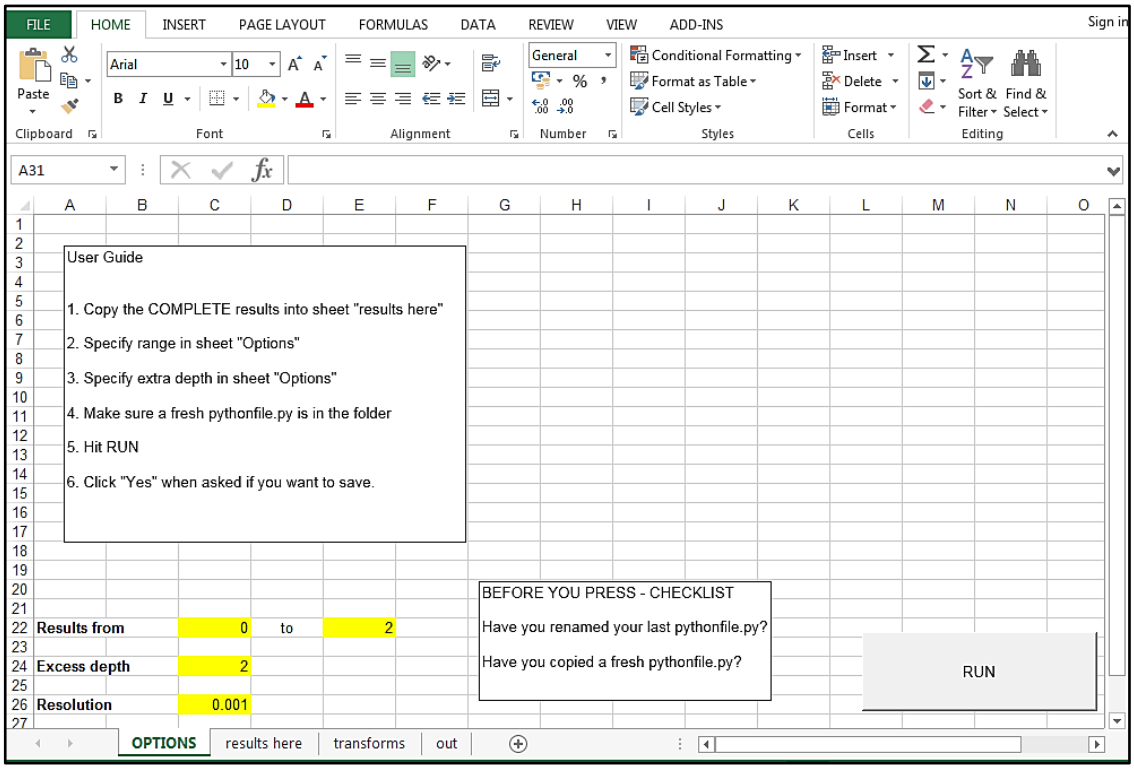


Figure 3-9 the option selection sheet from the Excel workbook.

The final model contains three separate parts that form the part of the physical disk to be analysed. The central part has the rough surface profile whereas its two neighbours, which do not make contact with the rigid plane under load, have smooth surfaces. Also, a single rectangular part which is located underneath the central part was created and set up to give greater depth as necessary to make the assembled part large enough to behave as a semi-infinite body. The group assembly function was used to create one single part as shown in Figure 3.10 that is then loaded in plane strain against a rigid plane. This is equivalent to loading the disk against a mirror image of itself. Figure 3.11 shows the portion indicated A in Figure 3.10 at a higher vertical resolution and Figure 3.12 shows the surface of portion A in the form that will be used for illustration of the loaded results.

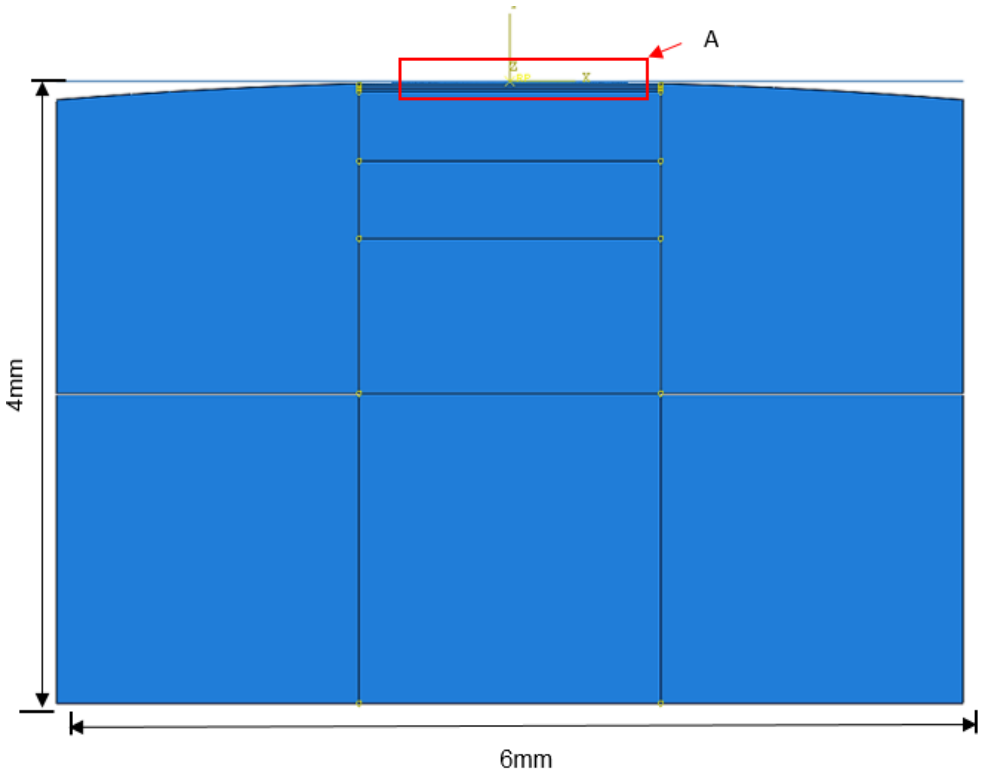


Figure 3-10 the rough surface contact model after the assembly.

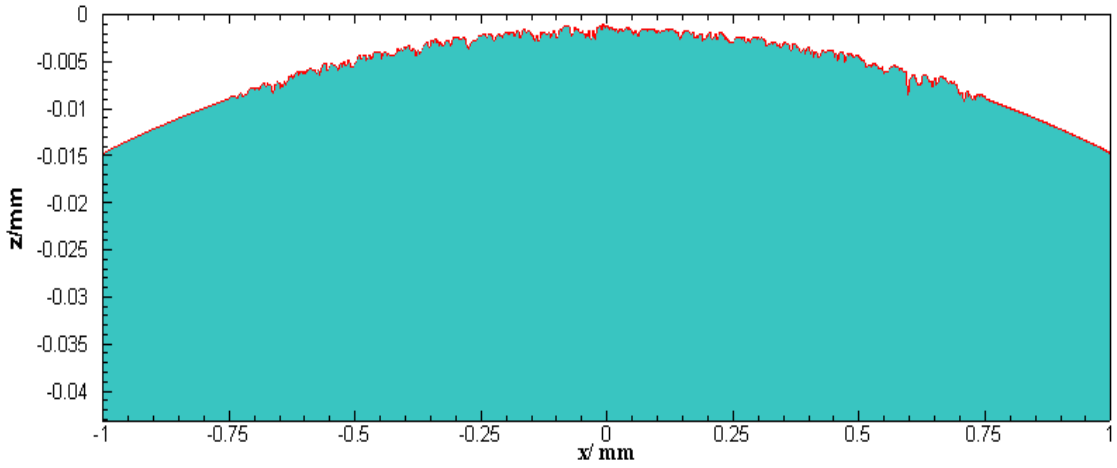


Figure 3-11 a large scale of section A of Figure (3-10)-red colour is 2 mm roughness profile.

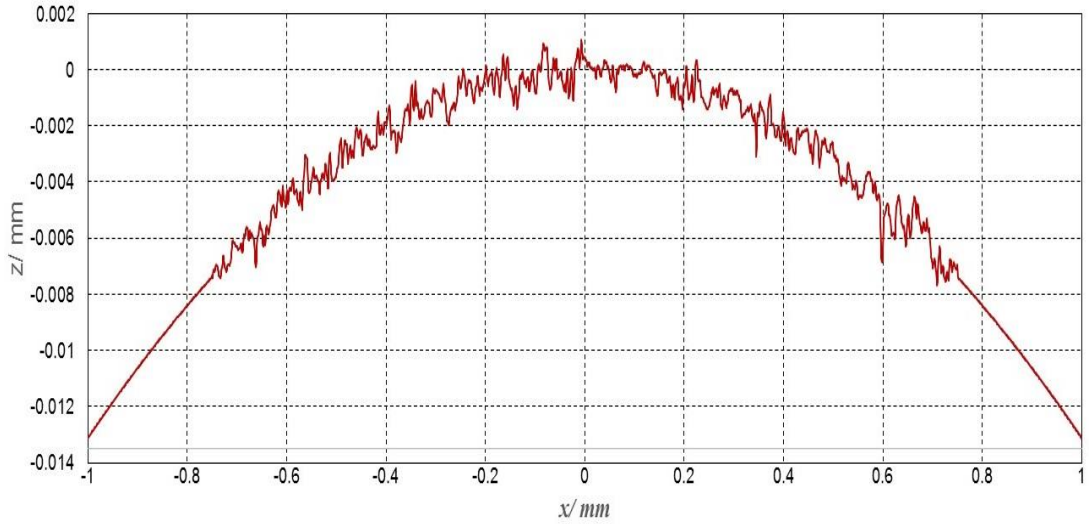


Figure 3-12 Unrun rough roller profile 2 mm long from Abaqus.

3.3 Determination of residual profile from Experiment and comparison with Abaqus finite element contact analysis

3.3.1 Introduction

In order to compare and correlate the amount of plastic residual deformation that asperities have experienced in the experimental work with the finite element analysis, the surface roughness profiles before and after loading were required. Thus, it is crucial to relocate the profiles such that the surface roughness is aligned in order to examine the level of plastic deformation. Those profiles, can be seen in Figure 3.13, where the black and blue profile in graph represent the fully loaded step and load removal step in the Abaqus analysis, respectively. The surface profile measurement before and after loading are shown in the Figure 3.14.

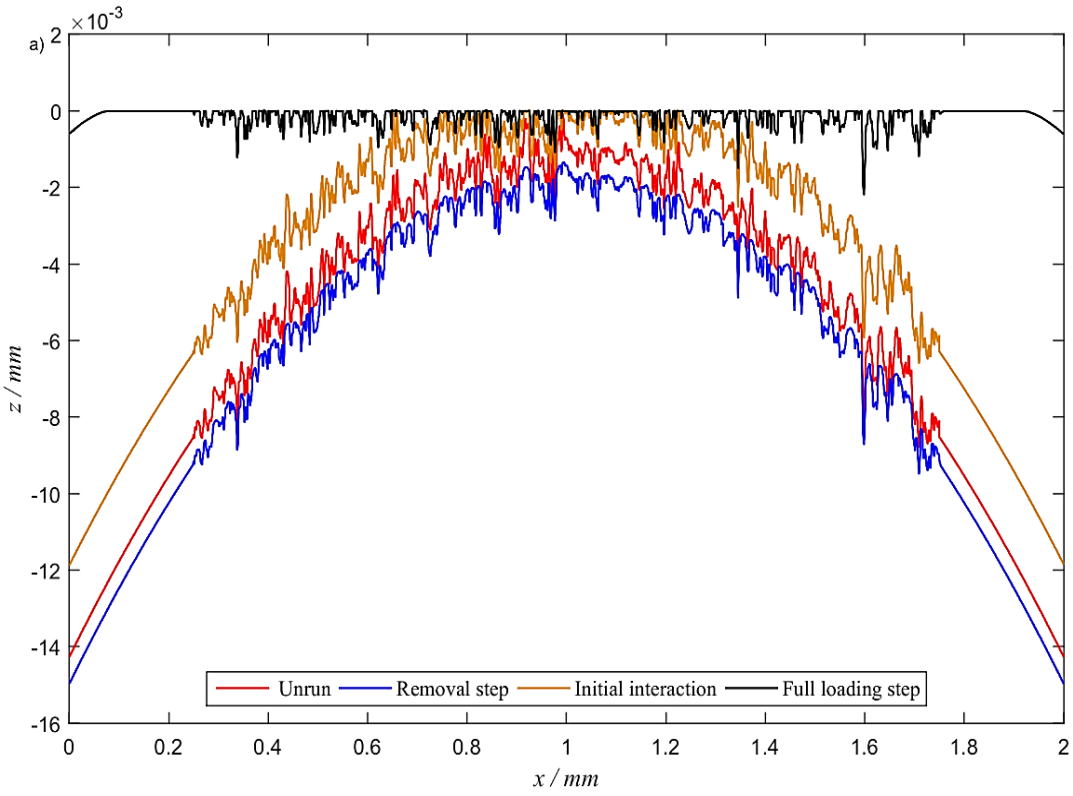


Figure 3-13 Rough roller at various stages of the elastic/plastic contact analysis

As has been previously discussed, only asperity peaks are affected significantly by the plastic deformation and this residual deformation does not extend to the valley features. For that reason the deep valley features in the pre and post running profile are used as standard reference points for realignment and comparison of the profiles in the trace direction. The level of residual deflection that occurred after the running in process was determined by comparing the profiles to quantify the proper level of residual stress to which the deformed asperities would be subjected. Therefore, comparison between the plastic residual deformations associated with the pre and post running profiles for the different applied loads extracted from the real surface and from the Abaqus simulation required a series of processes and a special technique for comparison. The following are the steps in the procedure for accomplishing an accurate comparison and evaluation of the pre and post running profiles in the running-in process.

3.3.2 Realignment of experimental profiles in the trace direction

Results from both experimentation and simulation profiles have clearly medicated plastic deformation at the asperity level. This is seen in Figure 3.14 which shows pre and post running profile results from the running-in process during one of the experiments.

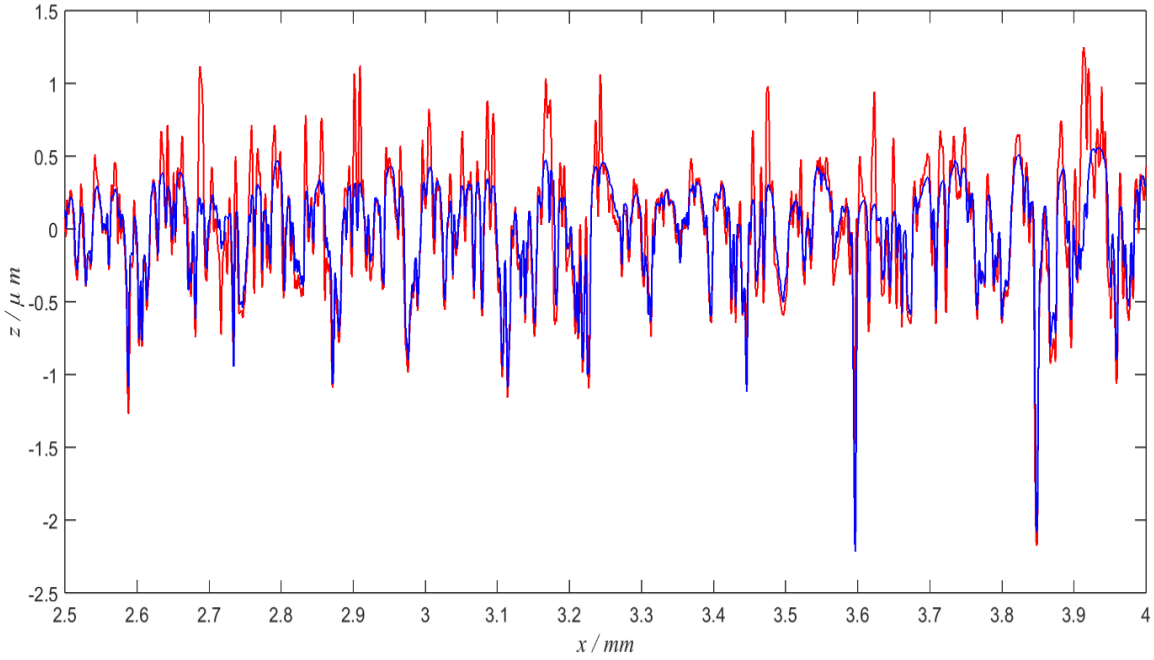


Figure 3-14 Profile for unrun (as manufactured) in red and following run in blue.

As noted above, the adjustment in the trace direction was achieved by using recognizable deep valley features located between asperity features. It can be seen in this figure that, the local valley features are relatively unaffected by the contact loading. However, there are some valley features that have been shifted in the vertical direction (z). For example, those between x = 2.55 mm and x = 2.6 mm, x = 2.95 and x = 3 mm, x = 3.7 mm and x = 3.75 mm. This tendency for deep valley features to rise could be due to the Gaussian filtering process, where the Talysurf software which was used would output data relative to the mean line which was calculated for each surface measurement. Moreover, it has been claimed by Pullen and Williamson, (1972) that a trend for valley features to rise may be due to other factors such as, a combination of material conservation; the surface material which is

deformed plastically being redistributed to regions not in considerable contact. Also it seems likely that the relative sizes between the profilometer stylus and the deep valley features may have a significant influence on the measured valley topography (Gohar and Rahnejat, 2008) as shown in Figure 3.15.

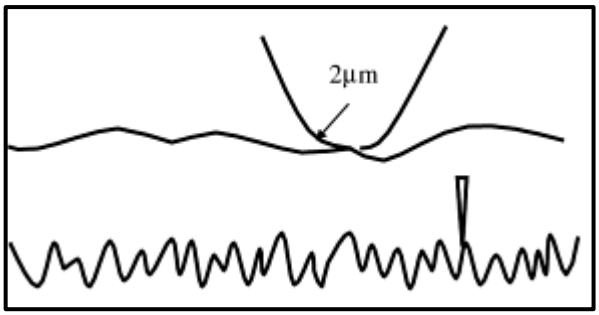


Figure 3-15 Profilometer stylus size effect, (Gohar and Rahnejat, 2008).

So the trace direction adjustment achieved by using identifiable local valley features that separate the significant surface asperity features was not enough to determine the level of residual deflection and the residual stress field that should be extracted for analysis. A further vertical adjustment was required to correlate and compare the profiles to give the actual level of residual profile shape and residual stress associated with asperity features.

The next section will highlight the necessity of adopting a new technique for comparison in order to accurately quantify the level of residual deformation experienced by the asperities.

3.3.3 Connection lines technique to relocate the deepest valley features vertically

This method relies on identifying the deepest valley feature for the pre and post running profiles then connecting those significant valley bottoms with a curve, as shown in Figure 3.16 a ,b and c. It can be clearly seen from the graph 3.16 c that, there is a noticeable fluctuation in the relative valley bottom heights between the run and unrun profile for some valleys. Factors contributing to this are variation in the sub surface valley plastic deformation, differences in the mean lines of the profiles obtained by Gaussian filtering, and

finally surface impurities such as, dissolved solids, iron or other metallic particles that have been retained on the disk surface during measurement process despite the cleaning process used.

The connection lines for the deepest valley bottom points for all profiles (run, unrun and Abaqus residual deformation profiles) were determined with the same discrete data points with a 0.5 μm resolution using the MATLAB simulation fitting tool. This, tool provides an interpolation type called shape-preserving interpolants. Thus, when the user selects 'shape-preserving interpolants' for fitting to data, Matlab uses a piecewise cubic Hermite interpolation for the fitting process. Therefore, in the resulting fitted data, each pair of consecutive data points is connected by a different cubic polynomial. The relative difference between the deepest valley connection curve line of the unrun profile and the other profiles were thus determined independently. This allowed the unrun deep valleys to be considered as benchmarks and the other profiles to be realigned relative to those valleys. Figure 3.17 illustrates the connection line curve and the difference between them for the run and unrun profile. This difference has then been added to the post running profile to achieve the final position for the realignment process, as shown in Figure 3.18. Comparing Figures 3.18 and 3.14 shows the effectiveness of this new approach.

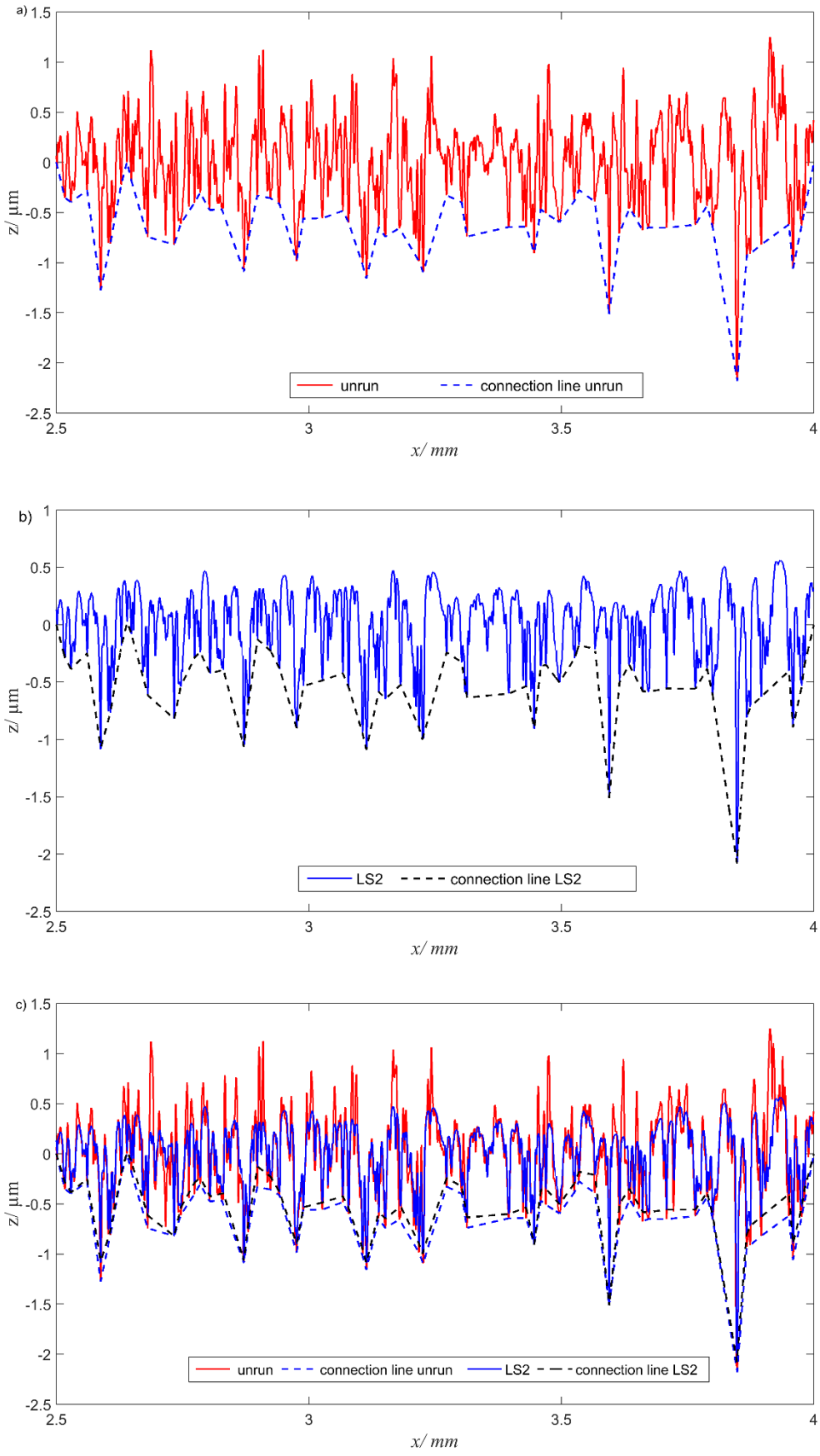


Figure 3-16 Connection line curves concerning the deepest valleys; a) unrun- as manufactured, b) following running and c) unrun , run profile and connection line curves.

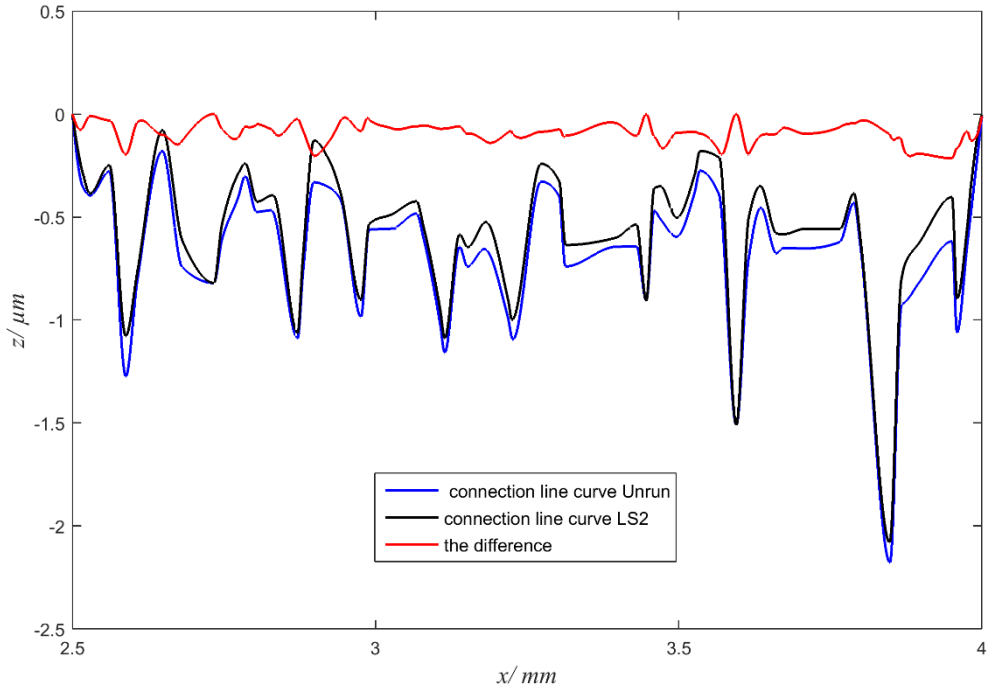


Figure 3-17 Connection line curves and their relative height difference

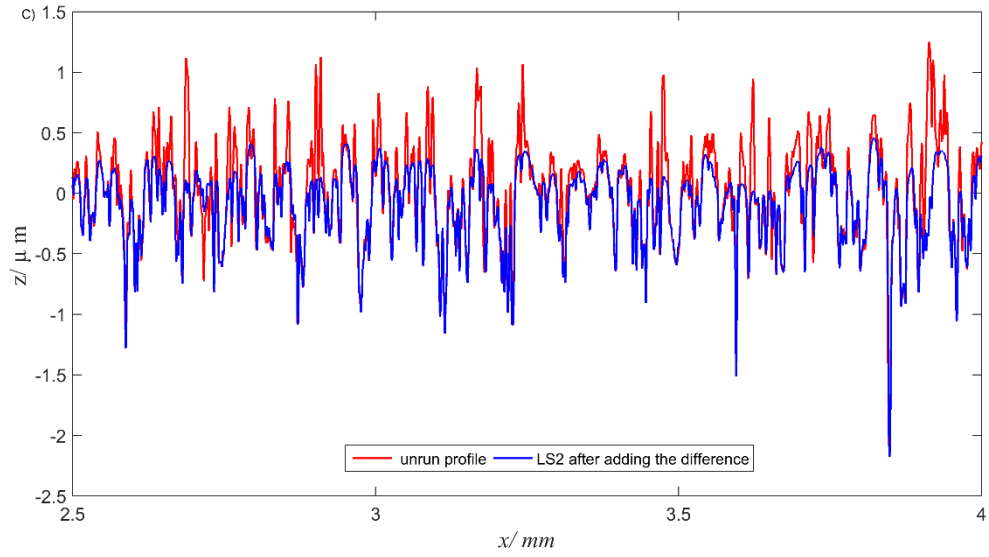


Figure 3-18 Final alignment for the unrun and run profile by adding the corralling line curve difference to the running profile.

3.3.4 Abaqus and experiment profiles alignment

The technique of using the connection line curves for vertical realignment was also used for the residual surface profiles obtained from the Abaqus analysis at the load removal stage. This was done for each of the sequence contact loads The loads used for the FEA analysis are referred to in terms of the distributed pressure of 500 MPa, 750 MPa, 1250 MPa, and 1500 MPa applied to the 6 mm length of the lower boundary the FEA model. Figure 3-19 shows the deep valley features remain undisturbed throughout the loading process. Moreover, it is interesting to note that , some valleys do not align accurately with each other, this is because some judgement needs to be made about choosing the appropriate valleys for the correction process, for example, those which are located approximately between x = 2.87-2.89 mm , x = 3.35-3.36 mm and x = 3.74-3.76 mm. This issue can be observed in Figure 3.20 which shows more details at a larger profile direction scale.

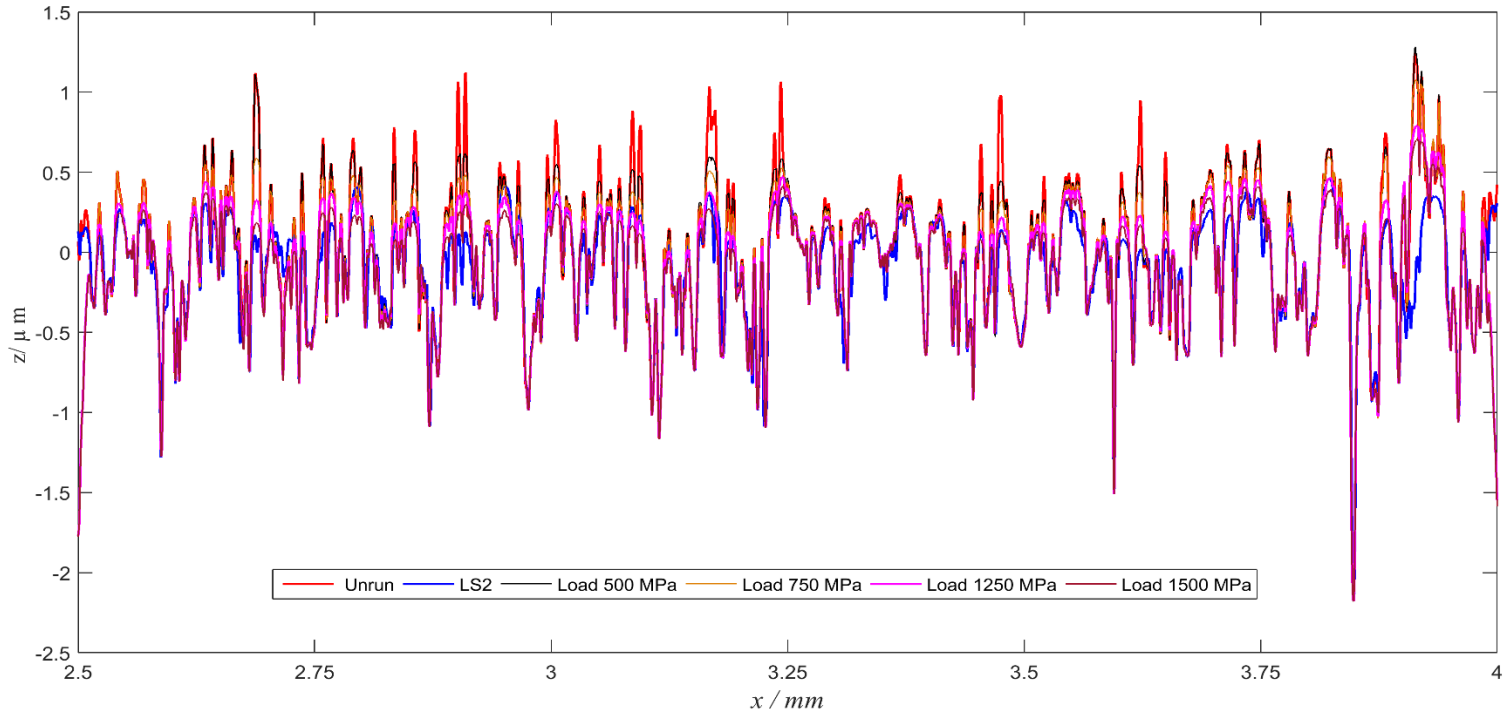


Figure 3 -19 Final alignment of residual shape profiles for the range of applied loads. Also shown are the unrun and load stage 2 experimental profiles

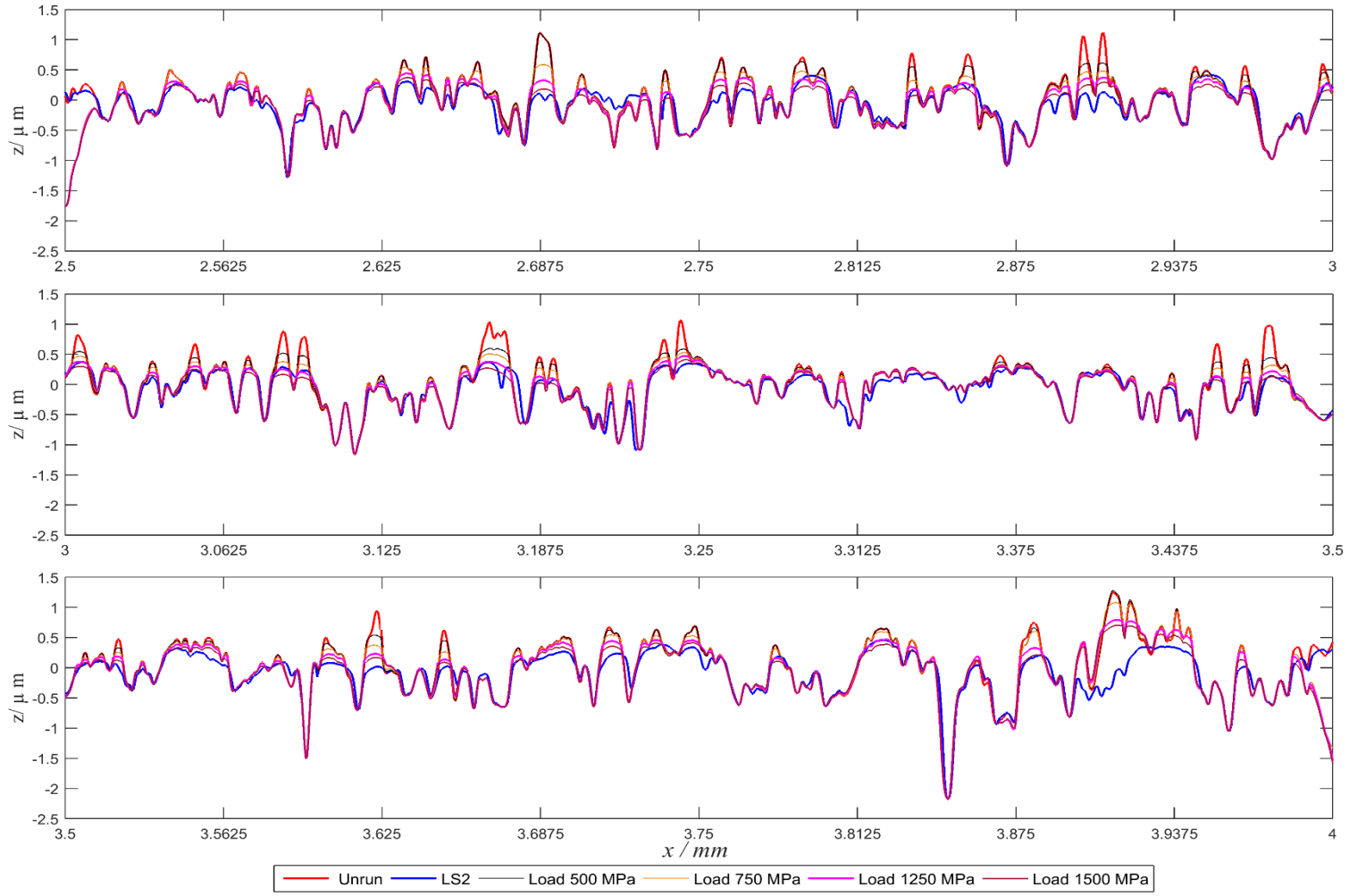


Figure 3-20 Final alignment of residual shape profiles of Figure (3-19) at a larger 3 sections profiles direction scale.

3.4 Modelling details of finite element analysis package used for the contacting problems.

3.4.1 Introduction

This section describes the methodology used for elastic-plastic contact modelling with the Abaqus finite element software. Generally, for finite element analysis contact conditions are considered a special parameter of discontinuous constraint behaviour, allowing nodal forces to be transmitted from one model part to another. This constraint has a discontinuous situation because it is utilized only when the two part surfaces come in to direct contact. Technically speaking, when the two model parts are separate, no such constraint is required. The contact analysis must therefore detect when and where two surfaces are in contact and apply the contact constraints accordingly.

This study highlights the contact problems in two dimensional, line contacts. This approach for modelling contact problems is considered appropriate for ground test gears where the surface topography is normally consistent with a two dimensional, plane strain assumption. This chapter describes modelling of a dry contact problem between a rigid plane and rough elastic- plastic roller. This is equivalent to loading the deformable rough body against a second deformable body that is the mirror image of the first.

The basic idea behind using a rigid part rather than a second deformable body is that it halves the number of degrees of freedom involved in the analysis. It is also an effective way of loading the asperities as each prominent asperity makes direct contact with an identical asperity on the counterface.

3.4.2 The Abaqus CAE software package / 6.12V

A full Abaqus analysis normally includes three prominent phases, and these are classified as: pre-processing which is an interactive CAE activity, simulation which involves using Abaqus/Standard and Abaqus/Explicit and post- processing where the finite element analysis results are visualised using the Abaqus viewer tool.

3.4.3 Pre-processing stage

This stage can be defined as an interactive pre-processor that is used to establish the finite element model. In addition, the model of the physical contact problem is created with the associated input file. This can be achieved by using the graphical tools available with CAE package. Also an Abaqus input file analysis can be created easily using a text editor directly. In this stage the discretised part geometry is defined, mechanical properties, boundary conditions, type of analysis and element type are specified.

3.4.4 Simulation with Abaqus/standard

This is the stage in which Abaqus/Standard or Abaqus/Explicit solves the finite element problem which is defined in the model. This is a nonlinear problem due to the unknown contact area and elastic plastic elements and is therefore an iterative process. The output data are stored in binary files in preparation for post processing. The simulation time depends on many parameters such as, the complexity of the contact problem being solved, the computer System requirements, and the computer power consumption. For the reasons outlined previously the running time may take from minutes, to several hours to complete the analysis specified in the input file.

3.4.5 Post- processing (visualizing the finite element analysis result)

Visualizing of the results generally occurs interactively using Abaqus/Viewer which provides a graphical display of the analysis. At this stage the fundamental variables can be specified for presentation such as, displacements, stresses and forces. There are a variety of

options for displaying the results in Abaqus/Viewer that include animation tools, deformed and undeformed shape plots, contour plots, material orientations and harmonic animation.

3.5 Abaqus /CAE working environment /6.12 V

Abaqus/CAE provides a complete modelling and visualization environment for finite element analysis. Users can efficiently manage and create, edit, monitor, diagnose, and visualize the result. Abaqus/CAE can be divided into modules, where each module characterizes a logical aspect of the modelling technique.

Figure 3.21 shows the graphical user interface of Abaqus/CAE version 6.12V. Abaqus has a number of methods and techniques in which modelling and presentational tools can be utilised. The majority of the basic modelling options can be achieved by using the buttons directly alongside the modelling viewport. There are sets of modelling tools available within Abaqus/CAE, and these tools and their key features are described and discussed in more detail below:

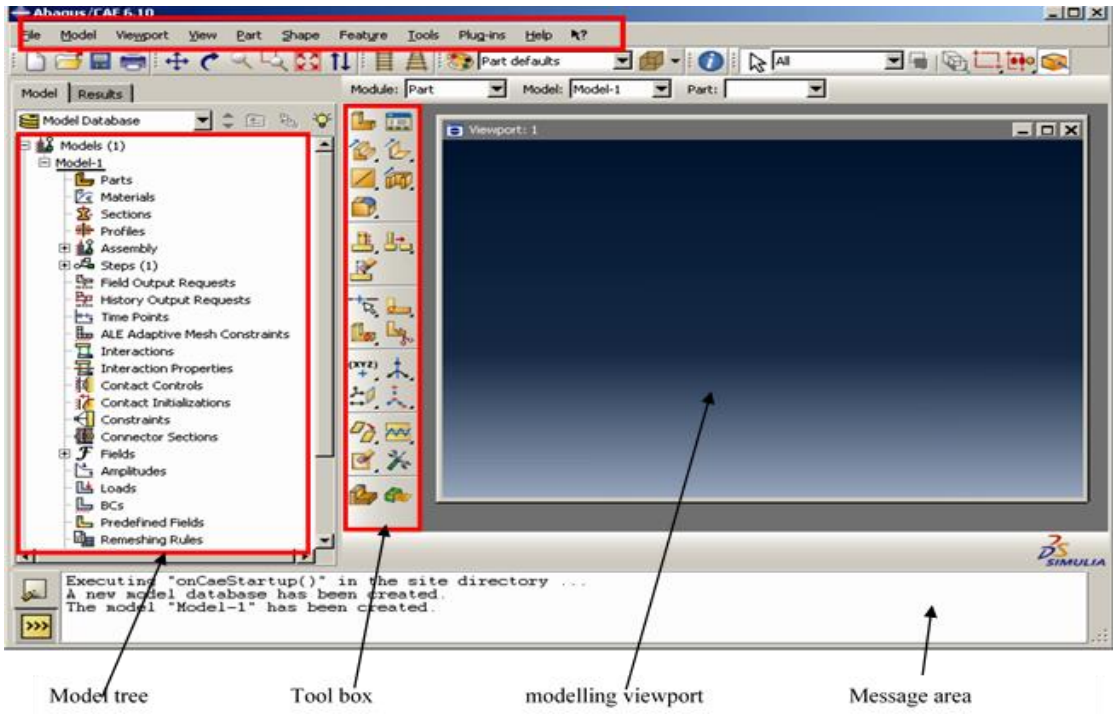


Figure 3 21 Graphical user interface of Abaqus/CAE.

***The part*:** This model is used to define and mange geometry of parts and for creating the model drawing by using simple CAD functions. It can also make partitions by splitting the created parts into regions. This tool allows the user to choose or create either a dependent part instance or an independent part instance, as illustrated in Figure 3.22.

***Property*:** The Abaqus user can use and specify the property model to achieve the following tasks: materials definition, assign sections, orientation of parts. Each part or model must have a section assigned to it. The material editor is used to specify all the information that defines each part, as shown in Figure 3.23.

***The step module*:** In this step parameters can be altered, and the user can control and define a sequence of analysis steps or changes in the way parts of the model interact with other parts. In addition the user can choose which outputs are required in each analysis step of the simulation from the field output request tool, the area of the model from which they will be output can be specified as well. This module is also used to define adaptive mesh regions and specify controls for adaptive meshing in those regions that experience significant element distortion during analysis.

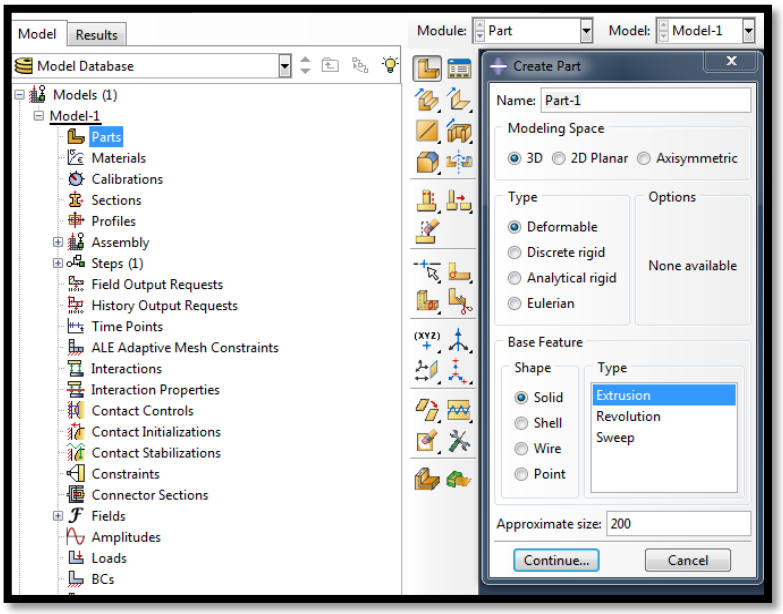


Figure 3-22 Abaqus/CAE Part module.

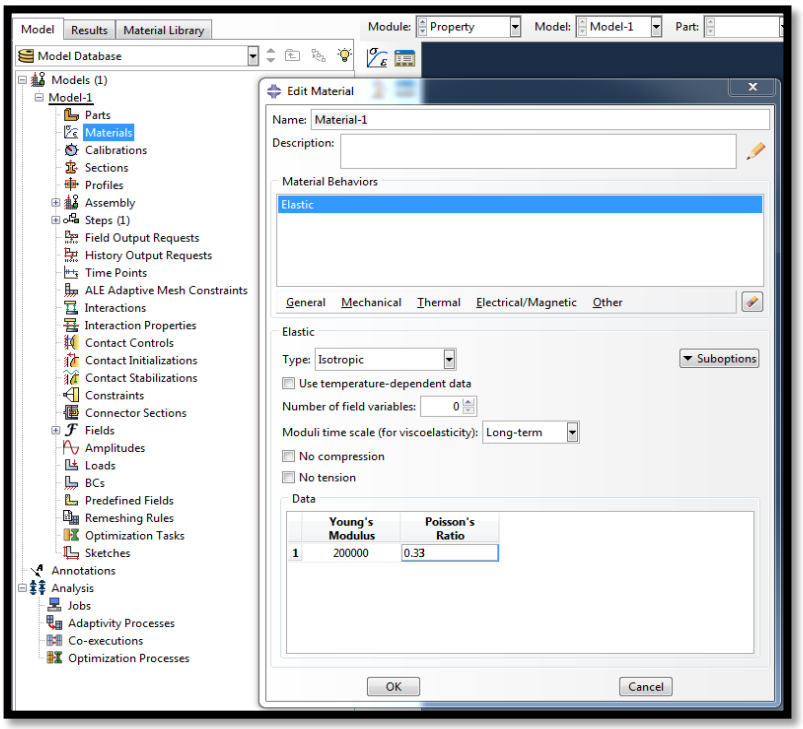


Figure 3-23 Abaqus/CAE Property module.

***Interaction*:** This module is used to define and control interactions between different parts in the assembly or between a specific region of a model and its surroundings, by selecting possibly interacting surfaces and defining the approach in terms of how the Abaqus package will deal with them. That is necessary because Abaqus/CAE does not have the ability to recognize mechanical contact between a part model and other regions. Abaqus includes the use of more than one mechanism of analysing and dealing with interactions, and these methods are classified as surface-to-surface contact, self-contact, general contact.

***Load module*:** This module defines the type of loading and application region, creating and modifying prescribed boundary conditions for the problem. There are many types of load cases, loads, predefined fields and boundary conditions that can be applied to model parts. The user has to specify in which analysis steps loads and boundary conditions are activated.

***Mesh*:** This module includes tools that allow the user to generate and control the finite element meshes on parts created within Abaqus/CAE by specifying mesh characteristics and mesh density. Furthermore, the mesh tool allows the user to select an element shape and meshing approach, create a mesh and seed a part instance that meets the requirements of the problem be analysed. The mesh module utilises a variety of finite element techniques to evaluate mesh quality based on element shape and to check for numerical analysis problems.

***Job*:** When the user has completed all of the task requirements including defining a model for analysis, defining the geometry of the model, applying load, assigning material properties, and defining the type of contact, the Job module is used to submit and analyse a model of the problem. In addition, this module is beneficial in terms of managing the analysis and viewing a basic plot of the analysis results, checking and monitoring progress of the execution of the analysis job.

***Visualization*:** This module provides graphical display for the results of a job analysis, in particular, contour plot on the deformed shape, and a symbol plot. The other benefits of this module are its ability to control and specify what information is required to be placed in the output database, scale factor animation and history or *x-y* data.

3.6 Conclusion

In this chapter the basic principles of the connecting line curves technique to relocate the deepest valleys features was introduced. This approximation is based on the concept of a smooth alignment curve that passes through deep valley bottom positions that are only slightly affected by the plastic deformation happening at the asperity tips, whereas the tips of prominent asperities are deformed plastically in the first few cycles of loading. The main finding is that the difference in connecting line curves for the profiles is clear and allows vertical realignment to be achieved so that the plastic deformation behaviour at the contact surface can be evaluated easily. The existence of a small scatter in the profile difference might be due to the noise during roughness measurement. This approach of profiles alignment allows the level of plastic deformation at significant asperity features to be observed and quantified correctly corresponding to the asperity shape changes evaluated by experimental measurements. There are some instances where profile features appear not to align. Reasons for inconsistency between roughness profiles could be due to misalignment in the axial direction when using the profilometer between tests. Alignment can still be made between the most noticeable asperity features, but as the surface profiles are not perfectly axially extruded there is inevitably some variation in the axial direction.

Chapter 4 Basic theories of fatigue: A review

4.1 Introduction

Contact fatigue is a common mechanism for failure of the surfaces in an EHL system, where the components of machines are frequently subjected to repeated loads, resulting in cyclic stresses that can lead to microscopic damage to the contacting surfaces. Contact fatigue differs from classic structural fatigue (bending or torsional) in that it results from a contact or Hertzian stress state. Conventional micropitting is considered as a major cause of failure in gears and is associated with roughness effects on the scale of surface asperities. Surface roughness is decreased through reduction of both asperity heights and asperity sharpness through pitting and local plastic deformation. Therefore, when roughness is present, it is necessary to assess the fatigue life quantitatively using a fatigue model, which takes account of the true surface loading.

The first section of this chapter presents an introduction to some of the principles of fatigue. The basic elements of fatigue are then introduced, such as mean stress effects, multiaxial fatigue and variable amplitude fatigue. Finally, some theories of fatigue and a review of some of the models that have been developed for rolling contact fatigue will be presented. The numerical evaluation of fatigue performance for real rough surfaces in lubricated rolling/ sliding contact will be reported in Chapters 5 and 6.

4.2 Principles of fatigue

4.2.1 Fatigue life models

Fatigue failure is initiated at local stress raisers within the material, such as impurities, inclusions, micro-cracks or dislocations in the crystal structure. Under cyclic or fluctuating loading, the local stress exceeds the yield stress of the material which then causes progressive failure associated with micropitting. These micropits have a diameter in the order of microns or tens of microns (Oila and Bull, 2005). There are three stages of fatigue crack: initiation, propagation, and final fracture (Verdu et al., 2008). The fatigue crack develops below the surface at the asperity level, until a region of metal is separated to some extent from the base metal by the crack growth. This ultimately results in the loss of pieces of metal from the contact surface (Bhushan, 2013). Hence, any cracks that initiate at the asperity surface as a result of tensile stresses can develop and lead to surface contact fatigue. The conditions for the nucleation of rolling contact fatigue and the rate of advance of the dominant fatigue crack are strongly affected by a wide range of mechanical, microstructural and environmental factors. Furthermore, there are many important factors which affect fatigue life, such as the type of the applied loading (bending or torsion), loading mode (constant or variable amplitude loading, surface topography (rough, smooth), stress concentration, residual stress, corrosion, mean stress and load (Loewenthal, 1984). It has been assumed that a thorough understanding of the factors that affect the fatigue life of components will allow for a more effective prevention of fatigue failure and this will enhance our understanding of micropitting nucleation. Consequently, many researchers have paid considerable attention to the rolling contact fatigue (RCF) that occurs in a variety of components subjected to rolling contacts, such as rolling-contact bearings, gears, cam, and rail wheel contacts in railways. RCF has been described as: *"*The mechanism of crack propagation caused by the

near-surface alternating stress field within the rolling-contact bodies, which eventually leads to material removal*"* (Ahmed, 2002).

The first analytical model to evaluate the RCF problem was introduced by Keer and Bryant, (1983). In this model, two-dimensional fracture mechanics were used to develop a fatigue life model for rolling/sliding Hertzian contacts. The authors assumed the crack initiation life to be small in comparison to the fatigue crack propagation life. In addition, their approach is based on assuming that stresses having a Hertzian distribution travel through the contact half-width such that one passage corresponds to one cycle of rolling. Vincent et al., (1998) also investigated the effect of crack initiation based on the dislocation pileup idea. They included in their model the effect of all of the stress components as well as residual stress effects. Crack initiation, crack propagation, and micropitting was examined by Zhou et al., (1989) using a two-disk rig in rolling and sliding contacts. The results of this paper have shown how surface cracks can initiate and propagate with increasing contact pressure along the whole contact area.

The rolling contact fatigue mechanism is affected by a large number of factors and many studies have been conducted to assess the roles of these parameters. For example, Oila and Bull, (2005) have carried out experiments using a twin disk machine to assess the impact of several factors, such as material, surface finish, lubricant, load, speed and slide-to-roll ratio on micropitting initiation. It was found that the load has a significant influence on the initiation of micropitting whereas the slide-to-roll ratio has a significant effect on micropitting propagation. Harris and Yu, (1999) investigated load ratings and fatigue life prediction of rolling bearings based on the Lundberg-Palmgren theory. They found that there was a significant positive correlation between the surface shear stresses and fatigue lives. The inclusion of surface traction along with Hertzian normal pressure was found to significantly increase subsurface octahedral shear stress.

The influence of the lambda ratio, which can be defined by $\frac{h_{min}}{\sqrt{\sigma_1^2+\sigma_2^2}}$, where h_{min} is the reference film thickness calculated using the Dowson and Higginson formula for smooth surfaces and σ_1 , σ_1 are the root mean square of roughness of the two surfaces in contact, was investigated by Sharif et al., (2012) in rolling fatigue contacts, where it is considered as a significant factor in determining damage performance. It is believed, based on results from this simulation, that increasing the lambda ratio significantly reduces the level of damage for the prominent asperity features, where higher lambda ratios results in less asperity contact. An additional highly significant factor in evaluating rolling contact fatigue performance was proposed by Li and Kahraman, (2013), which is the level of sliding experienced at the contact. The most striking result to emerge from the results data is that poorer micropitting performance has been associated with a higher level of sliding, where the reason for this could be due to a higher rate of cyclic load as the asperities pass one another along the contacting zone.

More recently, significant investigations have been undertaken into the rolling contact fatigue, some of which have included roughness affects (see, for example, Rycerz et al., 2017; Ekberg et al., 2016; Li and Wagner, 2016; Terrin et al., 2017; Paulson et al., 2017; Walvekar et al., 2017).

4.2.2 Fatigue stress cycle

The stresses observed in a fatigue test may be of a simple repetitive form as it is cycled between maximum and minimum values, as shown schematically in Figure 4.1. This figure shows the fatigue cycle of a sinusoidal waveform with a nonzero mean stress. In this case, the stress range, the stress amplitude and the mean stress, respectively, are defined as:

$$\Delta\sigma = \sigma_{\max} - \sigma_{\min}, \qquad \sigma_a = \frac{\Delta\sigma}{2} = \frac{\sigma_{\max} - \sigma_{\min}}{2}, \qquad \sigma_m = \frac{\sigma_{\max} + \sigma_{\min}}{2} \tag{4.1}$$

Generally, stress varies periodically over a given range ($\Delta\sigma$) between limits $\sigma_{\max}$ and $\sigma_{\min}$, which can also be regarded as the sum of a static stress (the mean stress, σ_m) and a cyclic stress of zero mean varying between two values opposite in sign but equal in magnitude (the stress amplitude, σ_a).

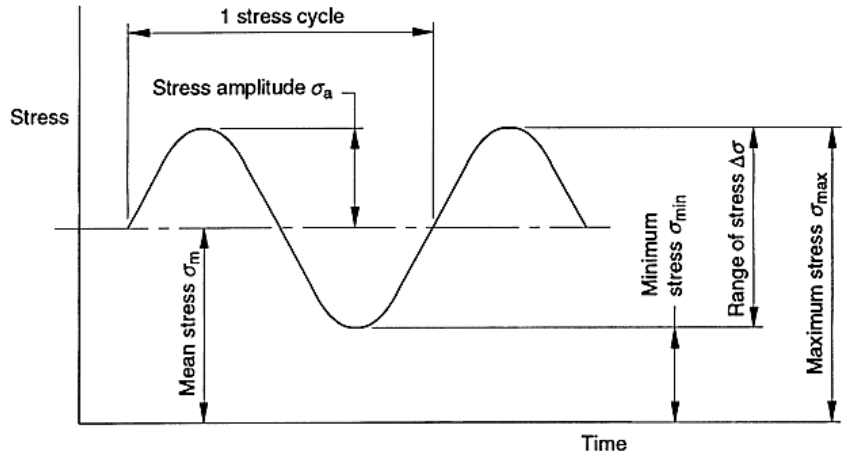


Figure 4-1 the fatigue stress cycle,(Dowling, 1998).

The following ratios of two of these variables are sometimes used:

$$R = \frac{\sigma_{\min}}{\sigma_{\max}}, \qquad A = \frac{\sigma_a}{\sigma_m} \tag{4.2}$$

where R is called the stress ratio and A is the amplitude ratio. With this definition, R can range from −1 to +1 (Suresh, 1998).

- Fully reversed loading, $R = -1$
- Partially reversed loading, $-1 < R < 0$
- Zero-to-tension fatigue, $R = 0$
- Loading between two tensile stresses, $0 < R < 1$
- Static loading, $R = 1$

Materials subject to cyclic loading generally exhibit either cycle hardening or cycle softening (Hassan and Kyriakides, 1994). A number of relations have been proposed to characterise the phenomenological cycle-dependent deformation and fracture behaviour of metals in terms of the initial state and the test conditions. The stresses that are needed to enforce the strain limits usually change as the test progresses. Some materials reveal cycle dependent hardening, which means that the stresses are increased, while others show cycle-dependent softening, or a decrease in stress with increasing numbers of softening cycles, as shown in Figure 4.2. The purpose of fatigue tests is to produce a stabilised stress-strain hysteresis loop (Dowling, 1998), which is shown in Figure 4.3. The steady state cyclic deformation resistance of a metal is conveniently described by the cyclic stress-strain curve, this is shown in Figure 4.4. Such a curve is obtained by connecting the tips of stable hysteresis loops for companion specimens tested at different strain amplitudes. A line connecting the tips of the loops is called the cyclic stress-strain curve, which represents a relationship between stress amplitude σ_a and strain amplitude ε_a for cyclic loading.

The stress range is $\Delta\sigma$, and the elastic portion of the strain range is related to $\Delta\sigma$ by the elastic modulus *E*. Summing the elastic and plastic portions gives the total strain range, $\Delta\varepsilon$:

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{\Delta\sigma}{E} + \Delta\varepsilon_p \tag{4.3}$$

If the resulting cyclic stress-strain curve is above the monotonic stress-strain curve, then the material is one that cyclically hardens, while cyclic softening exists if the cyclic curve is below the monotonic curve. A mixed behaviour may also occur, with crossing of the curves indicating softening at some strain levels and hardening at others.

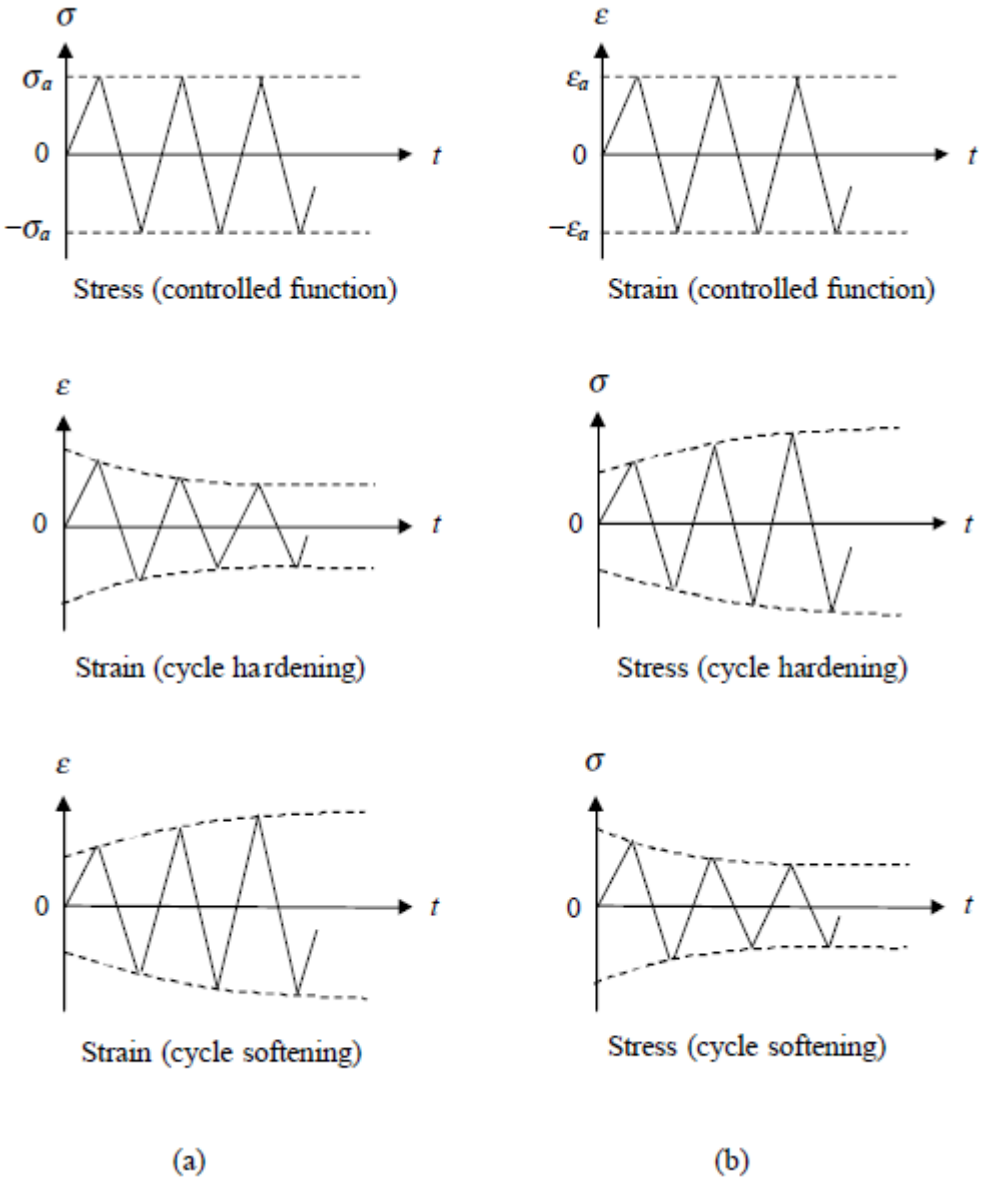


Figure 4-2 Cyclic material behaviour under (a) stress control and (b) strain control (Landgraf, 1970).

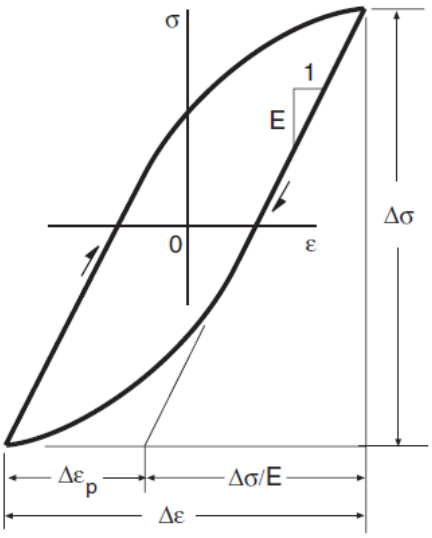


Figure 4-3 Stable stress–strain hysteresis loop, (Dowling, 1998).

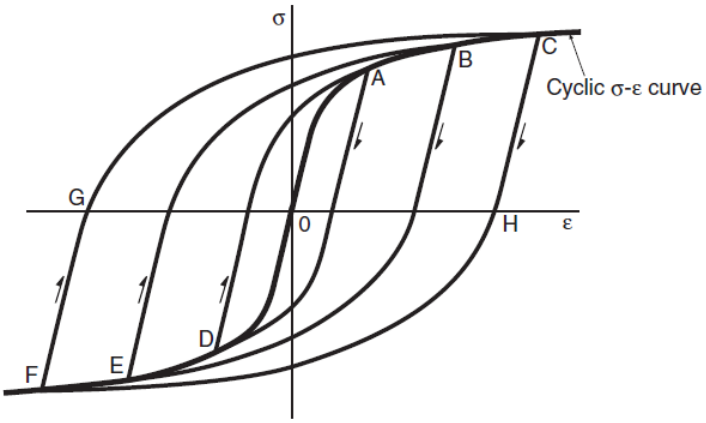


Figure 4-4 Cyclic stress–strain curve defined as the locus of tips of hysteresis loops. Three loops are shown, A-D, B-E, and C-F. The tensile branch of the cyclic stress–strain curve is O-A-B-C, and the compressive branch is O-D-E-F (Dowling, 1998).

The relation between the elastic strain amplitude ε_{ea} and plastic strain amplitude ε_{pa}, can be expressed by the Ramberg-Osgood equation, where H' and n' are the cyclic strength coefficient and cyclic strain hardening exponent, respectively.

$$\varepsilon_a = \varepsilon_{ea} + \varepsilon_{pa} = \frac{\sigma_a}{E} + \left(\frac{\sigma_a}{H'}\right)^{\frac{1}{n'}} \tag{4.4}$$

4.2.3 The fatigue life of materials

The fatigue life of a component is defined as "the total number of cycles or time to induce fatigue damage and to initiate a dominant fatigue flaw which is propagated to final failure" (Suresh, 1998). In this thesis, the fatigue life is expressed in terms of total life. Wöhler's S-N curves are the basis of the Stress-Life approach, where the fatigue data are usually obtained from a fully reversed, constant amplitude fatigue test.

The stress-life approach to fatigue is also introduced as a concept of a fatigue limit or endurance limit, which represents a stress level below which the material does not fail and can be cycled infinitely. A complete S-N curve may be divided into two modes: first, the high-cycle (low stress) range, which induces primarily elastic deformation in a component that is designed for long life (i.e. in the so called high-cycle fatigue (HCF) applications); and second, the low-cycle (high stress) range, which is characterised by a low cyclic fatigue (LCF) life time. In the latter approach, considerable plastic deformation occurs during cyclic loading; for example, the fatigue life is markedly shortened as a consequence of high stress amplitudes or stress concentrations.

HCF curves are usually displayed on a log-log plot with nominal stress amplitude as a y-coordinate against the total cycles to failure (N) as the x-coordinate, as shown in Figure 4.5. As mentioned previously, the HCF models are empirical (i.e. they are based on statistically processed results of a series of tests). The stress-life relationship can be described by relating the stress amplitude in a fully reversed cycle to the number of cycles leading to failure (Basquin relation), as in Dowling (1998) and Suresh (1998):

$$\sigma_a = \sigma'_f \left(2N_f\right)^b \tag{4.5}$$

where σ'_f is the fatigue strength coefficient and b is the fatigue strength exponent.

As shown in Figure 4.5, the fatigue limit or endurance limit, σ_e, is a property of materials where the specimen may be cycled indefinitely below this stress amplitude level without causing failure. For test specimens that do not exhibit such behaviour, a fatigue endurance limit is often defined as a specific long life; say 10^8 fatigue cycles (Dowling, 1998).

Another important term is fatigue strength, which is used to specify the highest stress that a material can withstand for a given number of cycles without breaking. For example, the fatigue strength at 10^8 cycles is the stress amplitude corresponding to N_f =10^8.

In LCF, loading that typically causes failure in less than 10^4 cycles is associated with localised plastic behaviour in metals; therefore, a strain-based parameter should be used for fatigue life prediction in metals. Logarithmic scales are used to allow a convenient graphical representation of the strain- life relationship, where the total strain amplitude ε_a is the ordinate and the number of cycles to failure N_f is the abscissa, as shown in Figure 4.5. A commonly used equation that describes the behaviour of low-cycle fatigue is the Coffin-Manson relation (Suresh 1998):

$$\varepsilon_{pa} = \varepsilon'_f (2N_f)^c \tag{4.6}$$

The elastic strain amplitude is related to the fatigue life by the Basquin relationship:

$$\varepsilon_{ea} = \frac{\sigma_a}{E} = \frac{\sigma'_f}{E}(2N_f)^b \tag{4.7}$$

where ε'_f is the fatigue ductility coefficient, c is the fatigue ductility exponent. (Note that powers b and c are intrinsically negative.)

Combining the equation of the cyclic strain amplitude, Equation (4.4), with the fatigue life correlations for the high-cycle regime, Equation (4.5), and for the low cycle regime Equation (4.6) leads to:

$$\varepsilon_a = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \tag{4.8}$$

Equation (4.8) is considered to be the basis for the strain-life approach to design against fatigue failure Suresh (1998). The strain life equation is shown schematically in Figure 4.6 as being asymptotic to the two straight lines corresponding to Equations (4.5) and (4.6) at long and short lives, respectively. Figure 4.6 also indicates the transition fatigue life, $2N_t$. This is the life at which the cyclic elastic strain range equals the cyclic plastic strain range.

$$2N_t = \left(\frac{\sigma'_f}{\varepsilon'_f E}\right)^{\frac{1}{c-b}} \tag{4.9}$$

At short lives, $2N_f < 2N_t$, plastic strain will predominate, and ductility will control the fatigue performance. At long life, $2N_f > 2N_t$, the plastic strain will be less than the elastic strain, and strength will control the fatigue performance.

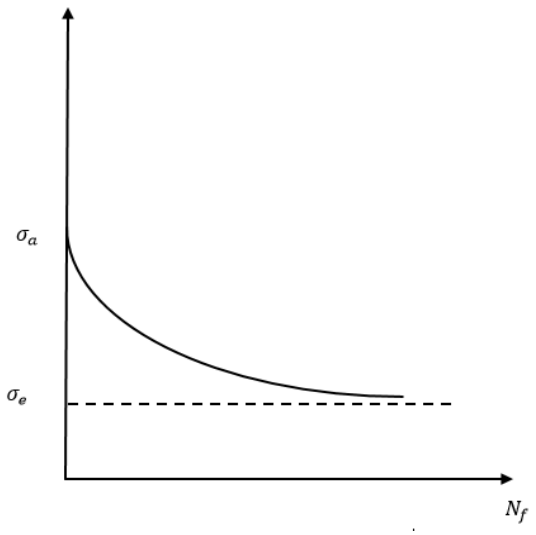


Figure 4-5 Stress-life relationship.

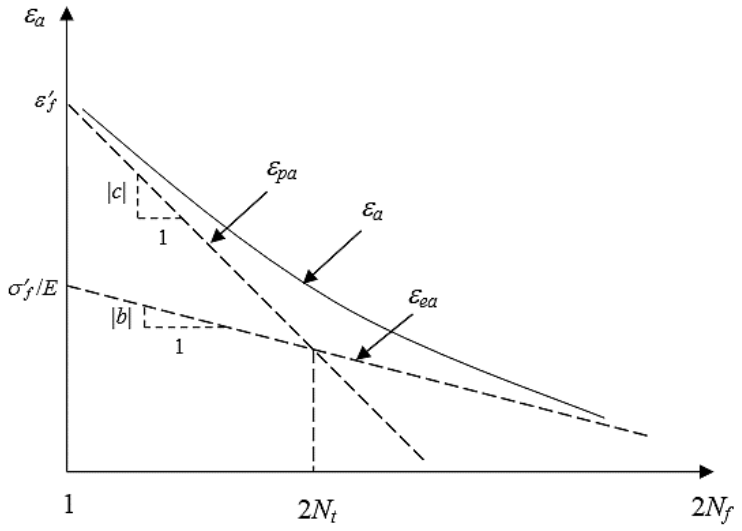


Figure 4-6 Strain-life relationship (Note that both $\varepsilon_a$and N_f are in log scales, and powers b and c are negative and their lines are accordingly illustrated), (Dowling, 2013).

4.2.4 Determination of fatigue properties

The strain-life Equation (4.8) requires four empirical constants (σ'_f, b, ε'_f, c). Several points must be considered in attempting to obtain these fatigue parameters from fatigue data:

- Not all material may be represented by the four-parameter strain-life equation.
- The four fatigue parameters may represent a curve fit to a limited number of data points. So, the actual values of these parameters may be changed if more data are included in the fitting process.
- The fatigue parameters are found from a set of data points over a given range. Therefore, when extrapolating fatigue life, a massive error may occur corresponding to data located outside the required range.

The parameter values obtained for several material tests are available in references Zahavi and Torbilo (1996); BS 3518-1 (1993); BS 3518-3 (1963); and ASTM STP 1122 (1992).From Equation (4.4), the plastic strain term of the cyclic stress-strain curve gives

$$\sigma_a = H' \varepsilon_{pa}^{n'} \tag{4.10}$$

From Equations (4.4) and (4.8) the following properties may be related and those of the cyclic stress-strain curve (Dowling, 1998).

$$n' = \frac{b}{c} \quad , \quad H' = \frac{\sigma_f'}{\left(\varepsilon_f'\right)^{n'}} \tag{4.11}$$

However, using power law relationships in Equation (4.11) is strictly a matter of mathematical convenience and is not based on a physical phenomenon.

The fatigue strength coefficient for steels with hardness below 500 Brinell Hardness Number (BHN) Zahavi and Torbilo (1996) can be approximated in MPa using the ultimate strength of the material, σ_u (Bannantine, 1990); that is,

$$\sigma_f' = \sigma_u + 345 \tag{4.12}$$

The fatigue ductility coefficient can be approximated by the true fracture ductility, ε_f, in a tension test (Bannantine, 1990), as

$$\varepsilon_f' \approx \varepsilon_f = \ln\left(\frac{100}{100 - \%RA}\right) \tag{4.13}$$

where *RA* is defined as the reduction in area at failure.

The fatigue ductility exponent (*c*) for steel is not as well defined as the other fatigue parameters, where Boardman (1990) found that *c* is about -0.6 and the fatigue strength exponent *b* is usually about -0.085. However, Suresh (1998) found that *c* varied between -0.5 and -0.7, and *b* is in the range of -0.05 to -0.12. Based on an energy approach, Morrow determined *b* and *c* (Hertzberg, 1976) as

$$b = \frac{-n'}{1+5n'} \quad , \quad c = \frac{-1}{1+5n'} \tag{4.14}$$

Where n' is the cyclic strain hardening exponent. Morrow found n' to vary between 0.15 and 0.18 for several alloys.

4.3 Mean stress effects on fatigue life

Mean stress is equal to the average of the maximum and minimum stress during a fatigue load cycle. Significant amount of data on fatigue life is obtained for the case of a zero mean stress, which means that the load cycle is completely reversed. However, fully reversed stress cycles with a zero mean stress are not always representative of many applications (Suresh, 1998). Scientifically, the effects of mean stress level on the cyclic deformation and fatigue life are considered to be significant for many applications, where it is seen to influence fatigue lives. The mean stress either increases the fatigue life with a nominally compressive loading or decreases the fatigue life with a nominally tensile loading, as shown in Figure 4.7 (Fatemi et al., 2000).

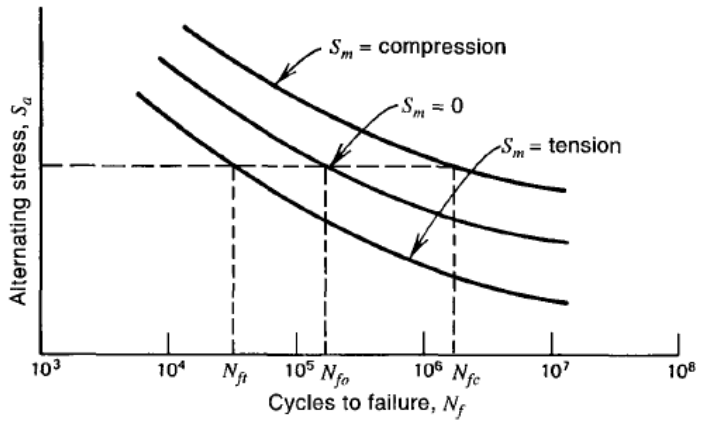


Figure 4-7 Effect of mean stress on fatigue life (Fatemi et al., 2000).

To take into account the mean stress effects, a normalised amplitude-mean diagram can be represented by employing different combinations of the stress amplitude and mean stress

providing a constant fatigue life, such as in Figure 4.8. The most interesting aspect of this graph is that most of the data are located between the straight and curved lines. The straight line is the modified Goodman line, and the curve is the Gerber parabola. Furthermore, the bottom left line of the graph is the Soderberg model, where equations for design are described, respectively, by the following expressions (Suresh, 1998):

$$\frac{\sigma_a}{\sigma_{ar}} + \frac{\sigma_m}{\sigma_u} = 1 \tag{4.15}$$

$$\frac{\sigma_a}{\sigma_{ar}} + \left(\frac{\sigma_m}{\sigma_u}\right)^2 = 1 \tag{4.16}$$

$$\frac{\sigma_a}{\sigma_{ar}} + \frac{\sigma_m}{\sigma_o} = 1 \tag{4.17}$$

The observations with the advantages and disadvantages of the foregoing models for the effects of mean stress on fatigue life can be found in Zahavi and Torbilo, (1996); Dowling, (1998); Suresh, (1998). Closer inspection of the curves in Figure 4.8 shows that the line ACB is considered to be the most suitable for design purposes (Zahavi and Torbilo, 1996). The reason for this choice is that in fatigue design the Soderberg line is overly conservative whereas the Gerber curve is dangerous.

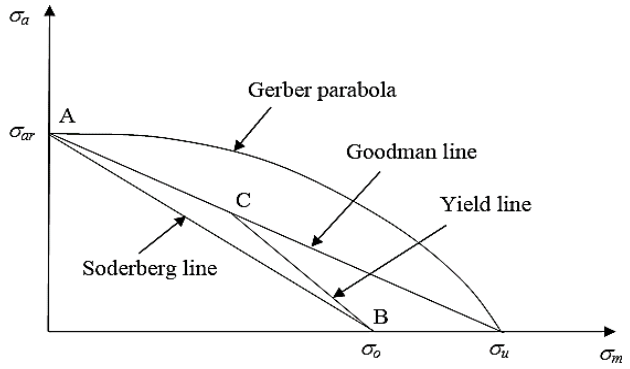


Figure 4-8 Normalised amplitude mean diagram (Dowling 1998).

4.4 Multiaxial fatigue approaches

The safety and durability of components has become more important than before because the sudden failure of complex systems (such as gear systems, automobiles, and bearing units) may cause fatigue for engineering structures and components, which are subjected to multiaxial cyclic loading. Multiaxial fatigue assessments are often complex due to complicated geometries and the mode of applied loads, whose directions change during a cycle of such loadings. In addition, the multiaxial loading has to be identified with the existence of more than one principal stress component.

Multiaxial fatigue models are usually dependent on uniaxial fatigue test data to estimate life corresponding to experimental results. Consequently, several multiaxial fatigue approaches have been developed to reduce the complex multiaxial loadings to an equivalent uniaxial loading. In this section, multiaxial fatigue criteria are discussed for both the stress-based approach and the strain-based approach.

4.4.1 Stress-based approach

4.4.1.1 Effective stress approach

The effective stress approach was proposed to predict multiaxial fatigue life at the early development of multiaxial fatigue research and still remains significant. According to this approach, a two- or three-dimensional stress state is reduced to an effective scalar parameter, usually by using either the Tresca criterion or the von Mises criterion. The stress-based criteria are more suitable in predicting long life (high-cycle) fatigue failure (Sines and Ohgi, 1981). The amplitude of the effective stress (Dowling, 2013) can be written as

$$\bar{\sigma}_a = \frac{1}{\sqrt{2}}\sqrt{\left(\sigma_{1a} - \sigma_{2a}\right)^2 + \left(\sigma_{2a} - \sigma_{3a}\right)^2 + \left(\sigma_{3a} - \sigma_{1a}\right)^2} \tag{4.18}$$

where σ_{ia} (i = 1, 2, 3) are the amplitudes of principal stresses, in terms of the amplitudes of the stress components.

An alternative form of the effective mean stress is given by the von Mises criterion based on the octahedral shear stress,

$$\bar{\sigma}_m = \frac{1}{\sqrt{2}}\sqrt{\left(\sigma_{1m} - \sigma_{2m}\right)^2 + \left(\sigma_{2m} - \sigma_{3m}\right)^2 + \left(\sigma_{3m} - \sigma_{1m}\right)^2} \tag{4.19}$$

For fluctuating multiaxial stresses, Equation (4.19) can be written in terms of the amplitudes of directional stresses:

$$\bar{\sigma}_a == \frac{1}{\sqrt{2}}\sqrt{\left(\sigma_{xa} - \sigma_{ya}\right)^2 + \left(\sigma_{ya} - \sigma_{za}\right)^2 + \left(\sigma_{za} - \sigma_{xa}\right)^2 + 6\left(\tau_{xya}^2 + \tau_{yza}^2 + \tau_{zxa}^2\right)} \tag{4.20}$$

To calculate the fatigue life, N_f, the Basquin exponent b and the fatigue strength coefficient σ'_f in Equation (4.5) are used in combination with the following stress-life equation for multiaxial stresses (Dowling, 2013).

$$\bar{\sigma}_a = \left(\sigma'_f - \bar{\sigma}_m\right)\left(2N_f\right)^b \tag{4.21}$$

One of the major criticisms of this effective stress concept is that the differing effects of axial tensile and compressive mean stresses, which are found from multiaxial fatigue experiments, may not be accurately estimated. Moreover, these criteria are unable to derive a quantitative orientation of fatigue cracks with respect to the loading axes (Suresh, 1998).

4.4.1.2 Criteria involving multiaxial fatigue

There is not yet a universally accepted model in spite of a great number of multiaxial fatigue criteria, which were presented by Garud, (1981); Brown and Miller, (1973); You and Lee, (1996); Papadopoulos et al. (1997); Miller et al.,(1985) . The major stress-based criteria (i.e. criteria for high-cycle fatigue) are described in the following subsections.

Sines Criterion:

Sines criterion (1959) uses the amplitude of second invariant of stress tensor deviator (which corresponds to the von Mises stress) as the basis (i.e. the mean value of the hydrostatic stress), and can be written as:

$$\sqrt{J_{2,a}} + \bar{\kappa}\sigma_{h,m} \leq \bar{\lambda} \tag{4.22}$$

Fatigue failure occurs if this inequality is not fulfilled. The parameters $\bar{\kappa}$ and $\bar{\lambda}$ can be obtained from a repeated bending test as:

$$\bar{\kappa} = \frac{3\tau_{af}}{\sigma_{ab}} - \sqrt{3} \; ; \; \bar{\lambda} = \tau_{af} \tag{4.23}$$

where τ_{af} is the fatigue limit for pure torsion and σ_{ab} is the repeated bending fatigue limit. The Sines high-cycle fatigue criterion correctly captures the fatigue test observations: first, the fatigue limit in torsion is independent of a mean shear stress; and second, the endurance limit in bending varies linearly with a static normal stress. However, the equations suggested by Sines need verification under out-of-phase loadings or other factors to predict the fatigue behaviour more accurately (You and Lee, 1996).

Crossland Criterion

Crossland (1956) proposed the amplitude second invariant of the deviatoric stress tensor and the maximum hydrostatic pressure during multiaxial loading as the governing variables in the crack initiation process, which can be written as:

$$\sqrt{J_{2,a}} + \bar{\kappa}\sigma_{h,\max} \leq \bar{\lambda} \tag{4.24}$$

The parameters $\bar{\kappa}$ and $\bar{\lambda}$ can easily be obtained from a torsion test or a fully reversed bending test. The limitations of the theory of Crossland are that it shifts the mean values of the fatigue index error to the non-conservative side. In addition, its mean values are under out-of-phase loading (Papuga, 2012).

4.4.1.3 Critical plane approaches

A critical plane approach is employed when the loading is nonproportional to a significant degree. In this approach, stresses during cyclic loading are determined for various orientations (planes) in the material. The plane on which the tensile normal stress or the shear stress is a maximum is used to predict fatigue failure in the materials. In general, determining the orientations of the critical planes for multiaxial loading with cyclic stress can become very complicated due to the changing fatigue properties. However, as reviewed by Chu et al. (1993), Fatemi et al. (2000) and Van et al. (2013) some simplifications can be applied for a given material or loading situation by noting that most fatigue cracking initiates at the surface level, as will be discussed in more detail in the following section.

A summary of the multiaxial fatigue criteria based on the critical plane approach, such as the Findley criterion, the Matake criterion, the Dang Van criterion and the McDiarmid Criterion is given in Table 4.1 and they can be written in a general form as:

$$A + \bar{\kappa}B \le \bar{\lambda} \tag{4.25}$$

Fatigue failure occurs if the inequality is not fulfilled. Parameters *A* and *B* are related to the stress components on the critical plane and $\bar{\kappa}$ and $\bar{\lambda}$ are material constants. These parameters can be found from the endurance limits, which characterises the applied stress amplitude below which a material is expected to have an infinite fatigue life.

Table 4-1 Parameters of Equation (4.25) for Findley, Matake's, Dang Van's and McDiarmid's criteria

Criterion	*A*	*B*	$\bar{\kappa}$	$\bar{\lambda}$	Hypothesis
Findley	τ_a	$\sigma_{n,\max}$	$\frac{2-\sigma_{af}/\tau_{af}}{2\sqrt{\sigma_{af}/\tau_{af}-1}}$	$\frac{\sigma_{af}}{2\sqrt{\sigma_{af}/\tau_{af}-1}}$	Fatigue damage of material is due to a linear combination of normal stress and shear stress amplitude in the critical plane.
Matake	τ_a	$\sigma_{n,\max}$	$\frac{2\tau_{af}}{\sigma_{af}}-1$	τ_{af}	The equivalent shear stress should be calculated on the plane with maximum shear stress amplitude.
Dang Van	$\tau_a(t)$	$\sigma_h(t)$	$3\frac{\tau_{af}}{\sigma_{af}}-\frac{3}{2}$	τ_{af}	The parameters responsible for crack nucleation along slipping bands are microscopic shear stress in the grain area and the microscopic hydrostatic stress.
McDiarmid	$\tau_{cp,a}$	$\sigma_{cp,\,max}$	$\frac{\tau_{af}}{2\sigma_u}$	τ_{af}	The maximum shear stress range and the normal stress in the critical plane are significant.

Where τ_a and $\sigma_{n,\max}$ are the shear stress amplitude and maximum value of the normal stress on the critical plane, respectively. σ_{af} and τ_{af} are the endurance limits under fully reversed bending and fully reversed torsion, respectively. The linear criterion proposed by Dang Van uses the current values of the microscopic shear stress amplitude on the critical plane, $\tau_a(t)$, and hydrostatic stress, $\sigma_h(t)$, which are independent of the critical plane orientation and *t* is time.

4.4.2 Strain-based approach

4.4.2.1 Critical planes of failure in multiaxial loading

Extensions of critical plane strain criteria have been proposed by many researchers. These developments are based on significant observations. First, the alternating shear stress and strain on the critical plane are the major cause of fatigue in multiaxial loading. Second, the mean or maximum values of the normal and shear stress and strain components on the critical plane are significant minor contributors to fatigue failure. Consequently, numerous models of the critical plane strain criterion that take the effect of the normal mean stress have been proposed. One interesting criterion was applied by Fatemi and Socie (1988) to incorporate mean stress using the maximum value of normal stress during a cycle to modify the damage parameter. In this criterion, fatigue damage is due to the shear strain amplitude influenced by the maximum normal stress, which is expressed as

$$\frac{\Delta\gamma_{\max}}{2}\left(1+k\frac{\sigma_n^{\max}}{\sigma_\gamma}\right)=\frac{\tau'_f}{G}\left(2N_f\right)^{b_o}+\gamma'_f\left(2N_f\right)^{c_o} \tag{4.26}$$

where $\frac{\Delta\gamma_{max}}{2}$ refers to the amplitude of shear strain on the critical plane, and σ_n^{max} refers to the maximum tensile stress which is normal to the critical plane. *K* is a material constant, *K* = 0.6 to 1.0, *G* is the shear modulus, and σ_γ is the yield strength for the cyclic stress-strain curve. The shear fatigue strength, τ'_f, and shear fatigue ductility coefficient, γ'_f, are found from fully reversed torsion fatigue tests in pure shear. In addition, b_0 and c_0 are shear fatigue strength and shear fatigue ductility exponents, respectively.

4.4.2.2 Smith, Watson and Topper (SWT) criterion

Smith and co-authors proposed a simple form of a damage parameter, SWT, which is a reasonable model for tensile stress dominated cracking (Dowling, 2013). This model was originally developed as a correction for mean stresses in uniaxial loading situations. The

SWT model is based on strain-life test data obtained with different mean stresses, which includes both the cyclic strain range and the maximum stress. It is expressed as follows (Dowling, 2013):

$$\sigma_{\max}\varepsilon_a = \frac{\sigma'^{2}_f}{E}\left(2N_f\right)^{2b} + \sigma'_f\varepsilon'_f\left(2N_f\right)^{b+c} \tag{4.27}$$

where $\sigma_{\max}$ the maximum normal is stress on the critical plane and ε_a is the amplitude of normal strain for the same plane as $\sigma_{\max}$. The product $\sigma_{\max}\varepsilon_a$ is assumed to be constant for a given life. If it is assumed to be zero, then Equation (4.27) predicts infinite life.

This model is used widely and is a common approach in most commercial fatigue analysis software, where it has been shown to better correlate mean stress data during multiaxial loading for a wider range of materials and it is, therefore, regarded as more favourable for general use (Carpinteri et al., 2003). Furthermore, the SWT model can also be used in the modelling of both proportionally and non-proportionally loaded components for materials that fail mostly due to planes of high tensile stress (mode I).

4.5 Fatigue from variable amplitude loading

Variable amplitude fatigue crack growth is a very significant aspect during each operating cycle, especially with respect to EHL contact fatigue problems. Hence, a form of variable amplitude time history analysis should be introduced to achieve accurate lifetime predictions. In general, loading above a certain level can eventually lead to a significant damage of the component, which causes microscopic rearrangements at the atomic scale. To deal with variable amplitude loading, several major issues have to be elucidated. First, how to relate the damage accumulation to a time history of load or nominal stress for a material considered. Second, how to identify cycles within a given time history and sum up partial

damages which result from these different stress levels. To achieve this, a cumulative damage theory must be considered. The basic steps of identifying cycles can be handled in a comprehensive manner by cycle counting, as will now be discussed.

4.5.1 Cycle counting for irregular histories

Cycle counting methods are often used to reduce the random load history into a series of discrete events, which can be analysed using the test data obtained for constant amplitude fatigue loads. A number of cycle counting methods have been introduced over the years, which include level crossing counting, peak counting, simple range counting, and the rainflow counting method, Suresh (1998). Some of these methods are employed in the analysis of fatigue data to reduce a spectrum of varying load into a set of simple, uniform data histograms with constant amplitude cycle. Amongst these cycle counting algorithms, the rainflow-counting algorithm is considered to be one of the best approaches for finding half-cycles within a given time history (Dowling, 2003). This method was invented by Matsuishi and Endo (1968), and is considered to be the basis of the development of an overall cycle counting algorithm for estimating fatigue damage of components that is caused by randomly fluctuating loading histories. A variety of variations of the rainflow counting method for highly irregular variations of load with time have been developed by Dowling (2003), Suresh (1998), Musallam and Johnson (2012), Downing and Socie (1982), and Amzallag et al. (1994).

The traditional rainflow counting method defines cycles as closed stress-strain hysteresis loops, without elaborate local stress-strain analysis. An example of the rainflow cycle counting method can be seen in Figure 4.9. The sequence of reversals 1 to 7 is registered in the order (4–3), (6–5), and (2–1). In addition, this plot shows an irregular stress history that consists of a series of peaks and valleys, which are points where the direction of loading

changes. Consequently, such an irregular stress history leads to the stress–strain path, as shown in the middle plot that forms the closed hysteresis loops shown at the right.

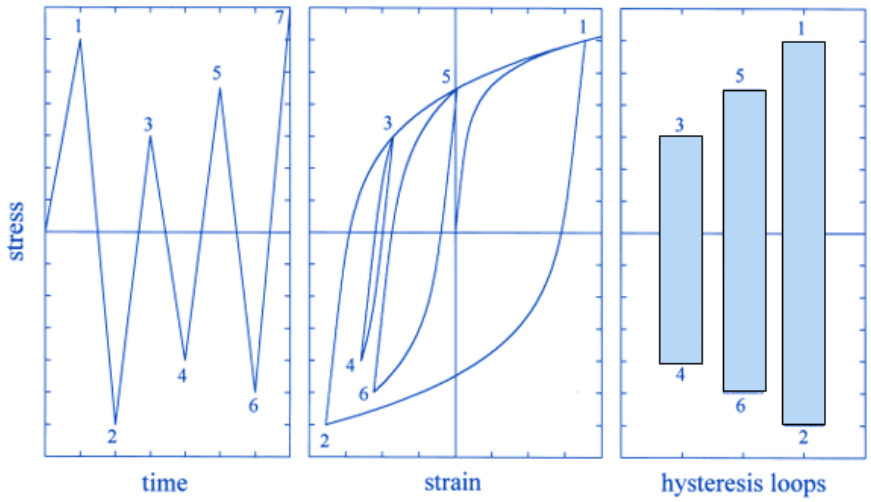


Figure 4-9 the stress-strain hysteresis loops extracted from the rainflow counting algorithm (Anthes, 1997).

These results can be used in conjunction with the Palmgren-Miner rule to predict the fatigue life of engineering components that are subjected to variable amplitude loading. The basis of this theory is the linear damage concept:

$$D_i = \frac{N_i}{N_{fi}} \tag{4.28}$$

Failure is predicted to take place when the total damage is

$$D = \sum D_i = \sum \frac{N_i}{N_{fi}} \geq 1 \tag{4.29}$$

Where the N_i is the number of cycles accumulated and N_{fi} is the average number of cycles to failure.

4.6 Cyclic shakedown limits

Under repeated application of the load, as for example occurs in rolling element bearings, gears, cam-followers, and rail-wheel contacts, the fatigue life of these components may be expected to depend upon the progress of plastic deformation. This process is described as shakedown and the maximum load for which it occurs is identified as the elastic shakedown limit (Williams et al., 1999). Generally, the shakedown process makes the pure elastic condition become the steady cyclic load state, whereby initial plastic deformation introduces residual stresses (Johnson, 1985). Such residual stresses are considered to be protective when the surface zone yields in tension during loading, so the residual surface stress will be in compression after unloading. On the other hand, these residual surface stresses can be developed in the sub-surfaces level of tribological contacts, which enables loads that are sufficiently large to cause initial plastic deformation to be accommodated purely elastically in the longer term. This means that the residual stress pattern, which was induced during shakedown process, builds up and reaches a stabilised state, so that it does not make any additional deformation under further cyclic loading. The stabilised stress cycle is then purely elastic. The influence of these residual stresses on crack growth and fatigue requires better understanding. From this perspective, the current research sheds light on the potential effect of the results obtained for plastic deformation and residual stress fields of the real surface asperities on fatigue damage under lubricated contact condition.

The development of plastic strain behaviour under cyclic contact loading is divided into four different regimes, as shown in the Figure 4.10, and as described in the following definitions:

(1) Perfectly elastic: this occurs when the maximum normal contact pressure is below the ‘elastic limit’, the behaviour will be perfectly elastic.

2) Cyclic plastic straining: here the material experiences steady, reversed cyclic plastic strains. In this mode of plastic strain development, the progressive accumulation of plastic strains is suppressed.

(3) Ratcheting: here the plastic strain magnitude increases continually with the load cycling, which eventually leads to incremental collapse.

(4) Elastic Shakedown: here plastic flow occurs during early cycles and subsequent material response is entirely elastic. No cyclic plasticity occurs in this regime.

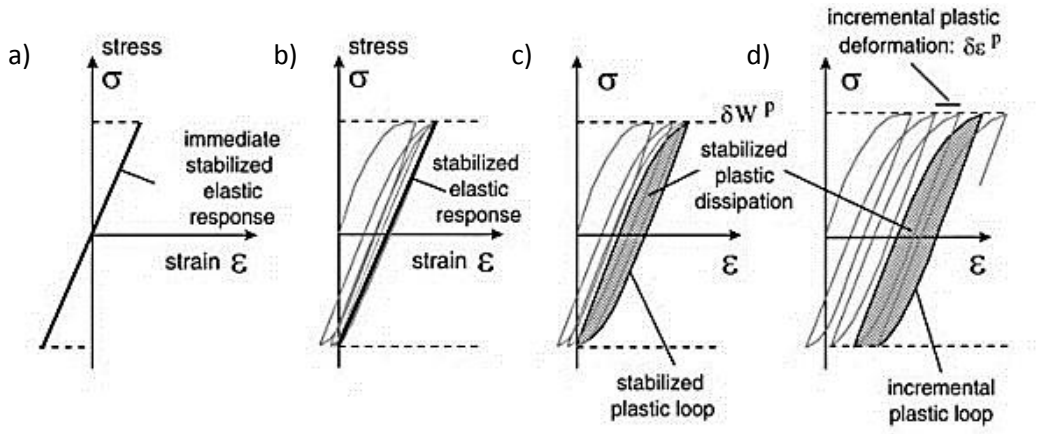


Figure 4-10 The effects of cyclic loading on the material: (a) perfectly elastic, (b) elastic shakedown, (c) plastic shakedown, (d) ratcheting (Fouvry et al.,2001).

4.7 Conclusion

In this chapter, some of the most widely used contact fatigue models are outlined and discussed to enhance our understanding of RCF. Some of these models are presented in this chapter for surface loading due to the EHL mechanism in the area near the surface asperities of the simulated body and they also incorporate methods for the accumulation of fatigue damage. A discussion of the various mechanisms by which crack propagation retardation occurs in components under variable amplitude cyclic loading was also presented in this chapter. It is suggested that variable amplitude-loading can have a distinctly different effect on the characteristics of fatigue damage growth than on fatigue crack initiation. One of the more significant findings to emerge from this review is that the strain-life theory is a comprehensive approach that is recommended for both low-cycle and high-cycle contacting fatigue problems.

Chapter 5 Fatigue analysis – applying multiaxial fatigue criteria based on critical plane approach

5.1 Introduction

It is now well established that during pure rolling lubricated conditions micro pits can be found on the surface of gears and rolling bearing (Johnson, 1987). This type of surface damage arises as a result of repeated deformation of micro-volumes of the mechanical components under cyclic contact loading (Kragelsky et al., 2001). Such contact scenarios involving repeated contacts and complex time-varying loadings are considered as multiaxial rather than uniaxial loading states. However, uniaxial fatigue test data can be used for prediction of fatigue life using a number of available models. Many multiaxial fatigue life estimation criteria have been proposed to reduce the complex multiaxial stress mode to an equivalent uniaxial stress state. The criteria based on the critical plane approach are considered the most appropriate to provide effective fatigue life estimation in these circumstances (Fatemi and Socie 1988). In this chapter, multiaxial fatigue criteria based on the critical plane approach are adopted to model fatigue analysis for the micro-EHL transient problem in rough gear contact problems. Also, the rough surface EHL in Line Contact analysis which provides the surface loading for the analysis is introduced.

5.2 EHL line contact fatigue analysis

5.2.1 Introduction

Results of the transient micro-EHL solutions for the two rough surfaces in rolling/sliding contact provide the transient variation of the surface loading applied to the disks. This enables the stresses below the rough contact surface to be calculated, and this stress history for the disk material enables evaluation of the fatigue models considered. Results of the micro-EHL modelling, which gives the full time-varying behaviour of lubricant pressure and film thickness, have been used to predict the fatigue life and accumulated damage at the scale of surface asperity features. To accurately understand the contact mechanics of the surface micropitting phenomenon it is essential to find the detailed tooth loading in terms of asperity pressure, shear stress and the local subsurface stress components under conditions of elastohydrodynamic lubrication (EHL). Dowson and Higginson, (1966) provided the first numerical solutions to this problem for smooth surfaces, which is considered the normal film creation mechanism in gear tooth contact analysis. Transient micro-EHL solutions taking surface roughness in to account have been used in the current research instead of smooth surface EHL because the local EHL mechanism may fail in the most severe cases of thin films/high roughness where transient direct solid interaction happens between prominent asperities. This is known as mixed lubrication regime where the tooth load is carried by a combination of fluid film pressure and boundary-lubricated asperity encounters. Different novel approaches have been made to predict transient asperity interactions for elliptical contacts and mixed lubrication problems. In this work the coupled differential deflection method developed by Elcoate et al., (2001) ; Holmes et al., (2003 a); Holmes et al., (2003 part b) ; Holmes et al., (2004) successfully modelled the micro-EHL transient problem in which the mean asperity height was at least an order of magnitude greater than the minimum lubricant film thickness.

5.2.2 Micro-Elastohydrodynamic theory

Real gear surfaces operating under typical industrial conditions of load, speed and oil viscosity are considered as operating under mixed lubrication conditions. This condition can be summarised as occurring when the roughness features are far greater in height than the predicted EHL film thickness. This section shows some numerical formulations for EHL line contact with real rough surfaces. The coupled numerical solution technique for estimating the pressure and film thickness distribution in micro-EHL conditions is briefly described in this section. This unified solution method will be used as the EHL analysis tool in this thesis. A classical EHL model comprises an equation describe the flow of the fluid such as Reynolds equation or the Navier-Stokes equation in this work the Reynolds equation is applied, the film thickness equation (also referred to as the elastic deformation equation), the force balance equation, and equations describing the viscosity-pressure-temperature and density-pressure-temperature relations.

5.2.2.1 The Reynolds equation

Reynolds' equation is the equation that describes the relationship between the film thickness and the lubricant pressure for conditions of full film lubrication. This equation was first proposed by Reynolds in 1886 and is referred to as Reynolds' equation. It can be derived from first principles by applying conservation of momentum, conservation of mass and conservation of energy (Pletcher et al., 2012). It is a combination of two physical principles; the Navier-Stokes equations for the motion of a viscous fluid and the continuity equation which is based on conservation of mass in the fluid.

A number of simplifying assumptions are made that result in the Navier Stokes equations being reduced to a balance between the pressure gradient forces causing motion of the fluid, and the viscous resistance to that motion. These are as follows.

1. The Reynolds number is small so that flow is laminar and the inertia terms (rates of change of momentum) for the fluid may be ignored.

2. There is no slip at the solid/lubricant boundaries.

3. The fluid is Newtonian so that shear stress is proportional to the shear strain rate.

4. Pressure, density and viscosity do not vary across the film.

5. The film thickness is much smaller than the Radius of Relative Curvature of the lubricated contact.

6. The film thickness is much smaller than the dimensions of the lubricated contact.

Assumptions 5 and 6 describe the geometric nature of lubricated contacts and contribute to the smallness of the Reynolds number set out in Assumption 1. Assumption 2 is general in fluid mechanics and is justified by experimental evidence. Assumption 3 is not strictly true when a contact is subjected to high sliding velocities but is generally reasonable in fluid mechanics. Assumption 4 may be relaxed without changing the nature of the resulting equation, but is generally made except when considerations of thermal effects within the film become necessary.

Reynolds equation, including non-Newtonian shear thinning based on the Eyring model, was developed by Conry et al., (1987) by considering the equilibrium of an element of flow subjected to viscous shear and using the continuity of liquid principle . With the adoption of rough surfaces, transient modes should be used due to the time varying geometries of contacting rough solids. In the following treatment the two- dimensional isothermal transient Reynolds equation is derived after Conry et al., (1987) incorporating the time-dependent squeeze-film action. The case shown in Figure 5.1 where the two surfaces are separated by a molecularly thin lubrication film is considered. By considering a small element of lubricant

in the form of control volume as shown in Figure 5.2, the equation of equilibrium in the *x*-direction can be written as shown in equation (5.1).

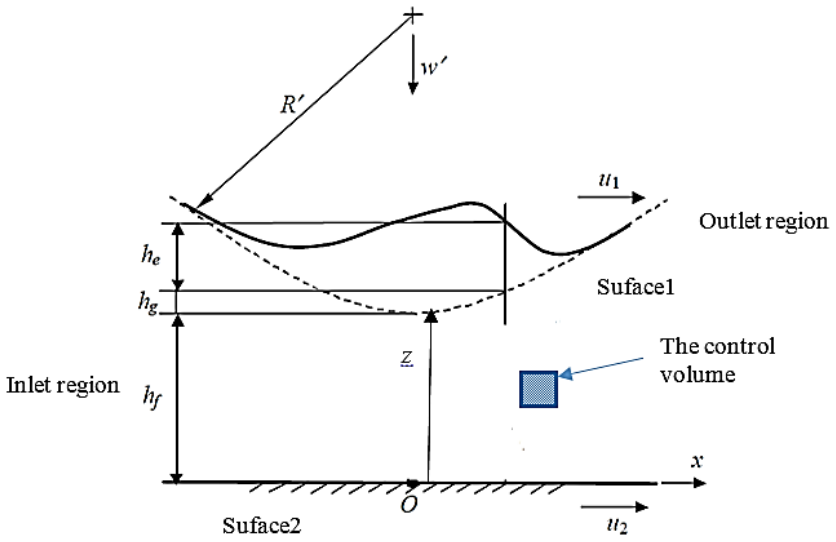


Figure 5-1 Section at the centre of the contact in the entrainment direction.

The control volume in the above Figure 5-1 is subject to forces due to pressure and shear stress acting on its boundary as shown in the Figure 5-2.

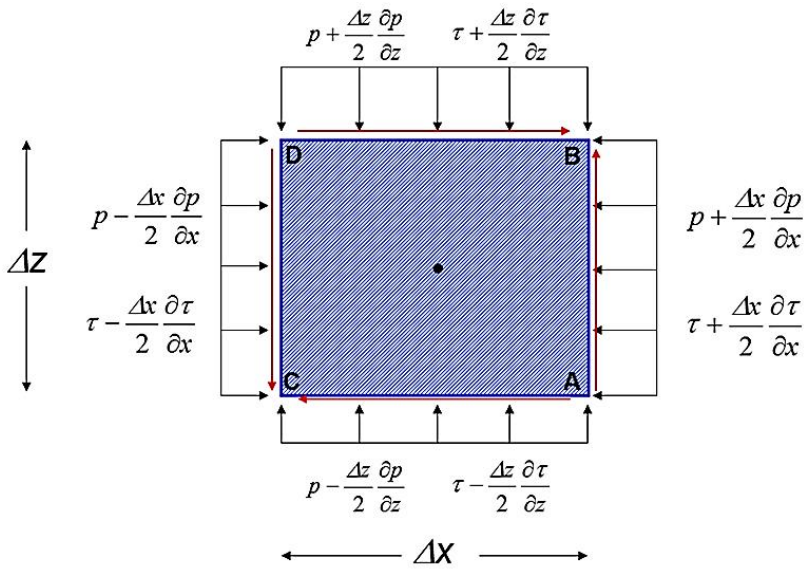


Figure 5-2 Forces acting on control volume.

$$\frac{\partial \tau}{\partial z} = \frac{\partial p}{\partial x} \tag{5.1}$$

Where the shear stress is $\tau = \tau_{xz}$ and τ_{yx} can be assumed negligible due to scale effects as lubricant shear happens predominantly in the *z* direction. Assuming that the pressure *p* stays constant across the film thickness, integration of Equation (5.1) results

$$\tau = \tau_2 + \frac{\partial p}{\partial x} z \tag{5.2}$$

where τ_2 is the shear stress acting on the boundary of surface2. The constitutive equations of the Eyring fluid (Kim et al., 2001), that relates shear stress to shear strain rate, neglecting the elastic shear term, is

$$\dot{\gamma} = \frac{\partial u}{\partial z} = \frac{\tau_0}{\eta} \sinh\left(\frac{\tau}{\tau_0}\right) \tag{5.3}$$

The parameter τ_0 in equation (5.3) is the representative stress or Eyring stress and is the shear stress that corresponds to the onset of significant non-Newtonian behaviour. The Eyring stress is assumed to be constant. Substituting Equation (5.2) into Equation (5.3) and integrating gives the velocity of the lubricating fluid as

$$\begin{aligned} u &= u_2 + \int_0^z \frac{\tau_0}{\eta} \sinh\left(\frac{\tau_2 + \frac{\partial p}{\partial x} s}{\tau_0}\right) ds \\ &= u_2 + \frac{\tau_0^2}{\eta \frac{\partial p}{\partial x}} \left[\cosh\left(\frac{\tau_2 + \frac{\partial p}{\partial x} z}{\tau_0}\right) - \cosh\left(\frac{\tau_2}{\tau_0}\right) \right] \end{aligned} \tag{5.4}$$

Defining the mid-plane at ($z = h/2$), Equation (5.4) can be simplified to

$$u = u_2 + \frac{\tau_0 h}{2\eta\Sigma}\left\{\cosh\left[\frac{\tau_m}{\tau_0} - \Sigma + \frac{2z}{h}\Sigma\right] - \cosh\left(\frac{\tau_m}{\tau_0} - \Sigma\right)\right\} \tag{5.5}$$

Where $\tau_m = \tau_2 + \frac{\partial p}{\partial x}\frac{h}{2}$ and $\Sigma = \frac{h}{2\tau_0}\frac{\partial p}{\partial x}$.

Applying the boundary condition $u\big|_{z=h} = u_1$, the relative sliding velocity of the two contacting surfaces can be found by

$$u_1 - u_2 = \frac{\tau_0 h}{\eta\Sigma}\sinh\left(\frac{\tau_m}{\tau_0}\right)\sinh(\Sigma) \tag{5.6}$$

The shear stress at the mid-plane can be found from

$$\sinh\left(\frac{\tau_m}{\tau_0}\right) = \frac{\eta(u_1 - u_2)}{\tau_0 h}\frac{\Sigma}{\sinh\Sigma} \tag{5.7}$$

The Mass flow rate in the *x*-direction between the surfaces per unit length at the centre of the control volume can be defined as

$$M = \int_0^h \rho u dz \tag{5.8}$$

Substitution of Equation (5.5) into Equation (5.8) and integrating gives

$$\begin{aligned} M &= \rho u_2 h + \rho\frac{\tau_0 h}{2\eta\Sigma}\left\{\frac{h}{2\Sigma}\left[\sinh\left(\frac{\tau_m}{\tau_0} + \Sigma\right) - \sinh\left(\frac{\tau_m}{\tau_0} - \Sigma\right)\right] - h\cosh\left(\frac{\tau_m}{\tau_0} - \Sigma\right)\right\} \\ &= \rho u_2 h + \rho\frac{\tau_0 h^2}{2\eta}\left[\frac{\sinh\Sigma}{\Sigma}\sinh\left(\frac{\tau_m}{\tau_0}\right) + \frac{\sinh\Sigma - \Sigma\cosh\Sigma}{\Sigma^2}\cosh\left(\frac{\tau_m}{\tau_0}\right)\right] \end{aligned} \tag{5.9}$$

Substitution of Equation (5.7) into Equation (5.9) leads to

$$M = \rho h \frac{u_1 + u_2}{2} + \frac{\rho h^3}{12\eta}\frac{\partial p}{\partial x}\left[\frac{3(\sinh\Sigma - \Sigma\cosh\Sigma)}{\Sigma^3}\cosh\left(\frac{\tau_m}{\tau_0}\right)\right] \qquad (5.10)$$
$$= M_C + M_P$$

The total flow M is divided into two terms, firstly Couette Flow M_C, which is caused by the movement of the surfaces. Secondly the term containing the pressure gradient is called the Poiseuille Flow M_P which is caused by the variations in pressure. By defining a dimensionless factor as

$$S = -\frac{3(\sinh\Sigma - \Sigma\cosh\Sigma)}{\Sigma^3}\cosh\left(\frac{\tau_m}{\tau_0}\right) \qquad (5.11)$$

Equation (5.9) can be simplified to the following form

$$M = \rho\bar{u}h - \frac{\rho h^3}{12\eta}S\frac{\partial p}{\partial x} \qquad (5.12)$$

The continuity equation in terms of *M* can be expressed as

$$\frac{\partial M}{\partial x} + \frac{\partial}{\partial t}(\rho h) = 0 \qquad (5.13)$$

Substituting equation (5.12) into equation (5.13) and rearranging provides

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{12\eta}S\frac{\partial p}{\partial x}\right) - \frac{\partial}{\partial x}(\rho\bar{u}h) - \frac{\partial}{\partial t}(\rho h) = 0 \qquad (5.14)$$

This equation is the transient Reynolds equation based on the Eyring equation for the lubricant relating its pressure with the fluid film thickness, and the elastic deflection of the surfaces under the action of pressurised lubricant. Term *S* is the non-Newtonian factor which

depends on pressure, film thickness, sliding speed and pressure gradient. Considering shear thinning using the Eyring model, the *S* factor is available in closed form taken from Hughes et al., (2000) as follows

$$S = \frac{3(\Sigma\cosh\Sigma - \sinh\Sigma)}{\Sigma^3}\sqrt{1+\frac{\eta^2(u_2-u_1)^2}{\tau_0^2 h^2}\frac{\Sigma^2}{\sinh^2\Sigma}} \tag{5.15}$$

The representative stress τ_0 , for the current thesis is taken as constant although it is known to be weakly dependent on the pressure. As τ_0 approaches infinity so Σ approaches zero and it can be shown that *S(x)* approaches unity, therefore returning equation (5.14) to the Newtonian fluid formulation. The last term in equation (5.14), which is the only time dependent term, is called the squeeze film term. This terminology arises from squeeze film dampers where the pressure is generated by the movement of parallel solid surfaces towards each other when the gap in between is filled with damping fluid.

5.2.2.2 Elasticity equation

A typical centreline section of film thickness in the entrainment direction for the EHL contact problem is shown in Figure 5.1. The film thickness profile is given by

$$\begin{aligned} h(x,t) &= \phi(x,t) + h_g(x) + h_e(x,t) + h_f \\ &= \phi(x,t) + \frac{x^2}{2R'} - \frac{4}{\pi E'}\int p(s,t)\ln|x-s|ds + h_f \end{aligned} \tag{5.16}$$

where the expression $\phi(x, t)$ is the measured roughness at x taken from a profilometer measurement for the fast and slow rough disks, h_g is the undeformed geometry that is assumed to be a parabola, and h_e is the elastic deformation of the contacting surfaces due to pressure distribution p. It is assumed that the contacting bodies behave as semi-infinite bodies, as the size of the contact is usually small compared to the remainder of the solid body. Finally h_f is a constant and its value controls the separation of the surfaces.

5.2.2.3 The viscosity and density equation

The equation which relates the lubricant viscosity to its pressure, is called the viscosity equation. In tribology the key effect of viscosity is the shear stress and it is considered as the most important lubricant property (Williams, 1994). The simple and most widely used form of this equation is known as the Barus equation, developed by Barus (1893).The equation is of the following exponential form:

$$\eta = \eta(p) = \eta_0 \ exp(\alpha p) \tag{5.17}$$

Values of the pressure viscosity exponent (α) and the viscosity at atmospheric pressure η_0 , vary with temperature . For systems that operate over a restricted range of temperatures the variation of viscosity with temperature is of minor importance so long as the appropriate viscosity can be obtained for the application. However, lubricant viscosity may change dramatically under a high pressure and this makes the Reynolds equation highly non-linear and significantly more difficult to deal with numerically. The Barus viscosity equation is not suitable for simulation at high pressure as it overestimates the viscosity. A more realistic equation for high pressure exists in the form of the Roelands equation (Stachowiak and Batchelor, 2005), which can be expressed as:

$$\eta = \eta_0 \exp\left\{\ln(\eta_0 / \kappa)[(1+\chi p)^Z - 1]\right\} \tag{5.18}$$

Where κ and χ are constants, Z is the Roelands pressure viscosity parameter, and p and η_0 are in units of Pa and Pas respectively. The constant Z can be found from α by comparing the Roelands equation and Barus equation as pressure distribution, p approaches to zero as shown below (Kreyszig 1999):

$$Z = \frac{\alpha}{\chi \ln(\eta_0 / \kappa)} \tag{5.19}$$

The pressure - density relationship is given by the Dowson and Higginson (1966) formula and Evans et al. (2013) .The equation takes the following form:

$$\rho = \rho_0 \frac{1+\gamma p}{1+\lambda p} \tag{5.20}$$

Where γ and λ are the pressure coefficients of compressibility.

5.2.2.4 EHL solution techniques

Solving the elastohydrodynamic lubrication (EHL) point and line contact problem is possible by discretization of the Reynolds and elastic deflection equations. The current research is based on numerical analysis of EHL contacts using a novel coupled differential deflection formulation developed by Cardiff Tribology Group (Hughes et al., (2000); Elcoate et al., (2001); Holmes et al., (2003a, 2003b, and 2005)). The efficiency of this method is validated under transient conditions by Hughes et al., (2000) and Holmes et al., (2003a).The advantage of the novel differential deflection method is that the full coupling of the hydrodynamic and elastic deflection equations is made possible by using this formulation where the two equations are solved as a coupled pair of differential equations.

5.2.2.4.1 Differential deflection method

The elastic deflection equation is utilized in the differential form developed by Evans and Hughes, (2000). The evaluation of surface deflection is needed in order to determine the film thickness in the lubricated contact analysis. This can be achieved by solving equation (5.16) which contains the elastic deflection which is a convolution integration of pressure multiplied by the appropriate weighting function:

$$u(x) = -\frac{4}{\pi E'} \int p(s) \ln|x-s| \mathrm{d}s \tag{5.21}$$

The integral in equation (5.21) can be calculated numerically for any general pressure distribution such as that which happens in the contact between two moving roughness features. As a result of numerical discretization, this integral can be reduced to a simple quadrature by summation of influence coefficients multiplied by the pressure value at all points of the solution domain (Evans and Hughes, 2000) ,which is given by the following expression:

$$u(x_i) = \sum_k g_{k-i}\, p_k \tag{5.22}$$

The influence coefficients g_{k-i} in equation (5.22) depend on the contacting surface material elastic properties and the quadrature form used for the pressure.The technique of evaluating the influence coefficients was developed by Kong (2001). Evans and Hughes, (2000) developed the conventional deflection equation to the form given in equation (5.22) for the second derivative of *u*. Adopting this method leads to localisation of the effect of pressure on the deflection calculation in comparison with the classical method described by equation (5.22). The differential form of the deflection is written as:

$$\frac{\partial^2 u(x_i)}{\partial x^2} = \sum_k f_{k-i}\, p_k \tag{5.23}$$

Where the f_{k-i} are defined as the differential influence coefficients. The influence coefficients in this quadrature formula are available in closed forms in Evans and Hughes, (2000). The advantage of such a modification can be seen by comparing the differential influence coefficients (*g* and *f*), where the *f* coefficients decay significantly faster than the *g* coefficients as the indices *j* increase from zero. This shows that the impact of pressure on the second order derivative of the deflection is highly localised. With the differential

deflection method, the film thickness equation (5.16) takes on the following differential form:

$$\frac{\partial^2 h}{\partial x^2} = \frac{\partial^2 \phi}{\partial x^2} + \frac{1}{R'} + \sum_k f_{k-i}\, p_k \tag{5.24}$$

5.2.2.4.2 The coupled method

The coupled method is used so that the Reynolds and elastic film thickness equations can be numerically solved as a set of simultaneous equations. Discretised equations (5.14) and (5.24) can be written as:

$$\sum_{k=0}^{n_c} A_k\, p_k + \sum_{k=0}^{n_c} B_k h_k = R_i \tag{5.25}$$

$$\sum_{k=0}^{n_c} C_k\, p_k + \sum_{k=0}^{n_c} D_k h_k = E_i \tag{5.26}$$

In equation (5.25):

p_k and h_k are the values at the neighbouring mesh points.

A_k are the Reynolds equation coefficients of pressure.

B_k are the Reynolds equation coefficients of film thickness.

n_c is the number of coefficients involved in the formulation.

$R_{i,}$ is the right hand side, that has a zero value for steady state conditions.

The coefficients A_k and B_k involve the fluid properties of density and viscosity i.e. ρ, η and S in equation (5.14). The value of viscosity is highly pressure dependent and to solve the equation the fluid properties are obtained at each mesh point using pressure and film

thickness distribution for an outer loop. Equation (5.25) and (5.26) are set up using the current outer loop values to determine the A_k and B_k values.

In equation (5.26) E_i , includes the near pressure contributions to the discretised equations which are included in the numerical calculation, and the other contribution are evaluated for the outer loop of pressure distribution. The discretised equations which are defined by the Reynolds and elastic film thickness can be then written in terms of the two key variables, pressure and film thickness, as a pair of simultaneous equations in the variables p_0 and h_0 in the form :

$$A_0 p_0 + B_0 h_0 = \hat{R}_i \tag{5.27}$$

$$C_0 p_0 + D_0 h_0 = \hat{E}_i \tag{5.28}$$

The coupled solution method solves this pair of equations simultaneously for each node in the computational mesh. There are two methods available for solving this pair of equations. Either equations (5.28) and (5.29) are solved iteratively using Gauss Seidel or a direct elimination method can be used. In the current research equations (5.27) and (5.28), are solved iteratively as a pair of simultaneous equations for p_0 and h_0 so that the 'new' values of p_0 and h_0 can be calculated from the following expressions:

$$p_0^{new} = \frac{\hat{R}_i D_0 - \hat{E}_i B_0}{A_0 D_0 - B_0 C_0} \tag{5.29}$$

$$h_0^{new} = \frac{\hat{E}_i A_0 - \hat{R}_i C_0}{A_0 D_0 - B_0 C_0} \tag{5.30}$$

These equation are applied to each point in turn to obtain a converged solution for the time step. Convergence is rapid with a relatively high under-relaxation factor possible.

In some contact situations in the iterative time step scheme the value of h_0^{new} obtained from equation (5.30) is negative at some mesh points, when this occurs the fluid film is unable to separate the surfaces. The film thickness then has its value set to zero and equations (5.29) and (5.30) are replaced by

$$p_0^{new} = \hat{E}_i / C_0 \ , \ h_0^{new} = 0 \qquad (5.31)$$

This effectively sets a new boundary condition, $h = 0$ at the point and applies the Reynolds equation and deflection equation subject to that boundary condition. In this way the method determines where asperity contacts occur dynamically at each mesh point and in each time step. This technique is considered as a further advantage of the application of the coupled differential deflection approach. Every mesh point through each iterative sweep is assumed to be in a full film condition at each time step unless the iterating equations (5.29) and (5.30) result in a negative value for h_0^{new} at that mesh point. The approach of the coupled iterative scheme is able to deal with the difficulties in simulation of EHL contact problems between surfaces that have aggressive asperity features. These cause significant variation in both pressure and lubricant film thickness as will be seen in the next section.

When the equations (5.25) and (5.26) have been solved in this way the outer loop pressure and film thickness distribution are updated using a relaxation process. The A_k and B_k terms are then re-calculated and the equations are solved once more.

This sequence of operations is repeated until the outer loop pressure has converged, which means that the solution for the current time step has been obtained. The outer loop pressure and film thickness distributions then become the first outer loop value for the next time step.

The right hand side of equation (5.25) is calculated for the newly converged time step and the procedure to solve the next-time step then starts.

5.2.2.5 Numerical results of transient EHL contact analysis

Evaluation of the fatigue model considered depends on the results of the transient micro-EHL solutions for the fast and slow rough surface disks in the rolling/sliding contact. Results of the micro-EHL modelling, which gives the full time-varying behaviour of lubricant pressure and film thickness, have been used to predict the fatigue life and accumulated damage at the scale of surface asperity features.

In the current research, the application of the coupled differential deflection method, developed by Tao et al., (2003) and Holmes et al., (2003a, 2003b), was adopted for the analysis of micro-EHL in test rough gear surfaces. This was used as part of a study of micropitting and fatigue damage accumulation calculations, with and without including the residual stress. A pair of run-in profiles was selected for simulating the "rough on rough" rolling/sliding contact conditions and they are shown in Figure 5.3. The roughness profiles were taken from profilometer traces from run-in axially finished crowned disks. Tests were run using these two disks running at different peripheral speeds of 0.798 m/s and 0.479 m/s. This corresponds to a slide-roll ratio of 0.5. These representative profile sections were repeated and extended by creating longer multiprofiles with the joins between the repeated representative profiles made at deep valley features to ensure that no new artificial asperity features were created in the multi profile. The purpose of using multiprofiles is to reproduce all the asperity interactions for the two surfaces during the Micro-EHL contact simulation.

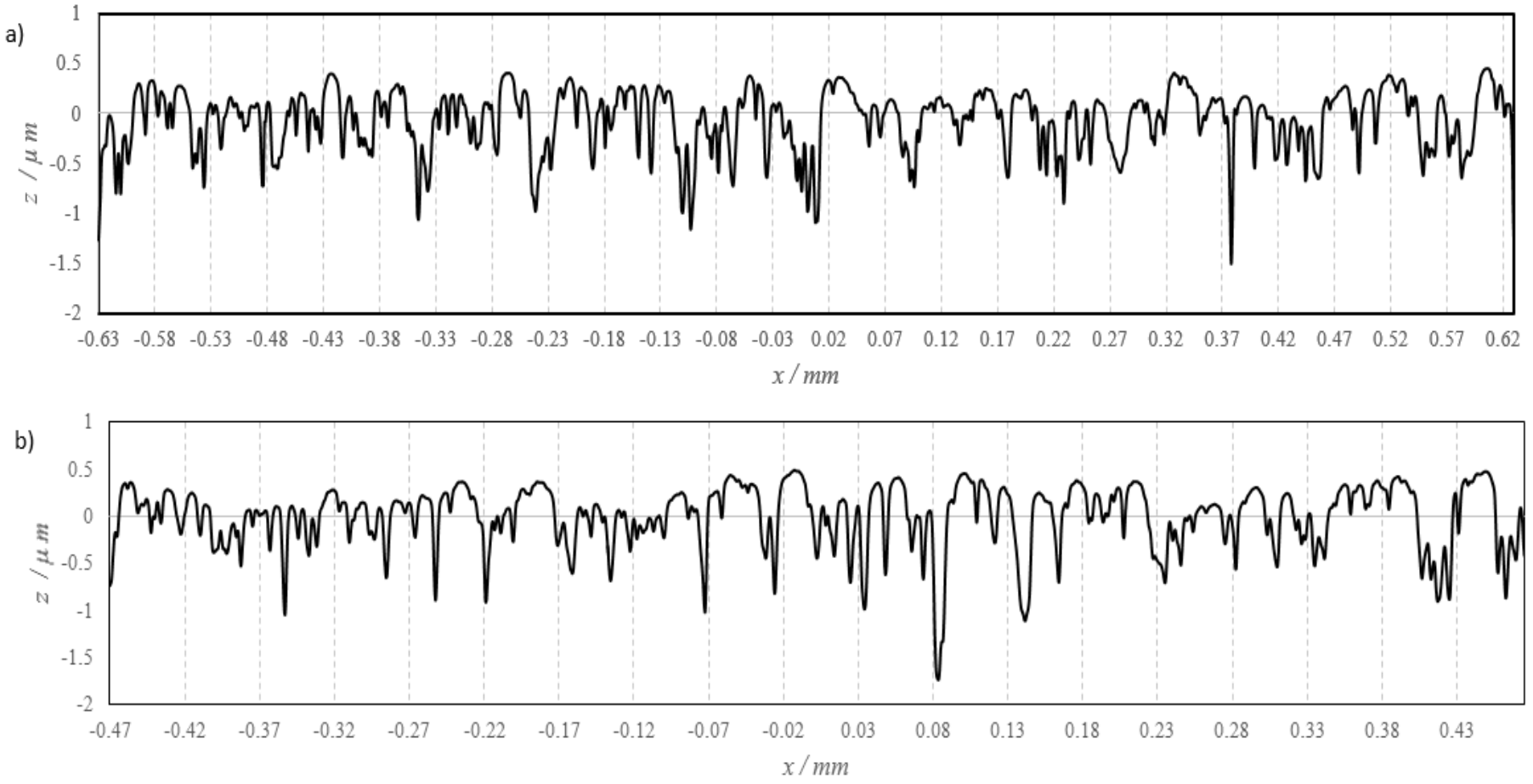


Figure 5-3 the representative profile section used in EHL simulations: a) Slow disk and b) Fast disk.

The results of the micro-EHL simulation are transient evaluations of lubricant pressure, surface shear stress and film thickness in the contact zone. The surface shear stress at direct asperity contacts is found from the asperity contact pressure. The time-dependent pressure and shear stress results are then used to predict the surface loading experienced by both contacting components. The transient analysis starts from the corresponding smooth surface steady state solution as an initial condition. The roughness profiles are then gradually fed into the contact from the inlet boundary. For each time step a converged solution of the pressure and film thickness with the current surface geometry is obtained. The lubricant pressure and film thickness values for the previous time step are used as starting values for the next time step during the solution process. An evaluation of the time varying surface and subsurface elastic stress distribution is then undertaken to build up the stress history for the disk material. This is a post processing step for the converged EHL analysis. The problem parameters assumed in the analyses are shown in Table1, which correspond to the operating conditions in the experimental test.

Table 1 Operating conditions assumed in micro-EHL simulation

Fast surface peripheral velocity(ms^{-1})	0.798
Slow surface peripheral velocity(ms^{-1})	0.479
Maximum Hertzian contact pressure (GPa)	1.4
Lubricant viscosity (Pas)	0.0257
Eyring shear stress (MPa)	10.0
Young's modulus (GPa)	207
Poisson's ratio	0.3
Hertzian contact dimension, a (mm)	0.469
Radius of surfaces 1 and 2 (mm)	38.1

The EHL simulation results of the steady state smooth surface are shown in the Figure 5.4 .This figure shows the pressure, film thickness distribution, in-contact cavitation and deflected smooth surface profiles for the smooth surface steady state solution. The mesh size adopted in the micro-EHL simulations is $\Delta x = a/200$, where a is the Hertzian contact semi-dimension (a =0.469 mm). The inlet zone of the contact is to the left hand side of the plot and the outlet area of the contact corresponds to the right hand side of the graph. It is clearly seen from this graph that the lubricated Hertzian contact area is divided into three regions, firstly, the inlet region which ends at approximately $x/a = -1$, where the lubricant is entrained and hydrodynamic pressure is generated. Secondly, the Hertzian zone that starts at $x/a = -1$ and ends just before $x/a = 1$. In this area the two surfaces are parallel and separated by an essentially constant film thickness. Finally, the outlet region, which starts at the film constriction and pressure spike before $x/a = 1$. Here the film pressure tends to push the lubricant flow toward the outlet region. Overall, these results indicate that the amount of lubricant within the EHL contact zone is managed by the inlet and continuity of lubricant can only be kept if there is a local restriction in the outflow, which causes a constriction to occur near the outlet. This constriction and pressure spike are close to the outlet region of the contact ($x/a > 1$) but are difficult to see on the graph due to the scale used. The smooth surface solution in Figure 5.4 includes all the key features of an EHL contact simulation as explained by Dowson and Higgingson (1966).

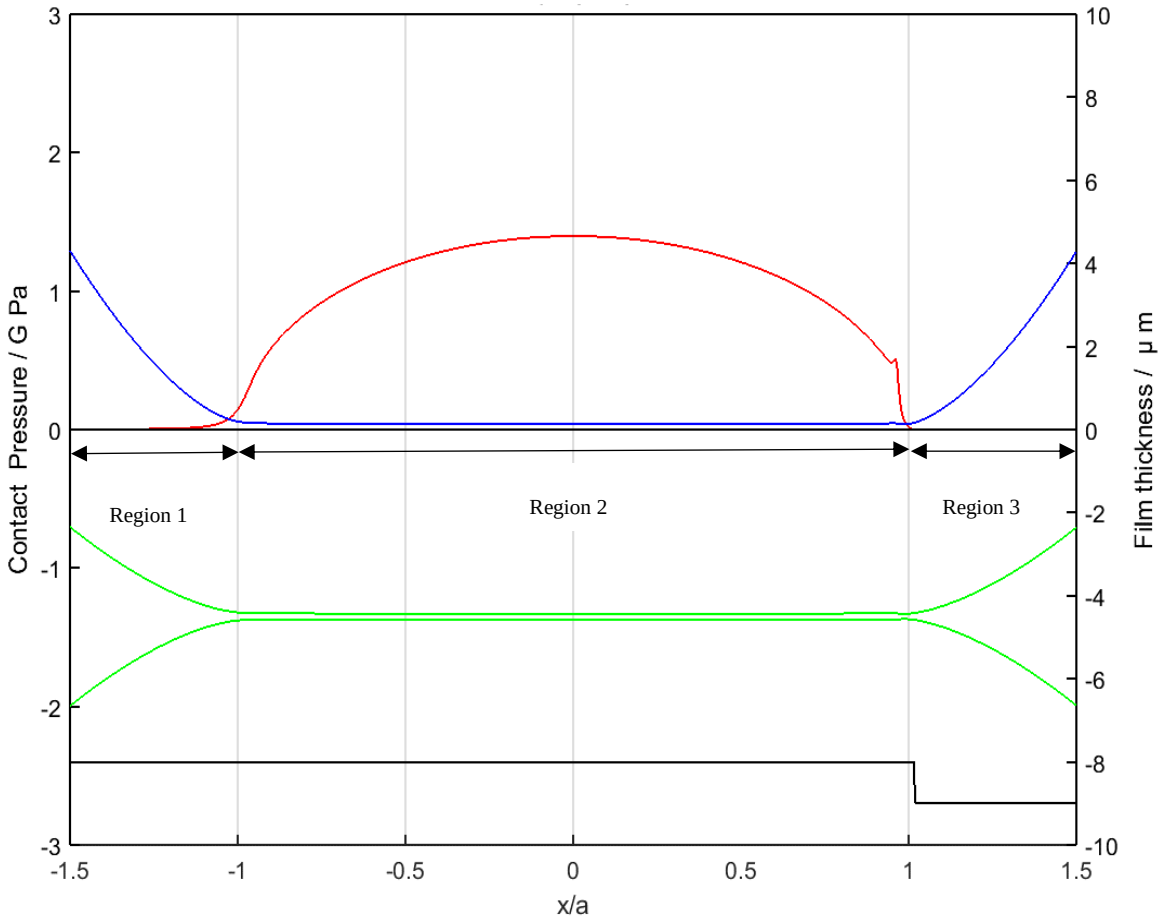


Figure 5-4 Steady state smooth surface solution at a typical timestep: Pressure distribution (red curve), film thickness distribution (blue curve), The black curve indicates full film conditions(upper position) or cavitated film conditions(lower position) at each mesh point and deflected surface profiles (green curves).

Figure 5.5 shows the pressure, film thickness distribution and deflected rough surface profiles at a single time step during a transient simulation of two roughness profiles. This is an example timestep when the roughness has passed through the contact zone. The boundaries for the calculation are x = -2.5 a and x =1.5 a , the profiles are offset for clarity so that the relative magnitudes of the surface roughness asperities and the film thickness can be appreciated. The pressure, p, at micro asperity contacts is clearly much higher than the Hertzian semi-ellliptical pressure and in this time step high extreme pressure spikes of 3, 4 and 3.25 GPa occur at positions x/a = - 0.8, - 0.23, 0.1, respectively, where the prominent surface asperities can be seen to be in close interaction. The black curve indicates the contact condition at each point in the mesh for the timestep. It has three levels corresponding to

cavitated film (lower level), full film(central level) and direct contact(upper level). For the smooth surface result in Figure 5.4, only the full film and cavitated conditions occur with cavitation occurring at the exit to the Hertizian zone. In Figure 5.5, there are three occurences of direct asperity contact of x/a = -1, x/a = - 0.8 and x/a = 0.3. The contact of x/a = - 0.8 is associated with an extreme pressure but the others are not.

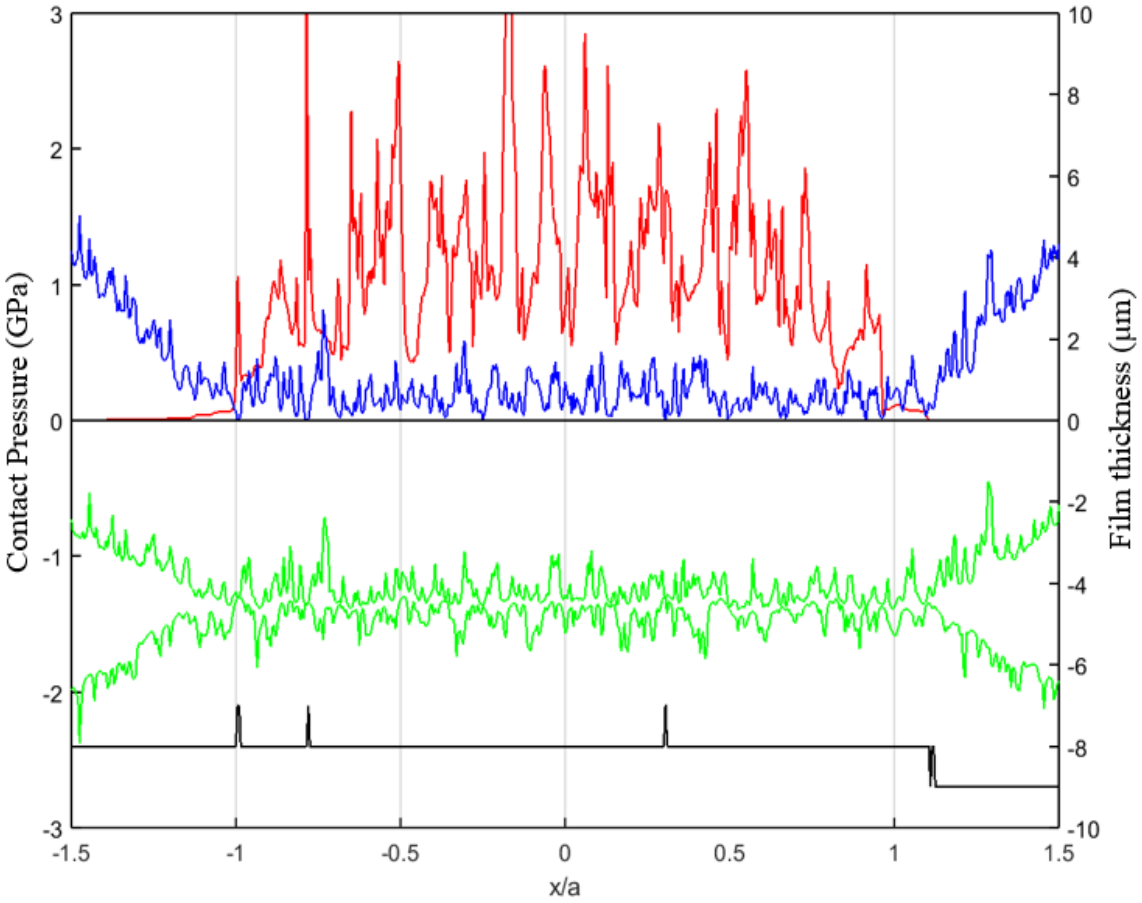


Figure 5-5 Illustration of a typical time step during the micro-EHL simulation: Pressure distribution (red curve), film thickness distribution (blue curve), the black curve indicates contact conditions (contact, full film, and cavitation), and deflected surface profiles (green curves).

A significant feature which is seen in micro-EHL solutions, is the phenomena of micro-cavitation which relates to cavitation at the roughness asperity level. An example of this is seen just befor the exit.

The EHL simulation results of the two rough surfaces in contact are shown in Figure 5.6 a. This graph shows the transient contact events happening during the micro-EHL analysis for the representative profile in a series of traverses of the contact zone. Contact between rough asperities in the EHL simulations were counted and accumulated relative to the surface roughness profile. In this way the number of time steps where contact happened was found for each point in the EHL profile. In this graph the results of five traverses are shown for the representative roughness profile. The roughness surface profile is shown as the lower of the curves in the figure with black colour. The five traverse count curves of surfaces are offset from each other by a 'count' value of 100 for clarity and they are aligned with the representative roughness section. The total number of time steps required for a point on the rough profile to pass the Herzian contact width is 800. Prominent asperities are seen to have high count values. Examples of this in Figure 5.6 (a) can be seen located at profile positions values of 274, 512 and 1164 μm. For other traverses the contact counts at those profile location are relatively lower. Figure 5.6 b, illustrates the corresponding count of the number of time steps where maximum contact pressure values $p > 3$ GPa occurred. The same prominent asperities are observed to experience high count values. The variability between traverse results are caused by the differences in the counterface asperity positions for each traverse.

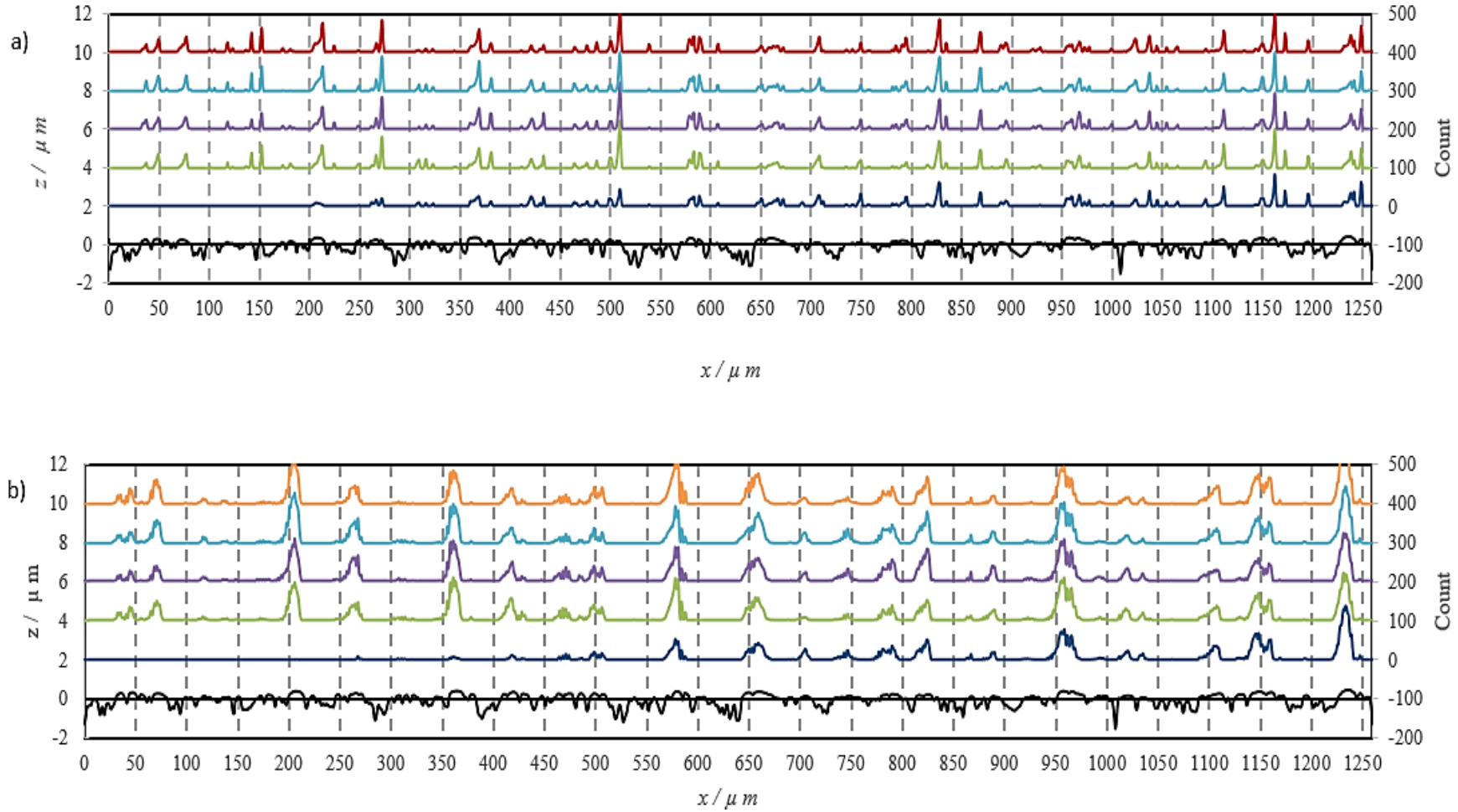


Figure 5-6 a) Profile contact count for five traverses of the contact zone and b) Profile high pressure count (p > 3GPa) for five traverses of the contact zone.

5.2.2.6 Contacting body stress analysis

This section describes the procedure for determining the surface loading history in terms of contact pressure and surface shear stress for stress analysis to determine material stress history at the surface and subsurface. The mixed EHL time step results are used to calculate the stress history at the lubricant/solid interface and for a block of the near surface material as it passes through the load bearing zone. At each timestep in the EHL analysis stored values of pressure, film thickness and surface traction are used to determine subsurface stress and strain history at each point in the representative block of material. The following convolution integrals are used to calculate the instantaneous distribution of subsurface elastic stress and strain.

$$\tau_{xz} = -\frac{2z^2}{\pi}\int_{p>0}\frac{p(s)(x-s)ds}{\left((x-s)^2+z^2\right)^2} - \frac{2z}{\pi}\int_{\tau>0}\frac{\tau(s)(x-s)^2 ds}{\left((x-s)^2+z^2\right)^2}$$
$$\sigma_z = -\frac{2z^3}{\pi}\int_{p>0}\frac{p(s)ds}{\left((x-s)^2+z^2\right)^2} - \frac{2z^2}{\pi}\int_{\tau>0}\frac{\tau(s)(x-s)ds}{\left((x-s)^2+z^2\right)^2} \qquad (5.32)$$
$$\sigma_x = -\frac{2z}{\pi}\int_{p>0}\frac{p(s)(x-s)^2 ds}{\left((x-s)^2+z^2\right)^2} - \frac{2}{\pi}\int_{\tau>0}\frac{\tau(s)(x-s)^3 ds}{\left((x-s)^2+z^2\right)^2}$$

The dimensions of the block are chosen to be 2.68 a parallel to the surface and a perpendicular to the surface. The mesh for the block of material is set up for stress evaluation in such a way that the finest mesh is used near the surface to capture the possible high stress gradients due to asperity contact and that significantly improves the accuracy of the numerical simulation results. In the current research, the blocks of material were discretized with a mesh of 201 uniform points in the x direction parallel to the surface as shown in schematic illustration in the Figure 5.7. In the vertical direction perpendicular to the surface the depth of the EHL block is divided into three different mesh density zones. The first zone

starts at the surface and has a uniform fine mesh with 31 grid points over $0 \leq z \leq 0.06\,a$. The second zone has a spacing of $0.005a$ over range $0.065a \leq z \leq 0.1\,a$. The third zone has a coarser spacing of $0.1a$.

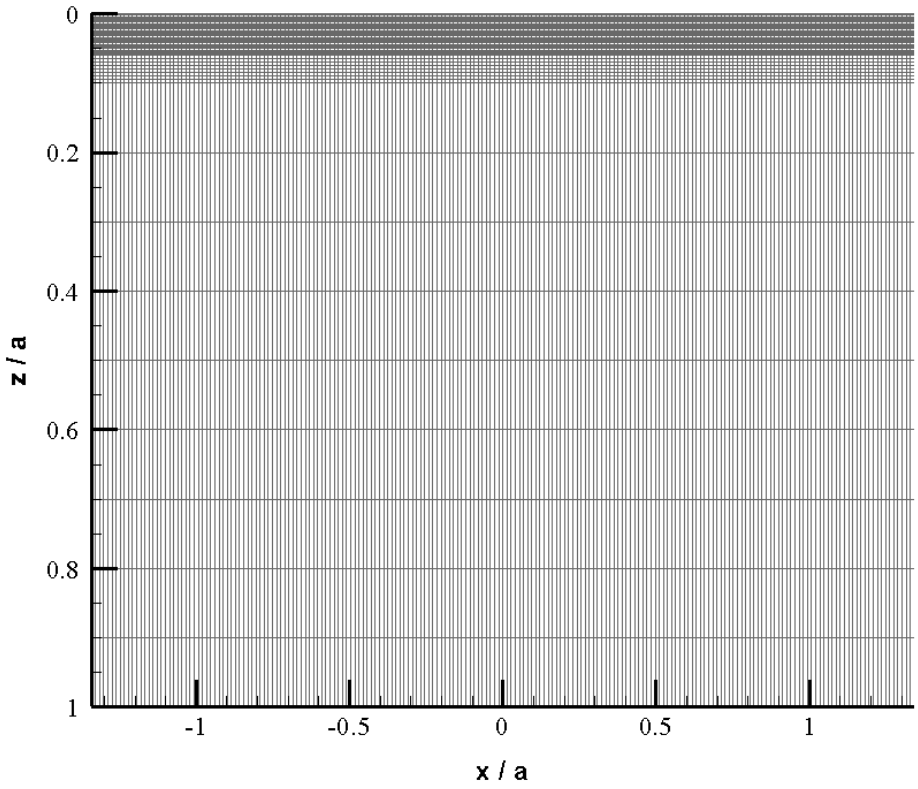


Figure 5-7 Schematic illustration of the material block mesh used in stress analysis.

The roughness irregularities can increase the local surface stresses to levels that are much higher than the maximum Hertzian pressure. This effect gradually decreases further down from the surface. This can be clearly observed in the results of stress analysis as shown in the Figure 5.8, its amplitude in the surface layer is evelated to the level comparable to its maximum alternating amplitude within the subsurface region. These observations are largely consistent with the earlier work done by (Sheng and Ahmet, 2011).

This result is significant at the subsurface level $z/a < 0.1$, where the high stress concentration is induced directly by the surface roughness. It has been believed that such localized extreme stress concentration is associated with the occurrence of the fatigue damage, in the form of

surface and subsurface micropitting, surface crack formation (Fujita and Yoshida, (1981);Yoshida et al., (1994); Yoshida and Konishi, (1995)). It can be seen from the Figures 5.9, 5.10 and 5.11 that *S11*, *S22* and *S12* all have negative values since the contacting bodies are under compression. Here $S_{11} = \sigma_{xx}$, $S_{22} = \sigma_{zz}$ and $S_{12} = \sigma_{xz}$. Basically,the compressive stress is greatly increased and its location moves toward the rough surface when the roughness feature is considered. Figure 5.10 shows σ_{zz} which is equal to the pressure at the surface. The subsurface stress and strain history at each point in a representative block of material will be used to search for the critical plane and assess the fatigue damage as will be discussed in more detail in the next section.

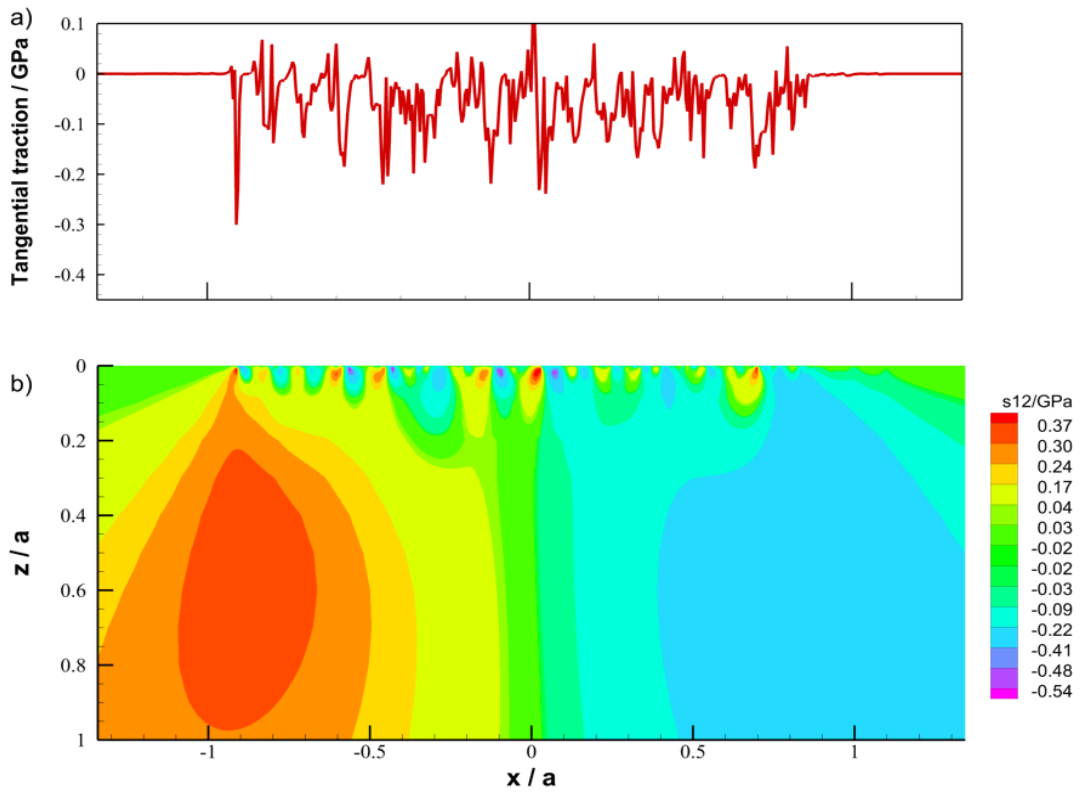


Figure 5-8 Rough surface contacts: a) Tangential traction; b) contours of horizontal stress component S12.

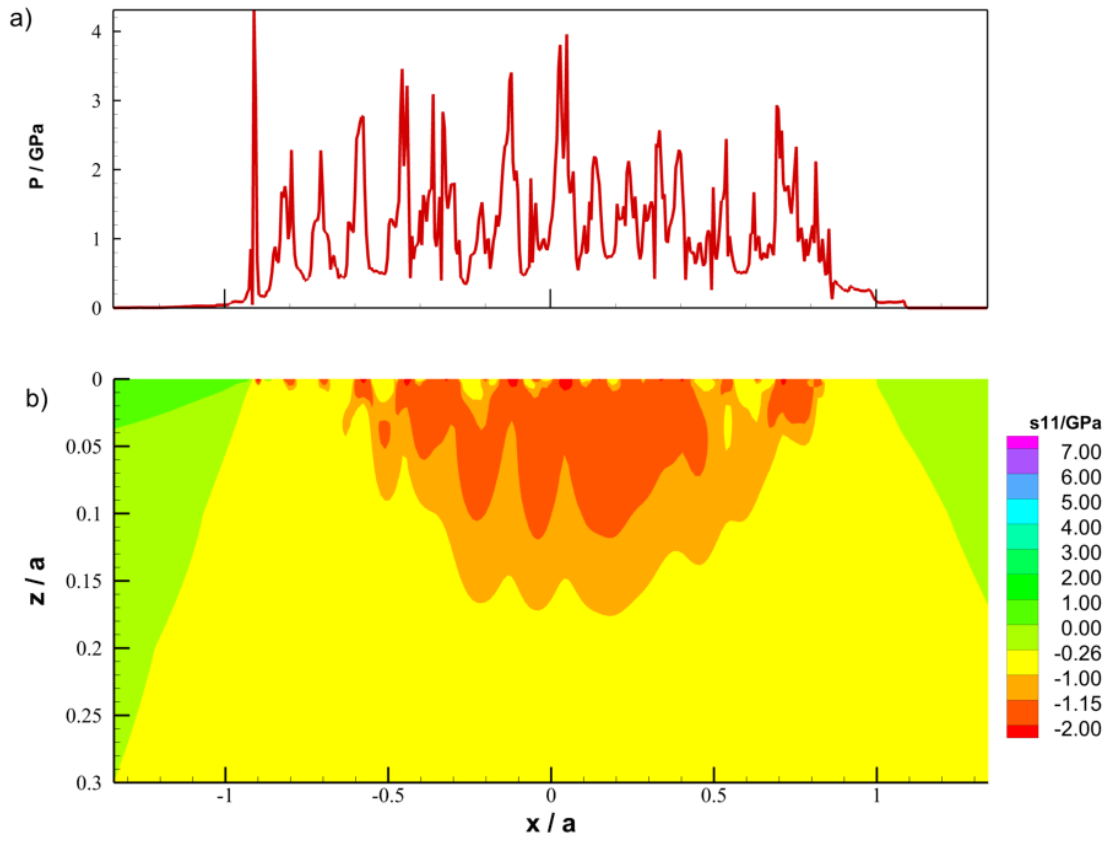


Figure 5-9 Rough surface contacts: a) EHL pressure distribution; b) contour of horizontal stress component S11.

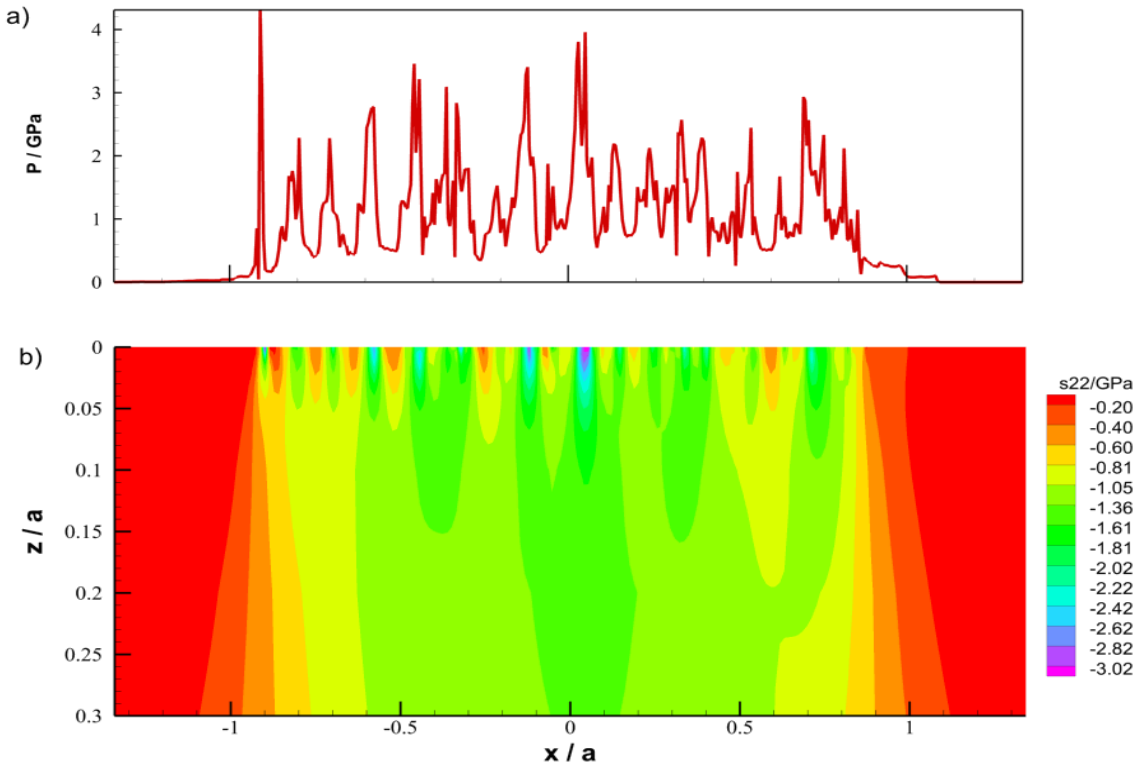


Figure 5-10 Rough surface contacts: a) EHL pressure distribution; b) contour of vertical stress component S22.

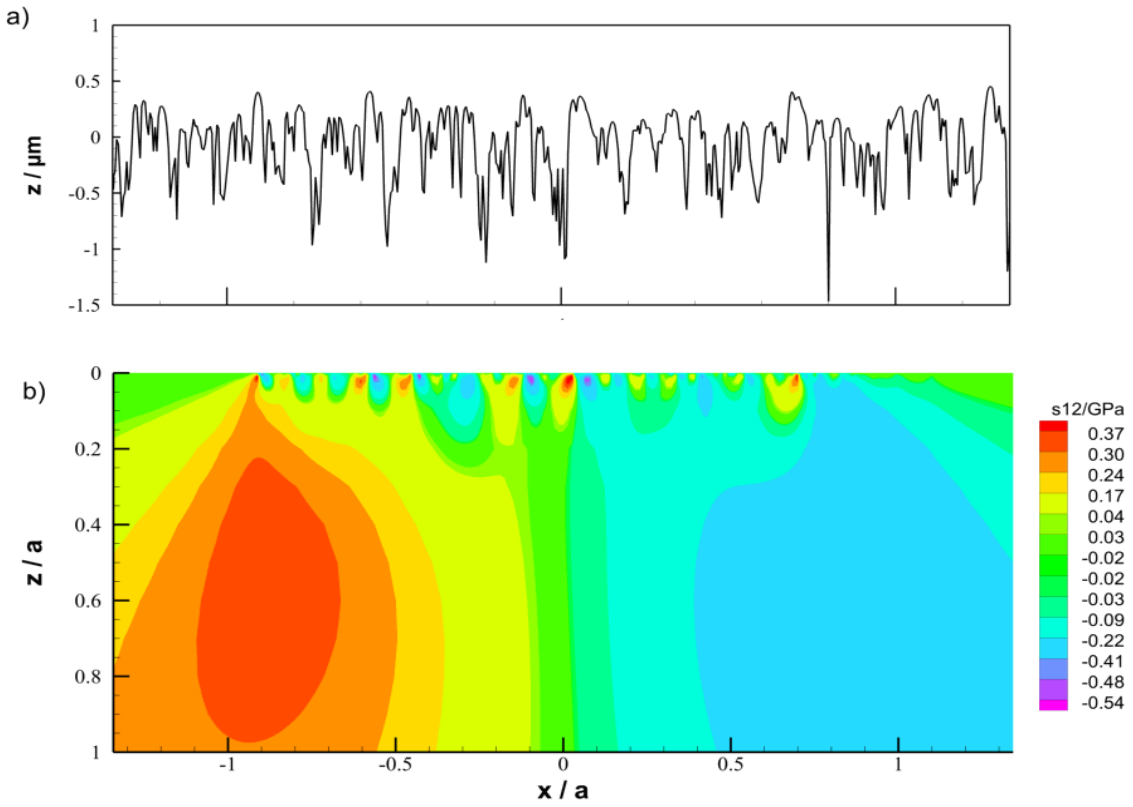


Figure 5-11 Rough surface contacts: a) EHL roughness profile; b) contour of horizontal stress component S12.

5.3 Fatigue and damage theory based on critical plane approach

Multiaxial fatigue failure criteria can be divided into three approaches i.e. stress, strain and energy-based criteria. Among these approaches, one criteria called the critical plane approach is considered significant. This plane is defined as the plane that is subjected to maximum damage. The Multiaxial fatigue failure criteria based on the critical plane approach reduces a multiaxial stress state to the equivalent uniaxial one. The aim of this approach is to combine the normal and shear stresses acting on a fixed plane within a material and to use these to identify the plane most vulnerable to fatigue failure. Findley (1956) introduced the term ” critical plane” and verified fatigue stress criteria based on the critical plane approach. Fatigue model criteria based on the critical plane approach are used to calculate fatigue life as well as fatigue fracture plane orientation. Three criteria based on the critical plane approach, the Findley criterion, the Matake criterion, and the Dang Van

criterion, are applied to the Micro-elastohydrodynamic line contact fatigue calculations in the current research.

Since many influencing parameters affect the fatigue analysis, there is no universally accepted multiaxial fatigue approach to achieve fatigue analysis (Fatemi and Socie, 1988). However, it has been reported that more accurate fatigue damage predictions are often achieved using a critical plane approach (Chu, 1995). Therefore, the critical plane approach is considered to be the most effective criteria and will be used in this chapter for analysing fatigue damage in mixed lubrication line contacts. The concept of the critical plane approach is based on the hypothesis that the fatigue crack initiates in slip systems of maximal shear stresses and, in the short crack stage, it propagates along these planes which have normal stresses that open the crack tip (Bannantine and Socie, 1992).

The criteria of Findley, Matake, and Dang Van are based on the assumption that the fatigue process is driven by a linear combination of a measure of shear stress, *A*, and a measure of normal stress, *B*. Fatigue can be expected to occur in a set number of cycles if this linear combination exceeds a fixed level. They can be written in the general form of as

$$A + \bar{\kappa} B \leq \bar{\lambda} \tag{5.33}$$

where the condition corresponds to a known probability of enduring 10^7 cycles without fatigue occurring. Parameters $\bar{\kappa}$ and $\bar{\lambda}$ are material coefficients and take different values for the different criteria. They are determined from the endurance limits under fully reversed bending and fully reversed torsion tests so that equation (5.33) is an equality when applied to those tests. The parameters that are used in these criteria are given in Table 2, where τ_a is the shear stress amplitude, $\sigma_{n,\max}$ is the maximum value of the normal stress on the plane, and σ_h is the hydrostatic stress. σ_{af} and τ_{af} are the endurance limits under fully reversed bending and fully reversed torsion tests, respectively.

Table 2 Expressions of parameters for criteria based on the critical plane approach

Criterion	A	B	$\overline{\kappa}$	$\overline{\lambda}$
Findley	τ_a	$\sigma_{n,\max}$	$\frac{2-\sigma_{af}/\tau_{af}}{2\sqrt{\sigma_{af}/\tau_{af}-1}}$	$\frac{\sigma_{af}}{2\sqrt{\sigma_{af}/\tau_{af}-1}}$
Matake	τ_a	$\sigma_{n,\max}$	$\frac{2\tau_{af}}{\sigma_{af}}-1$	τ_{af}
Dang Van	$\tau_a(t)$	$\sigma_h(t)$	$3\frac{\tau_{af}}{\sigma_{af}}-\frac{3}{2}$	τ_{af}

These criteria suggest that the critical plane orientation coincides with the maximum shear stress amplitude, where the maximum value of this linear combination occurs.

There are two aspects to be considered when applying the critical plane approach to estimate fatigue lifetime under multiaxial approach. Firstly, determining the orientation of the critical plane based on parameters *A* and *B*. Secondly, to assess the fatigue behavior using Equation (5.33). From Table 2, it can be clearly seen that calculation of the amplitude and mean value of the shear stress acting on the plane under consideration is needed to apply these fatigue criteria.

Since the transient micro-EHL solution is a plane problem, only the *x*-*z* plane is examined in the current research. As shown in Figure 5.12, the normal and shear stress history on the critical plane through a material point can be calculated at time (t) from the directional components by using Timoshenko and Goodier, (1973).

$$\sigma(t)=\sigma_x\cos^2\theta+\sigma_z\sin^2\theta+2\tau_{xz}\sin\theta\cos\theta \tag{5.34}$$

$$\tau(t)=\tau_{xz}(\cos^2\theta-\sin^2\theta)+(\sigma_z-\sigma_x)\sin\theta\cos\theta \tag{5.35}$$

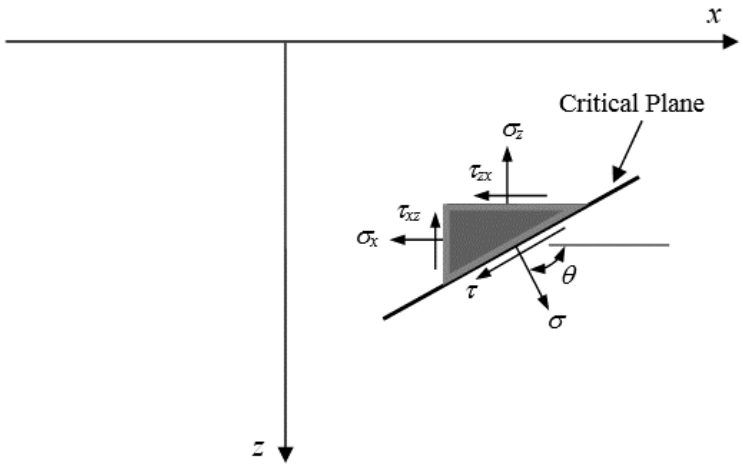


Figure 5-12 the orientation of the critical plane

Consider a candidate plane (the critical plane is not known) which is defined by θ in Equations (5.34) and (5.35) which is the angle between the normal to the plane and the *x*-axis. From the stress history of each material point the amplitude and the mean shear stress on the plane are found by

$$\tau_a = \frac{\tau_{\max} - \tau_{\min}}{2} \tag{5.36}$$

$$\tau_m = \frac{\tau_{\max} + \tau_{\min}}{2} \tag{5.37}$$

where $\tau_{\max}$ and $\tau_{\min}$ are the maximum and minimum values of the shear stress on the plane, respectively obtained during the loading history in the orientation,θ, under consideration. After parameters *A* and *B* are calculated the critical plane at a point is the plane which has the maximum value for $A + \bar{\kappa}B$. The next step is to evaluate the fatigue failure condition and a fatigue parameter (*FP*) can be determined from equation (5.38) as

$$FP = \frac{A + \bar{\kappa}B}{\bar{\lambda}} \tag{5.38}$$

Thus when the *FP* in equation (5.38) is greater than unity, fatigue failure can be expected to occur with corresponding probability .

For the Dang Van fatigue criterion, since both A and B use current stress values, the fatigue parameter is

$$FP = \max_{t}\left(\frac{A + \bar{\kappa} B}{\bar{\lambda}} \right) \tag{5.39}$$

The critical plane for each particular case can be determined by maximisation of the right-hand side of equation (5.39).

Since the critical plane through a given material point is not known before the analysis, *FP* is calculated on all the candidate planes in order to establish the critical plane. The above analysis methology has to be repeated for all the material points to be analysed.

5.4 Interpolation of residual stress

In order to use the residual stress calculated in the Abaqus FEA analysis for consideration in the fatigue calculation, it is essential to transfer the residual stress field calculated for surface asperities using the ABAQUS finite element (FE) package to the fatigue calculation. Such residual stresses which have developed in the material may have a significant effect on the fatigue life. This residual stress field is combined with the stress resulting from EHL analysis to give an improved indication of damage and to assess surface fatigue life. This is done by using an interpolation procedure that enables each asperity feature to have the correct residual stress distribution according to the loading that gives the best fit to the measured residual deflections. The stress information for Abaqus could only be plotted using the contour plotting tool with in Abaqus. The requirement was to interpolate this stress field onto the mesh which would be used for the evaluation of stress based on the EHL model's

surface loading from timestep to timestep. The EHL mesh used had a spacing of $a/200$ which was 2.45 μm for the contact load adopted.

The interpolation process involves two steps:

(a)- Expressing each Abaqus stress point position (x,z) in terms of x (unchanged) and a modified z, the distance from the rough surface.

(b)- Interpolating each component of the Abaqus stress field from the new x,z coordinates to the regular mesh of x,z coordinates adopted for the EHL stress evaluations.

Step (a) is referred to as, flattening and is illustrated using Tecplot's triangulation process for comparison purposes only.

The best settings for interpolation from the non-structured Abaqus mesh to the structured EHL fatigue calculation mesh have been determined by a lot of careful evaluation and checking using a typical Abaqus stress field. This process includes two steps defined above. Step (a) is to extract the residual stress field at each FEA node by tabulation within the Abaqus system. This table also gives the nodal coordinates for each of the mesh points in the residual unloaded position at the end of the analysis FEA. The coordinates of the surface points are extracted separately from Abaqus using the "path tool". The surface position is established for each x coordinate required by linear interpolation of the surface points. This z value is subtracted from the z coordinate of the stress point to achieve the objective of step (a). Figure 5.13 shows the rough surface flattening process, i.e. the form and residual roughness is removed and the data remains on an unstructured mesh.

The second step is to interpolate from the unstructured mesh to the uniform mesh to be used for fatigue failure analysis. An existing Gaussian interpolation software programme was developed to accomplish this process in the current research. This code finds the

interpolated value at a mesh point in the regular mesh by considering a circle centred at the point and considering the values of the stress field component at each of the unstructured mesh points that are located within the circle. These contribute to the interpolated value according to the reciprocal of their distance from the regular mesh point. The radius of the interpolation circle is changed according to the fineness of the unstructured mesh. The radius is specified in terms of the finest resolution of the high density mesh and this is used for the interpolation along with a minimum number of points condition. This was a minimum of four points for the current analyses. If there are fewer than the minimum number of points specified for interpolation, the programme doubles the radius until the condition of the point is satisfactory. The interpolation programme was designed to differentiate between the low and high density of interpolation points and then a convenient interpolation circle radius was assigned for each zone (the high density mesh dimension and low density mesh dimension).

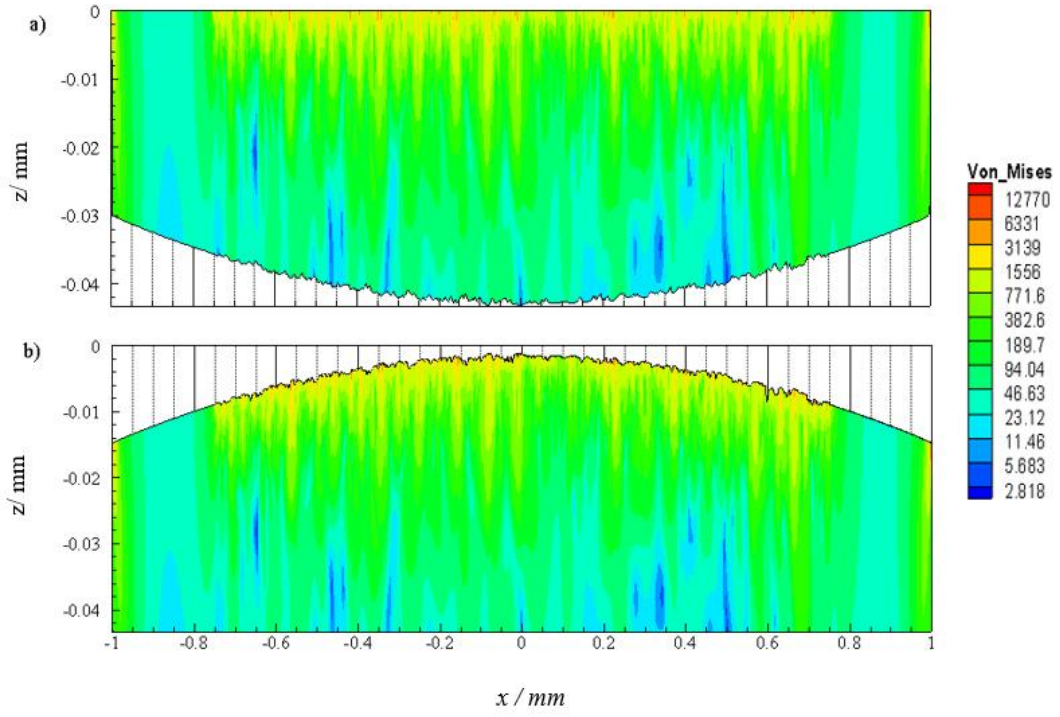


Figure 5-13 Rough surface flattening: a) model after flattening process b) model before flattening process, Von_Mises stress unit is MPa.

Figure 5.14 shows a comparison between the contour plot result gained by the Gaussian interpolation software programme and a contour plot of the von Mises stress as produced by the Abaqus FEA system using the calculated values at each mesh point, these figures should have the Abaqus result as (a) and the interpolated result as (b). Figure 5.14 a is for a 0.5 mm sample length of the surface and for 201 points in the *x*-direction in the moving material box inside the EHL analysis. In this figure the radius of interpolation for the fine and course mesh area was 2 μm and 3 μm , respectively. These settings were able to provide a good reproduction of the intensity value of stress and smooth contours in the area located between z = - 0.02 mm to z = - 0.044 mm when compared to the data as produced by the Abaqus FEA system in Figure 5.13 (b). However, it gives poor contours in the area which has a high mesh density in between z = 0.0 and z = -0.01 mm. In addtion, this plot does not capture the detailed information of the possible high stress gradients due to asperity contact near the surface that can be seen in Figure 5.14 a at x = - 0.01 mm, -0.08 mm, -0.16, -0.34 and at x = -0.41 mm. In order to achieve the appearance of fine detail on the interpolated version of the stress data in the area which has a high density of mesh near the surface, it is necessary to increase the number of a points inside the EHL mesh moving box to 401 and 501 as shown in Figures 5.15 and 5.16, respectively. The advantage of the latter technique is that more detail can be presented at the subsurface level and using approximately the same mesh size of EHL. In this way, the interpolation software programme can give smooth stress contours on the subsurface of the model as shown in Figure 5.15. This figure used 401 points in the moving box and using 2 μm and 3 μm radius of interpolation values for fine and course mesh, respectively. This can effectively capture detailed information about the stresses in the near surface area which has a high mesh density and examples of this can be seen at at x = - 0.01 mm, -0.08 mm, -0.16, -0.34 and at x = -0.41 mm. Figure 5.16 which used 501 points in the moving box was not able to pick up the values of a high stress component and give smooth contours in the same area which has a high density near the

subsurface level. Thus, to capture detailed stress data the current research used 401 points in the moving box and 2 µm radius for the FEA's densely meshed area and used the 3 µm radius for the low density area as shown in Figure 5.15.

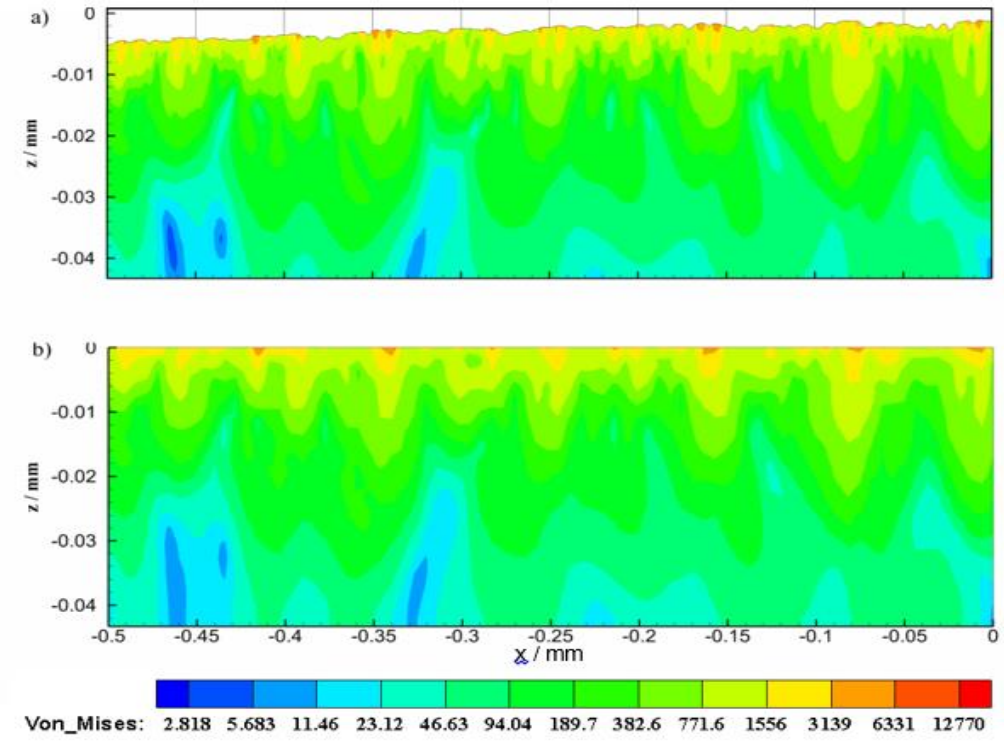


Figure 5-14 von Mises stress contours 201 points with radii of interpolation for fine and coarse meshes of 2 µm and 3 µm, respectively: a) Abaqus result, b) Interpolation result.

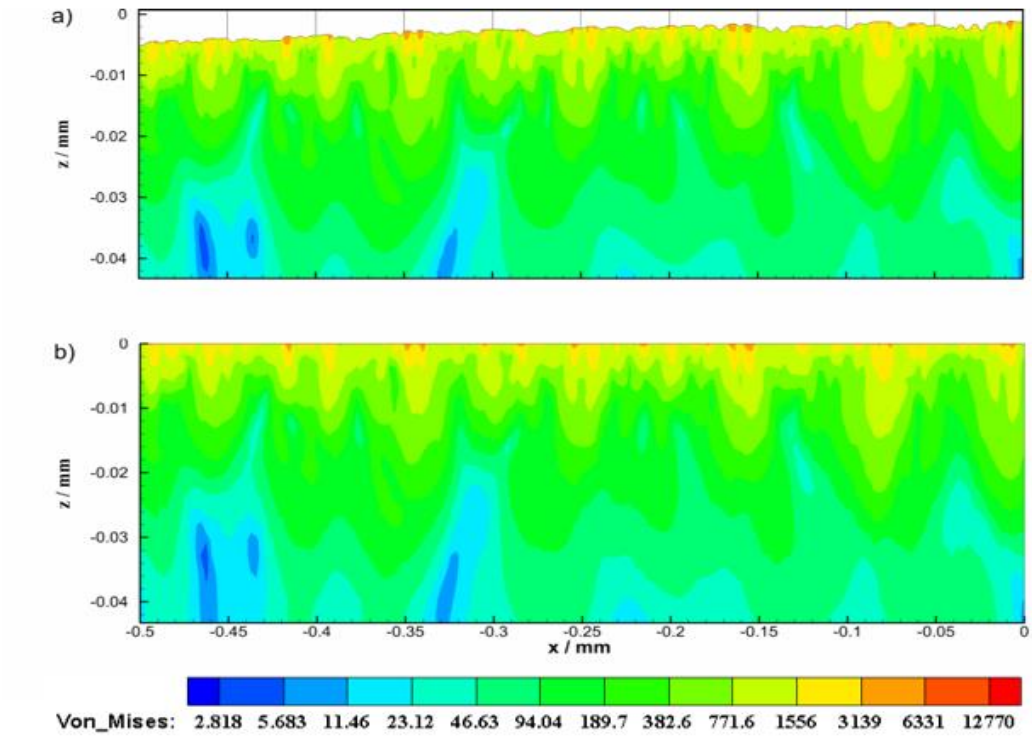


Figure 5-15 von Mises stress contours 401 points with radii of interpolation for fine and coarse meshes of 2 µm and 3 µm, respectively: a) Abaqus result, b) Interpolation result.

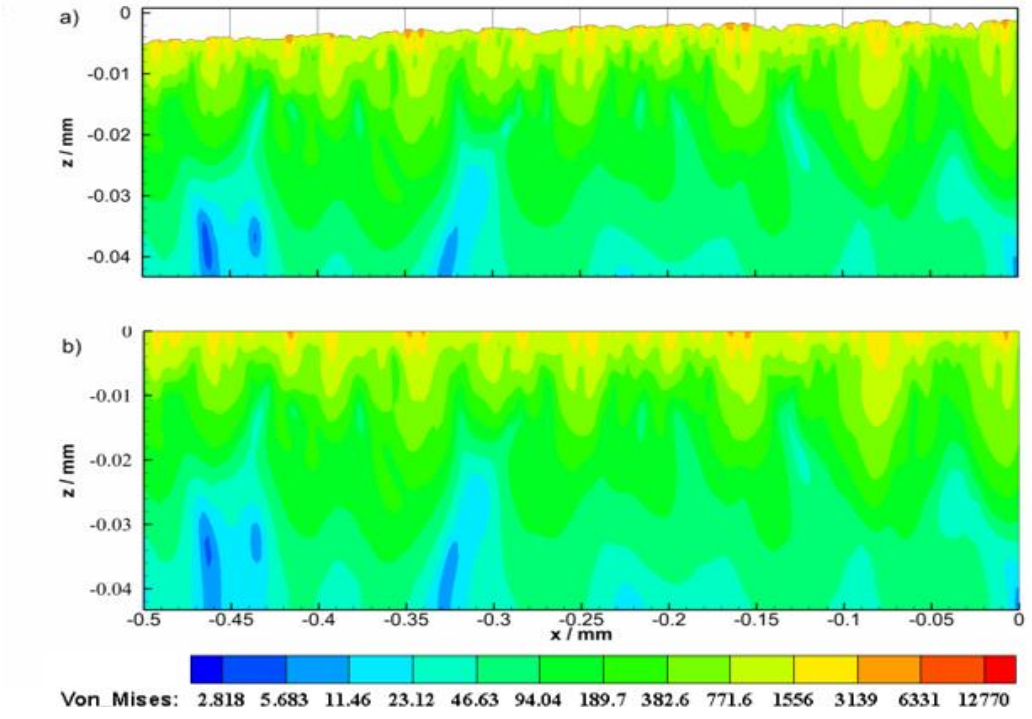


Figure 5-16 von Mises stress contours 501 points with radii of interpolation for fine and coarse meshes of 2 µm and 3 µm, respectively: a) Abaqus result, b) Interpolation result.

5.5 The Process of applying the critical plane model

The evaluation of any particular fatigue model using multiaxial fatigue criteria based on the critical plane concept involves using the stress history for the material as it passes through the contact area. In such a model if the *FP* is unity or above, fatigue failure is likely to happen after 10^7 loading cycles. The fatigue parameter is evaluated based on the stresses encountered during one pass of the representative volume through the EHL contact area. To evaluate the fatigue failure based on the critical plane approch, such as the Findley criterion, the Matake criterion, and the Dang Van criterion the following steps are taken for all the material points to be examined:

1. Read the stress history for all the evaluation points in the material block considered for EHL analysis.
2. For each evaluation point in turn use the stress component history to calculate the strain history by using Hook's law as follows :

$$\varepsilon_x = \frac{1}{E}\left\{\left(1-\nu^2\right)\sigma_x - \nu\left(1+\nu\right)\sigma_z\right\} \tag{5.40}$$

$$\varepsilon_z = \frac{1}{E}\left\{\left(1-\nu^2\right)\sigma_z - \nu\left(1+\nu\right)\sigma_z\right\} \tag{5.41}$$

$$\gamma_{xz} = \frac{\tau_{xz}}{G} = \frac{2\left(1+\nu\right)}{E}\tau_{xz} \tag{5.42}$$

3. Consider each potential critical plane orientation in turn which is defined by θ, the angle between the normal of the plane and the *x*-axis. Obtain normal and shear stress history, and in the same way calculate normal and shear strain on the critical plane respectively according to Timoshenko and Goodier (1973) as:

$$\sigma = \sigma_x \cos^2\theta + \sigma_z \sin^2\theta + 2\tau_{xz}\sin\theta\cos\theta \tag{5.43}$$

$$\tau = \tau_{xz}\left(\cos^2\theta - \sin^2\theta\right) + \left(\sigma_z - \sigma_x\right)\sin\theta\cos\theta \tag{5.44}$$

$$\varepsilon = \varepsilon_x \cos^2\theta + \varepsilon_z \sin^2\theta + \gamma_{xz}\sin\theta\cos\theta \tag{5.45}$$

$$\frac{\gamma}{2} = \frac{\gamma_{xz}}{2}\left(\cos^2\theta - \sin^2\theta\right) + \left(\varepsilon_z - \varepsilon_x\right)\sin\theta\cos\theta \tag{5.46}$$

4. Evaluate τ_a , $\sigma_{n,\max,}$ $\tau_a(t)$ and $\sigma_h(t)$ for the plane considered and calculate the fatigue parameters for each of the fatigue models.
5. *FP* has to be examined on all the candidate planes in order to identify the critical plane for the point considered for each model. The critical plane is defined as the material plane where $FP = \dfrac{A + \bar{\kappa}B}{\bar{\lambda}}$ is a maximum for that point.
6. Evaluate the fatigue failure using equation (5.38). If *FP* is greater than unity at a point, fatigue failure is likely to happen in N_f cycles at that point.
7. Produce contour plots of the *FP* values obtained at each point in the material.

5.6 Numerical results for critical plane models without and with residual stress

A number of fatigue analyses of the critical plane approach were used and compared for the EHL line contact with rough surfaces. The results of three such models of the fatigue process are presented here namely the Findley, the Matake, and the Dang Van criteria. The fatigue failure is postulated to occur for 10^7 loading cycles when the fatigue parameter (*FP*) is unity. The parameters for the fatigue model applied are given in Table 3.

Table 3 Parameters used for transient EHL line contact analysis.

K	Material constant in equation (5.29)	1.0
G	Shear modulus (GPa)	80
σ_o'	Yield strength for the cyclic stress-strain curve (GPa)	2.293
τ_f'	Shear fatigue strength coefficient (GPa)	1.15
γ_f'	Shear fatigue ductility coefficient	0.831
b	Fatigue strength exponent	- 0.091
c	Fatigue ductility exponent	- 0.6
σ_{af}	Fatigue limit for bending (MPa) ($N = 10^7$)	695
τ_{af}	Fatigue limit for torsion (MPa) ($N = 10^7$)	401
a	Hertzian semi-contact dimension (mm)	0.469
ξ	Slide roll ratio = (u1 − u2) *2 /(u1 + u2)	0.5
p_o	Hertzian contact pressure (GPa)	1.4

The parameters values of τ_f' and γ_f' were estimated from the fatigue strength coefficient σ_f' = 2 GPa and the fatigue ductility coefficient ε_f'= 0.48 for pure shear conditions as $\sigma_f'/\sqrt{3}$ and $\sqrt{3\varepsilon_f'}$, respectively,(Dowling 1998).

The EHL block material used for the two rough surfaces is SAE4340 (BS970:En24) whose hardness is HB = 409 and ultimate tensile strength is σ_u= 1470 MPa (Zahavi and Torbilo 1996). The steel used in the fatigue experiments was not the same, but it was assumed these parameters as typical values. In addition, its fatigue limit at $N = 10^7$cycles under fully reversed fatigue loading (bending) can be assumed by (Ciavarella and Maitournam, 2004)

$$\sigma_{af} = 1.7HB = 1.7 \times 409 = 695 \text{(MPa)}$$

and the fatigue limit under fully reversed torsional loading through each cycle can be approximated by (Dowling 1998)

$$\tau_{af} = \frac{\sigma_{af}}{\sqrt{3}} = 695 \times 0.577 = 401 \text{(MPa)}$$

The fatigue damage value can be considered progressive and localised over the material and its variation can be illustrated as a contour plot. The results reported in this section are based on two rough surface profiles for the fast and slow discs discussed in the experimental work. The two rough surface profiles run against each other in a transient EHL analysis so that comparisons could be made between the same portions of roughness profile as shown in the previous plot Figure 5.3. The slower rough profile is examined critically because its asperities are subject to a higher number of stress cycles during the EHL contact analysis than are those of the faster disk.

Figures 5.17 and 5.18 show the overall results for the slow disk surface for the Dang Van fatigue parameter, the Matake fatigue parameter and the Findley fatigue parameter without residual stresses and with residual stresses , respectively. Figure 5.17 b, c and d shows the contours of high values of the fatigue parameters without residual stresses for the three fatigue criteria. The rough surface profile is seen in the upper plot and the black dotted headed arrows labelling to the rough surface correspond to ten asperities which have highest *FP* values. It can be seen that high levels of damage occurred at the scale of the asperity

features at positions corresponding to aggressive asperity loading within 10^7 rotating cycles, such as depths of 0.07*a* (33 µm). It can be seen from Figure 5.17 b that the calculated fatigue parameters by using the Dang Van criterion are slightly lower than the Matake and Findley fatigue parameter. It can be noticed that the Matake criterion and the Findley criterion give almost the same results. In addition, they predict a more aggressive damage, because they are highly dependent on the shear stress, therefore, the roughness effect is more responsive by these criteria. Fatigue failure regions (in red colour) are those where fatigue parameter ≥ 1. It is clear that damage regions are concentrated near the surface of particular asperity features for the three criteria considered; at x/a = -0.57, x/a = -0.92, x/a = 1.11 and x/a = 1.28, for example. Figures 5.18 b, c, d shows the corresponding result when the Abaqus residual stress is included in the fatigue analysis. It was found that for some asperities the fatigue life for the models involving residual stress was shorter than the fatigue life obtained using the same model without the induced residual stress. However, there are differences of an order of magnitude in the calculated fatigue parameter for the three criteria adopted.

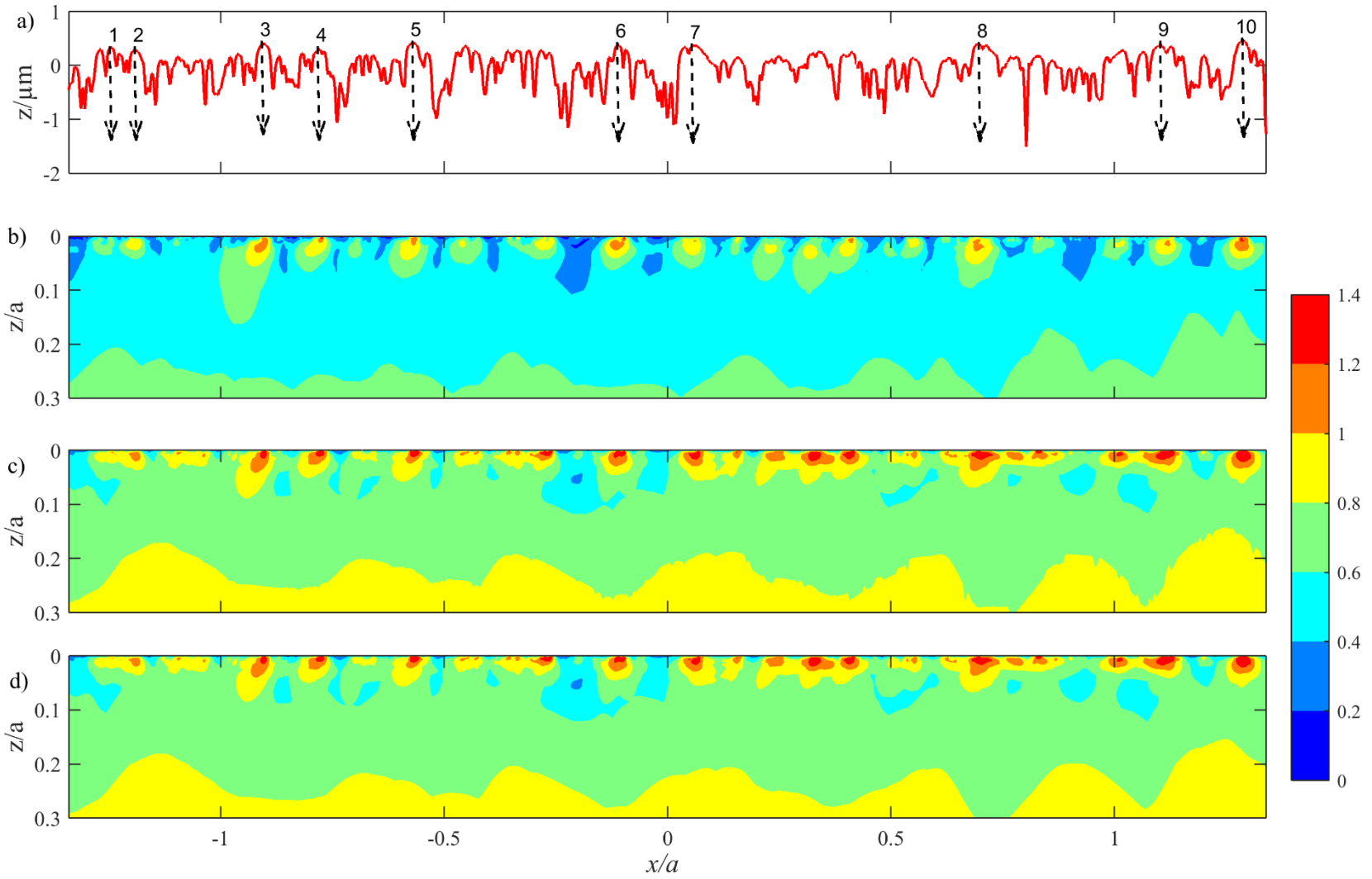


Figure 5-17 Contours of fatigue parameter (FP) for fatigue at 10^7 cycles – without residual stresses: a) EHL profile, b) Dang Van, c) Matake, and d) Findley criteria

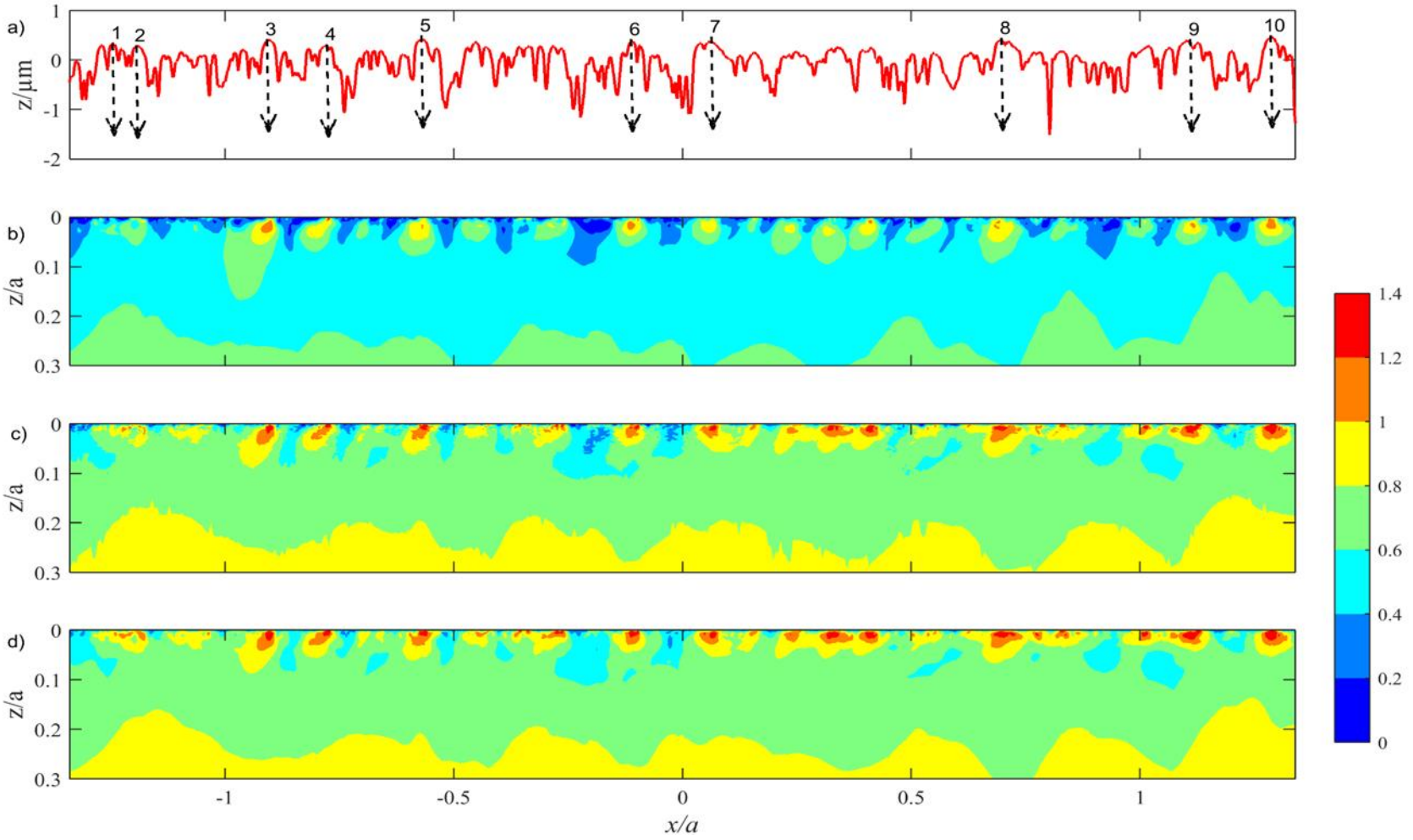


Figure 5-18 Contours of fatigue parameter (FP) for fatigue at 10^7 cycles – With residual stresses: a) EHL profile, b) Dang Van, c) Matake, and d) Findley criteria

Figures 5.19 through 5.30 present the fatigue parameter calculated with and without residual stress at a larger scale and more detailed contour comparisons for the three criteria considered. The roughness profiles which are seen in the upper part in the Figures 5.19 *a*, 5.23 *a* , and 5.27 *a* were termed as load stage 2 shown in black and load stage18 shown in red. Load stage 2 is following a further 30 seconds of loading of the two rough surfaces, the asperities during this stage experience comparatively little plastic deformation. In addition, the asperity features become almost uniformly flattened having rounded lands while the valley features remain relatively unchanged as high contact loads happen predominantly between interacting asperities. However, Load stage 18 appears to show a large amount of modification and it is thought that surface fatigue is the mechanism causing these prominent modifications at the level of valley regions, which are the result of the removal of material these can be most clearly identified at x/a = 1.28, x/a = 1.11, x/a = 0.69, x/a = 0.06, x/a = - 0.27 and x/a = -1.2. Some of the experimentally observed micropits seen in Load stage18 were found to occur at positions that had high calculated fatigue parameters. This shows that residual stress can have a significant effect on the fatigue and more detailed comparisons indicate that where this happens it tends to occur near the surface of the material. This observation is in good agreement with the fatigue results obtained from the critical plane approach models and with the fatigue experimental profile (Load stage 18). It has been shown clearly in all magnified sections the compressive residual stresses are often intentionally introduced into material to improve the fatigue strength at long lives and this is considered a protective environment. However, the risk comes from the high tensile residual stresses that increase the failure near the surface of aggressive asperity feature as shown clearly in the magnified section in Figure 5.20. Furthermore, deeper in the subsurface the material is hardly affected by the roughness, for example at depths of 0.2*a* where the damage is decreasing at this level.

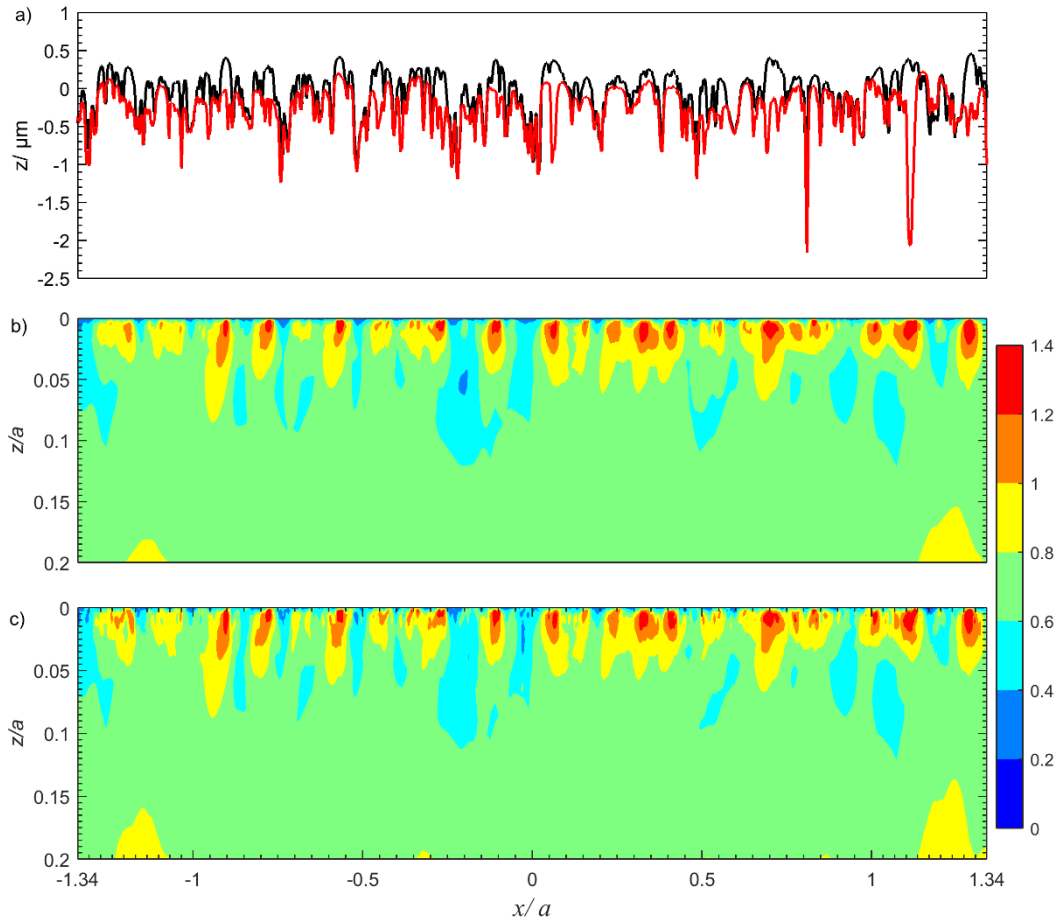


Figure 5-19 slow disk- contours of fatigue parameter (FP) at 10^7 cycles using the Findley Criterion, target load is 1.4 GP: a) Surface profile; EHL – black, LS18 (experimental) - red; b) Damage without residual stress, c) Damage with residual stress.

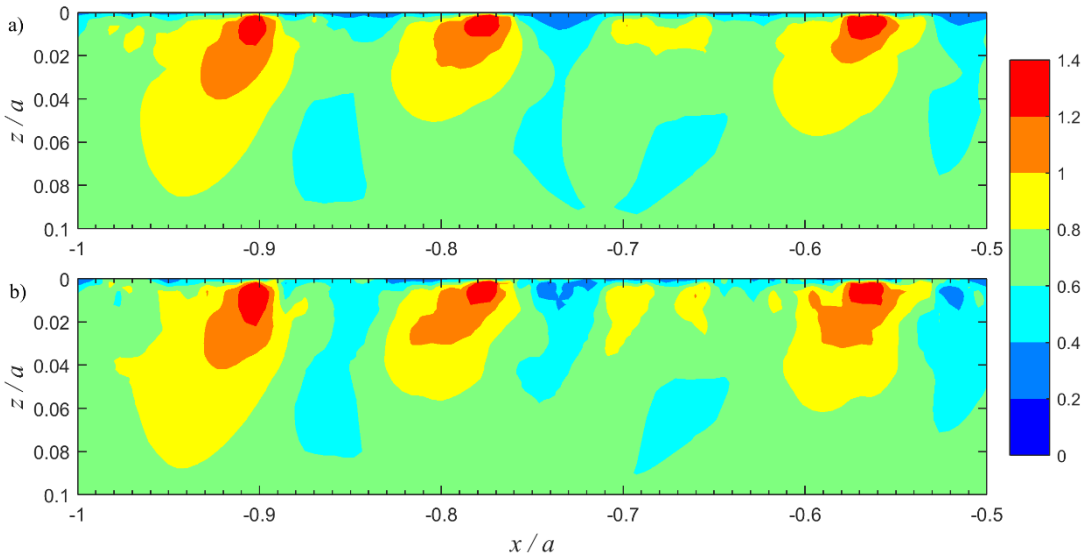
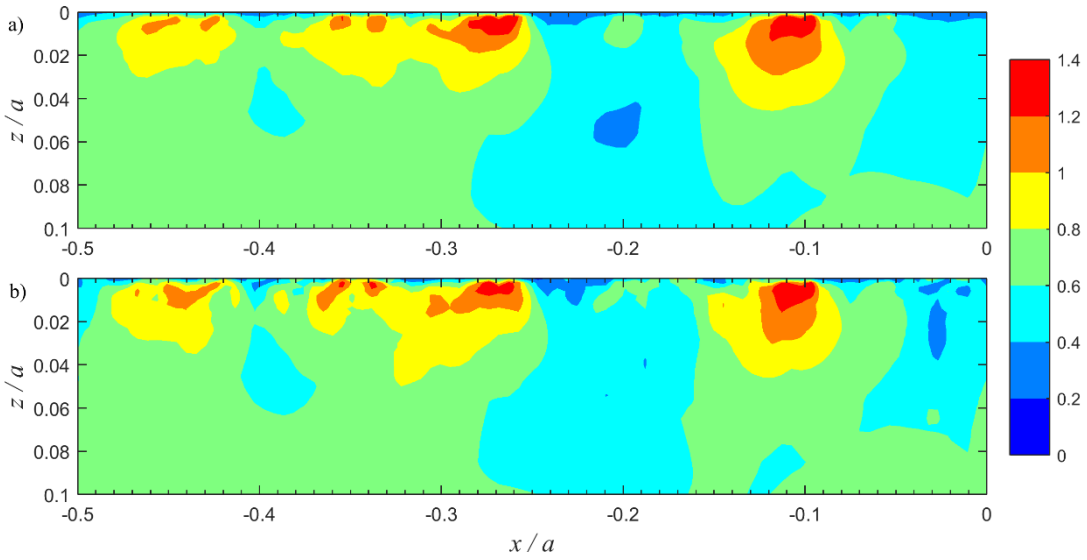


Figure 5-20 slow disk: Findley fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.



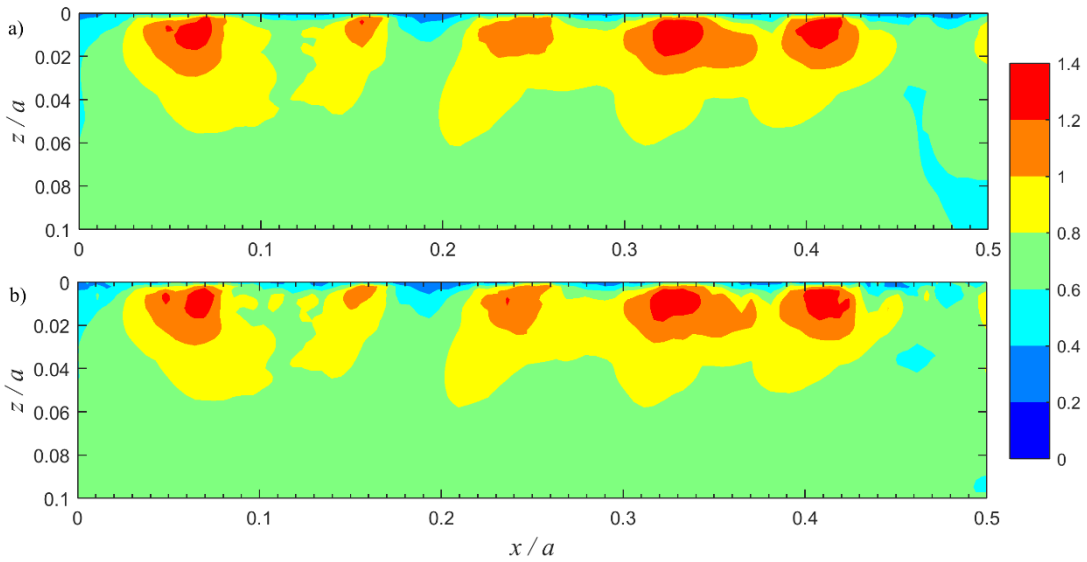
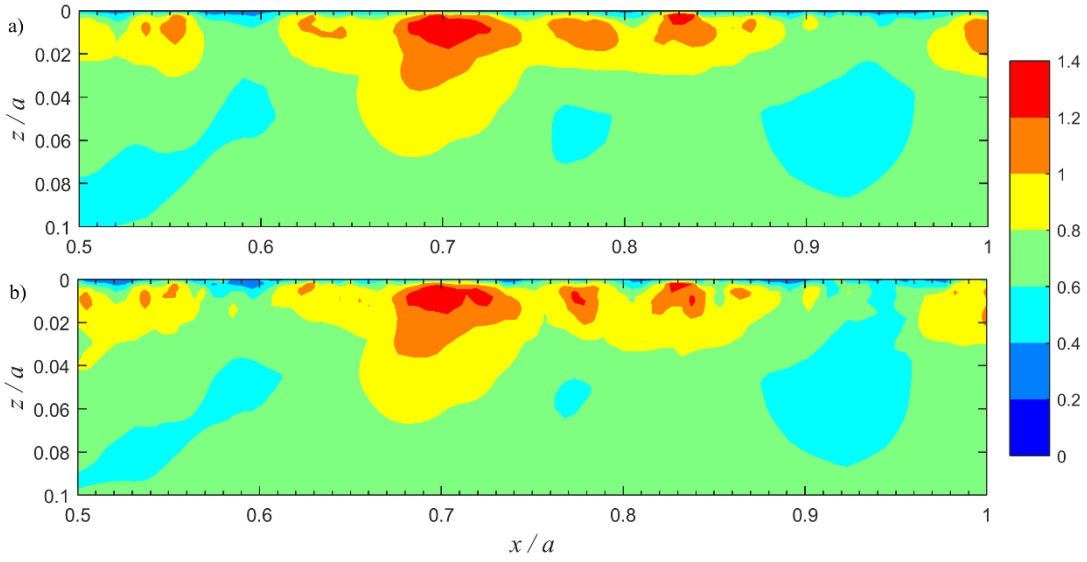


Figure 5-21 slow disk: Findley fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.



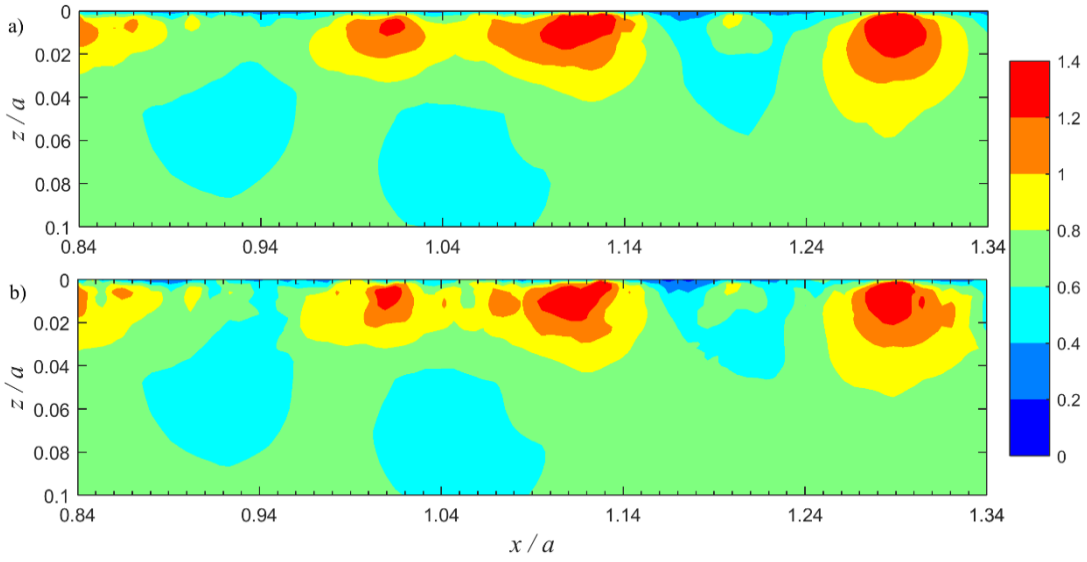


Figure 5-22 slow disk: Findley fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

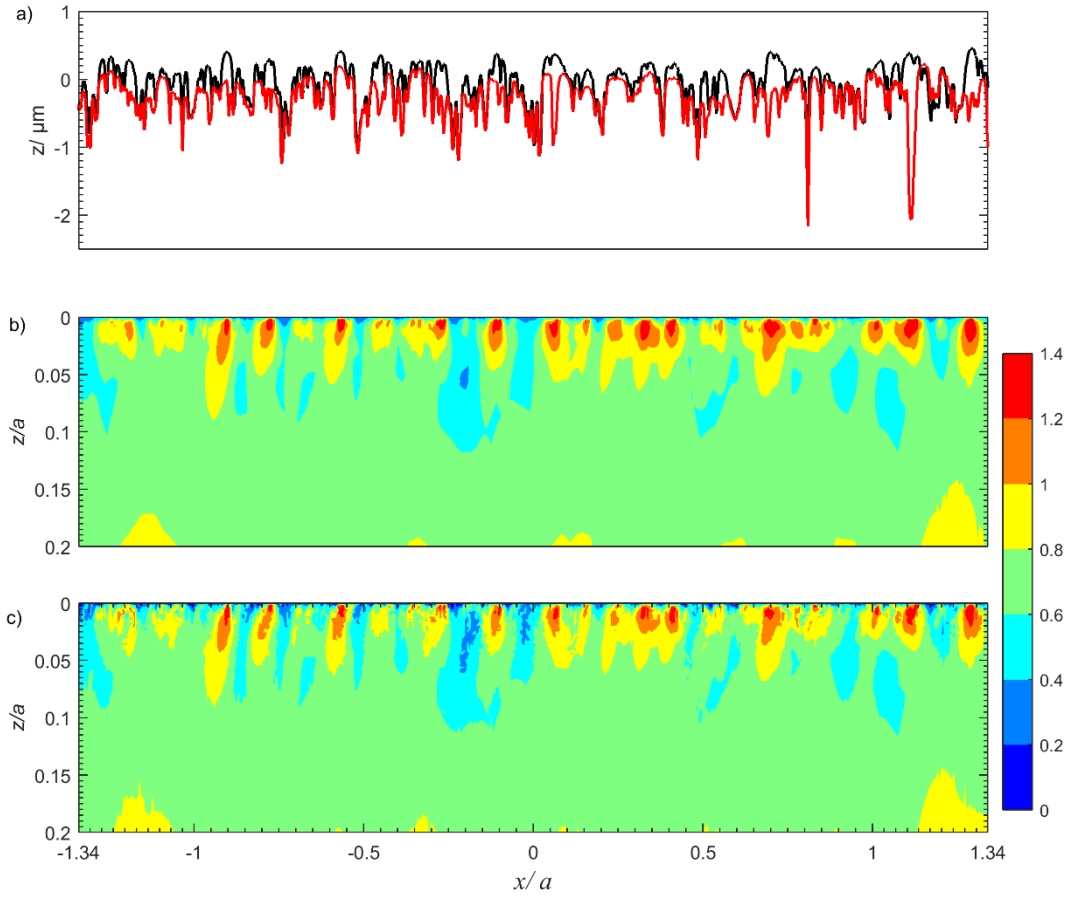


Figure 5-23 slow disk- contours of fatigue parameter (FP) at 10^7 cycles using the Matake Criterion, target load is 1.4GP: a) Surface profile; EHL – black, LS18 (experimental) - red; b) Damage without residual stress, c) Damage with residual stress.

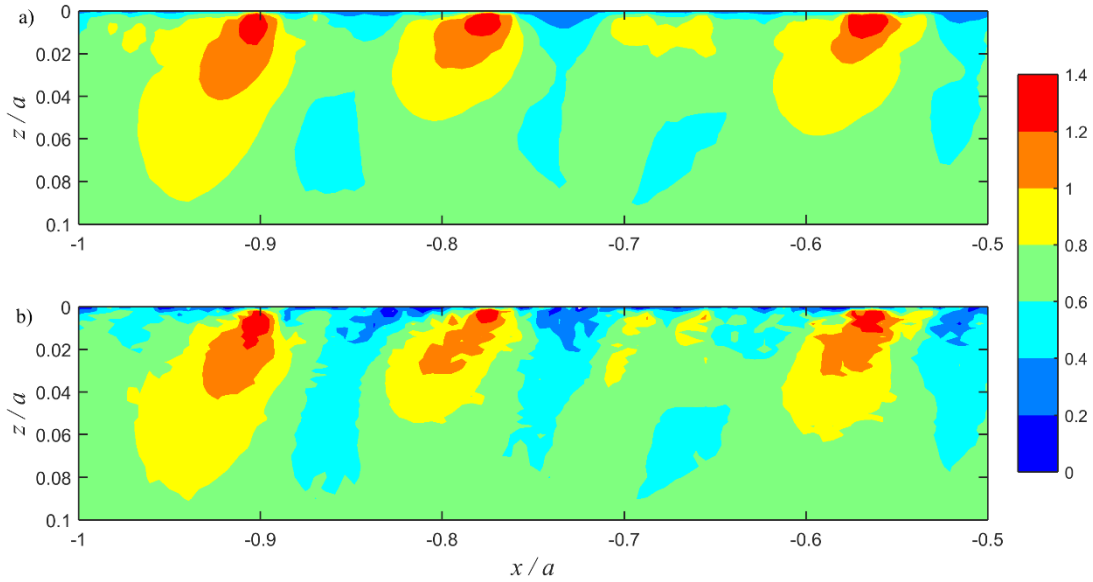
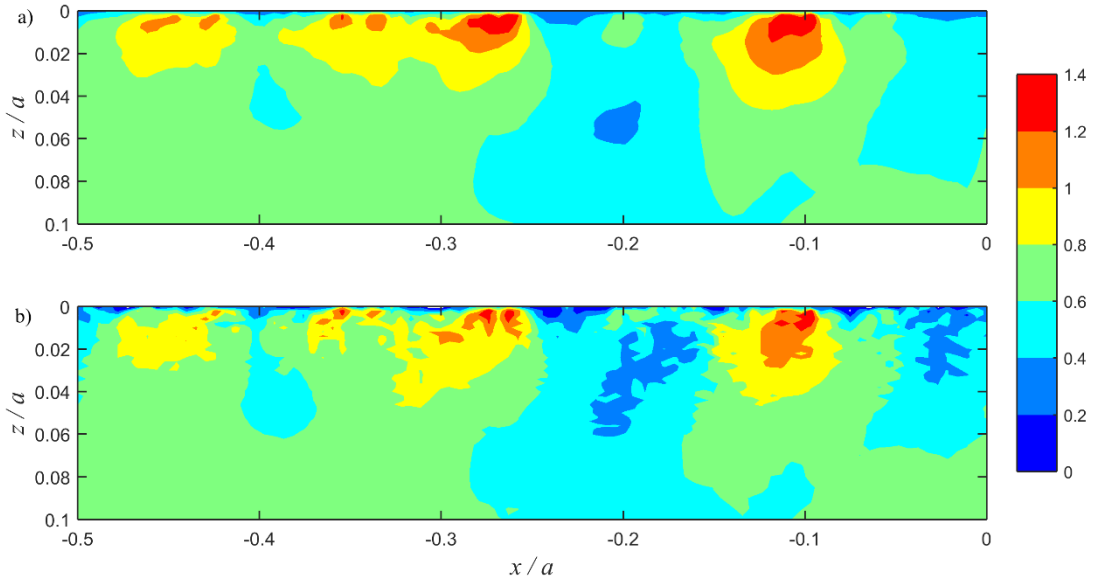


Figure 5-24 slow disk: Matake fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.



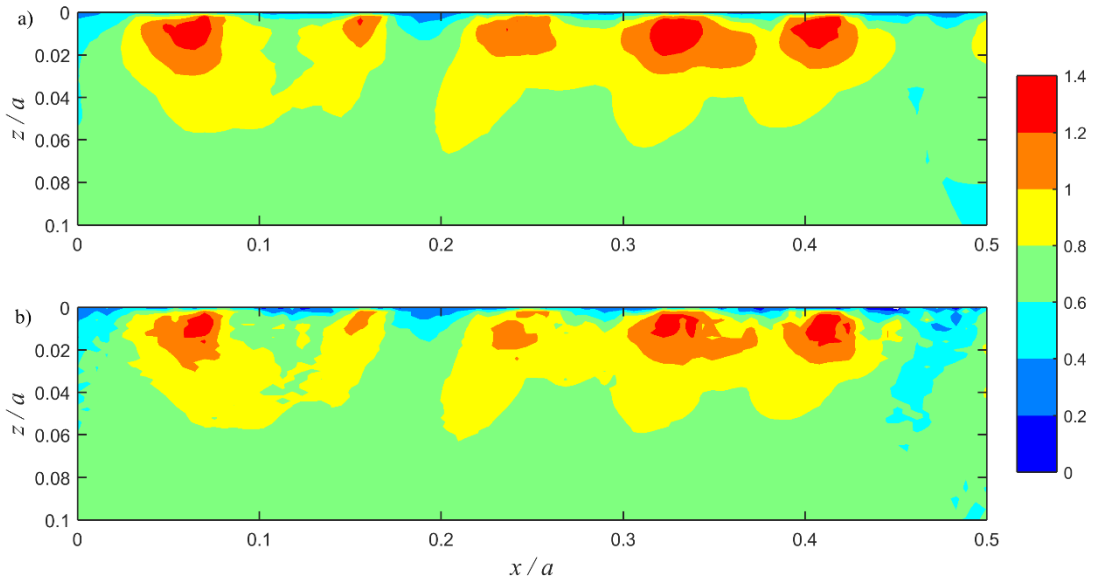


Figure 5-25 slow disk: Matake fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

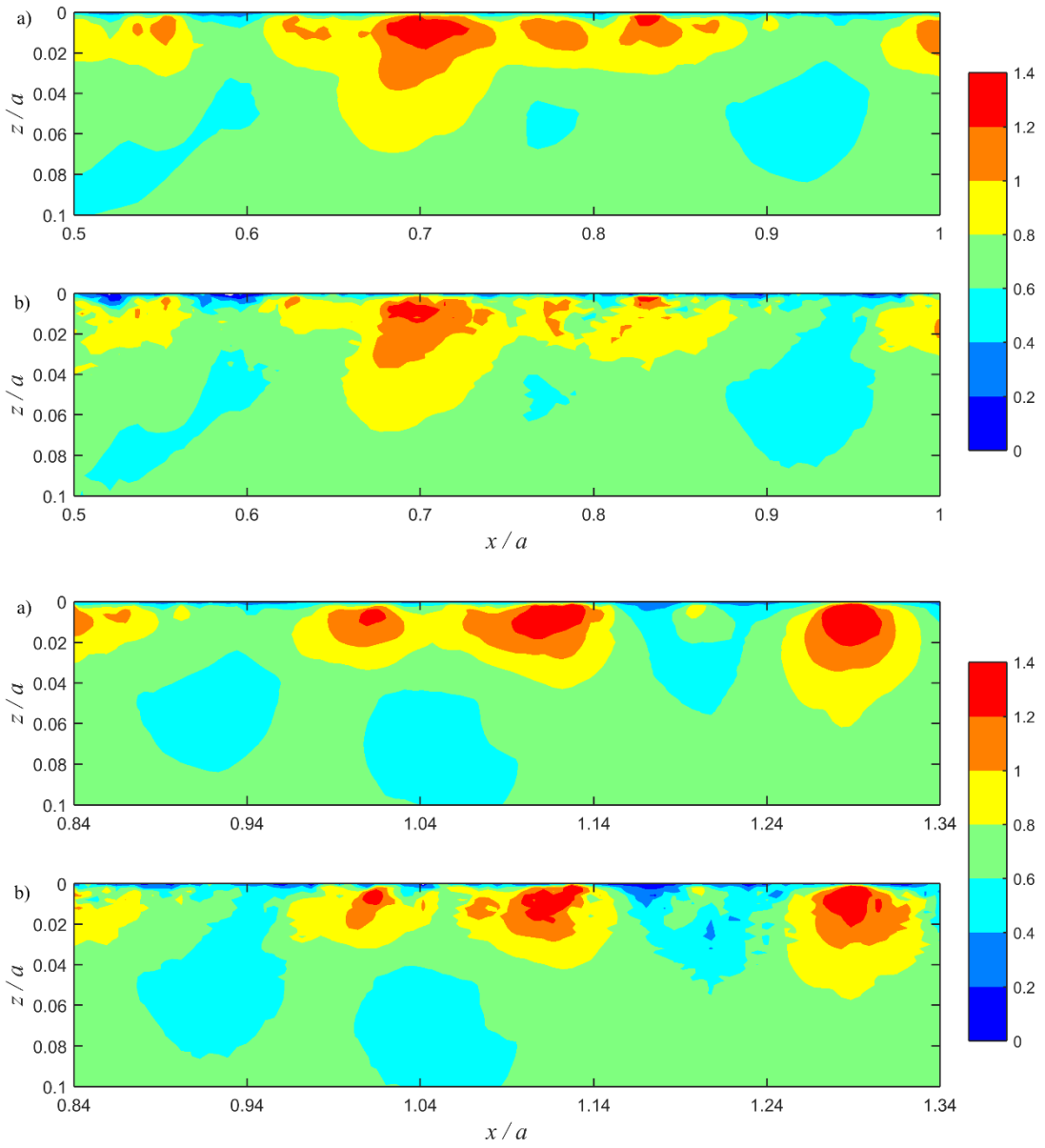


Figure 5-26 slow disk: Matake fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

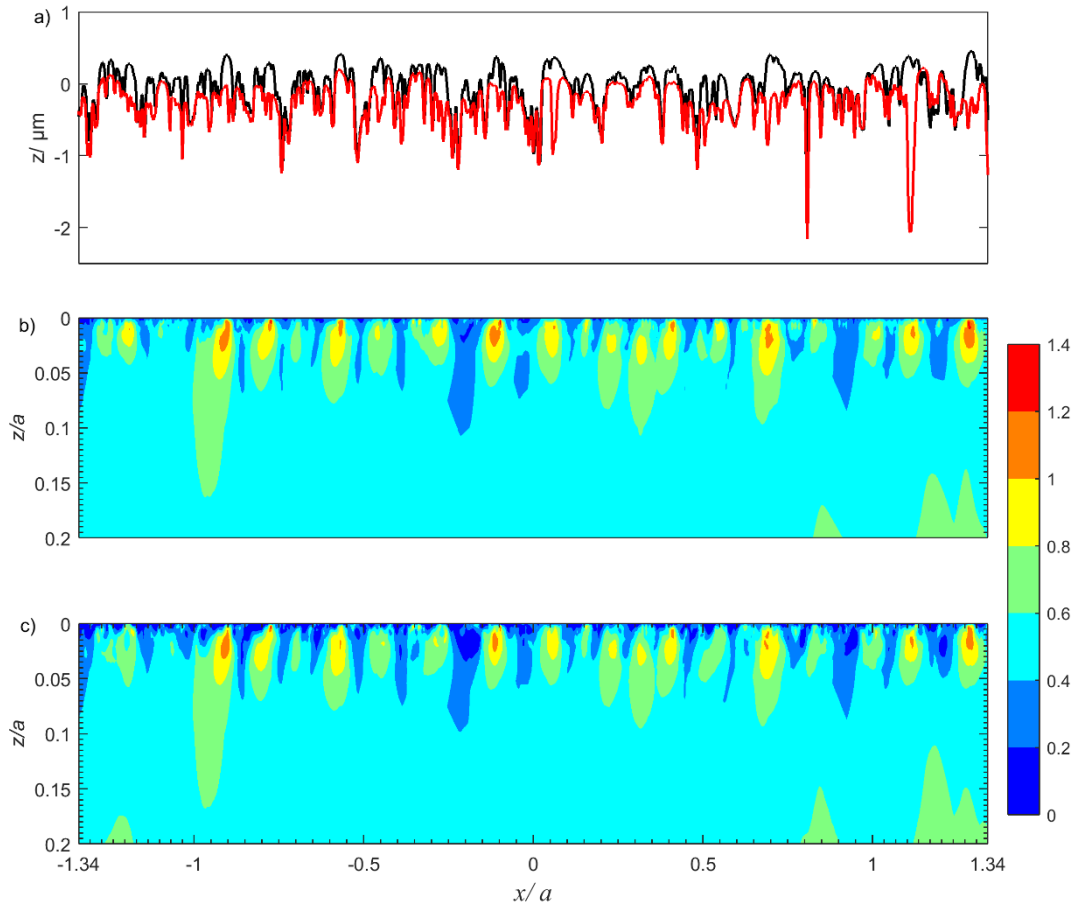


Figure 5-27 slow disk- contours of fatigue parameter (FP) at 10^7 cycles using the Dang Van Criterion, target load is 1.4GP: a) Surface profile; EHL – black, LS18 (experimental) - red; b) Damage without residual stress, c) Damage with residual stress.

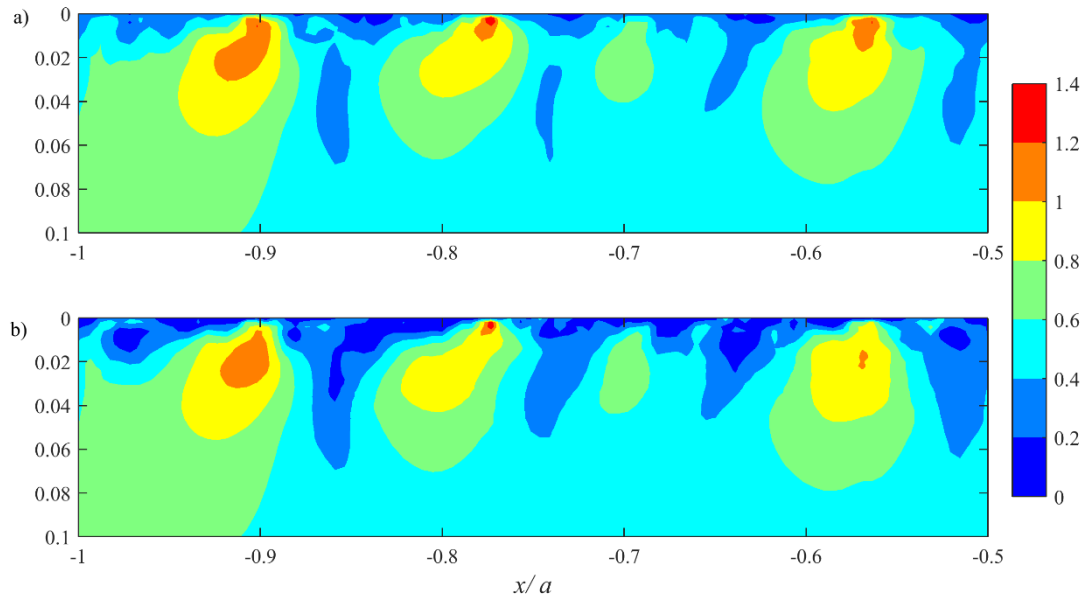


Figure 5-28 slow disk: Dang Van fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

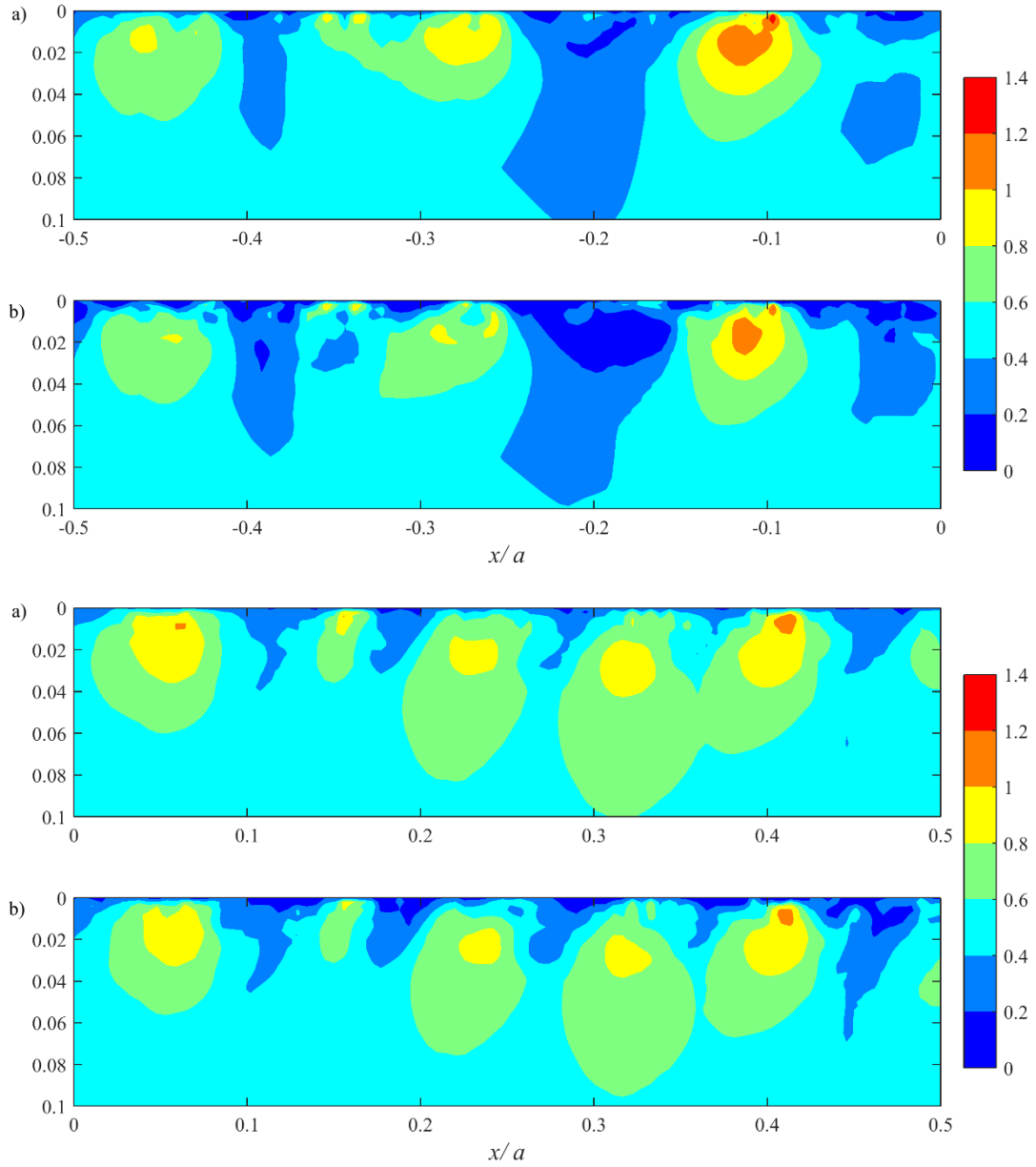


Figure 5-29 slow disk: Dang Van fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

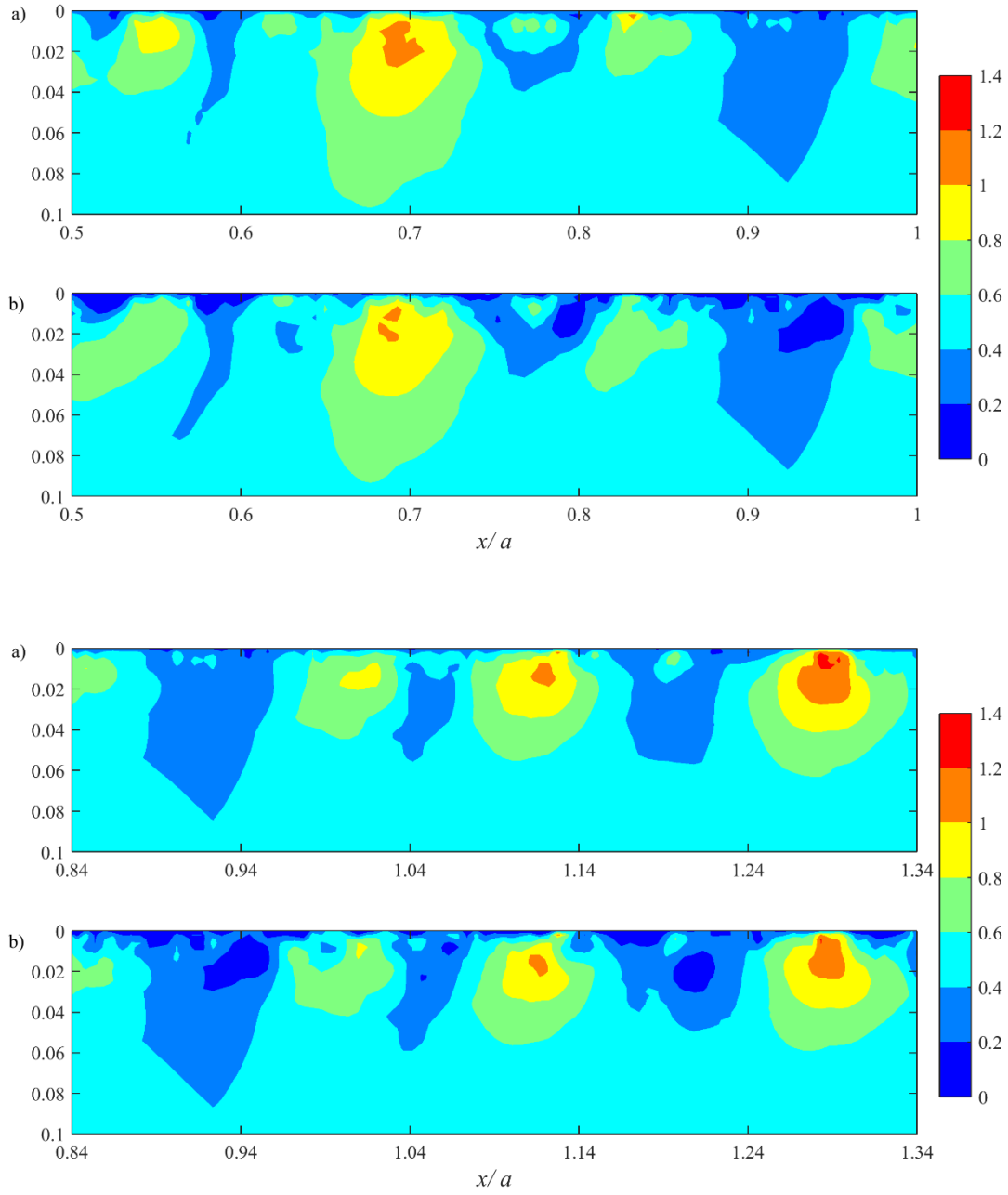


Figure 5-30 slow disk: Dang Van fatigue parameters for 10^7 loading cycles at the close proximity to the contact surface; part a) Damage without residual stress, part b) Damage with residual stress.

A summary of the application of the critical plane approach criteria to this analysis is presented in Table 4, which records ten asperities which have highest fatigue parameter (*FP*) values. The ten asperities are labelled with numbers of 1 to 10 and identified by the distinguishing arrows in Figure 5.18 *a*. It can be seen from the data in Table 4 that the Findley criterion and the Matake criterion reported almost the same *FP* values. Using the Dang Van criterion, the calculated results are slightly lower for this case study. What is interesting about the data in this table is that all three critical plane approach criteria agree closely on which asperities are those most likely to experience fatigue and, which are not likely to experience fatigue.

Table 4 Ten identified asperities with highest fatigue parameter values by three fatigue criteria.

Index	*Findley*		*Matake*		*Dang Van*	
	FP	Asperities No.	*FP*	Asperities No.	*FP*	Asperities No.
1	1.9305	3	1.953	3	1.4742	3
2	1.8486	10	1.8683	10	1.4319	10
3	1.7222	9	1.7421	9	1.3369	9
4	1.6721	4	1.6921	4	1.3174	4
5	1.6409	1	1.6605	1	1.2183	1
6	1.5807	5	1.5996	5	1.2018	5
7	1.5207	6	1.5311	6	1.1536	6
8	1.3957	8	1.4044	8	1.0966	8
9	1.3256	7	1.3323	7	1.0414	7
10	1.1715	2	1.1831	2	0.96081	2

Figures 5.31 to 5.34 show the *FP* values plotted against *x/a* at a different depth levels *z/a* = 0.0, 0.008, 0.01, 0.016, 0.024 and 0.5 using these criteria. When the fatigue parameter exceeds unity at a position, fatigue failure is expected to occur. Closer inspection of these plots show that the *FP* values have their greatest variation at and near the surface, such as depths of 0.008*a* (3.75 µm) and 0.01*a* (4.69 µm), due to the effect of roughness where the highest damage values are found at this depth level. At increasing distance from the roughness surface, for instance at depth of 0.5*a* (234.5 µm), *FP* curves become flat and lower where the stress is seldom influenced by the roughness.

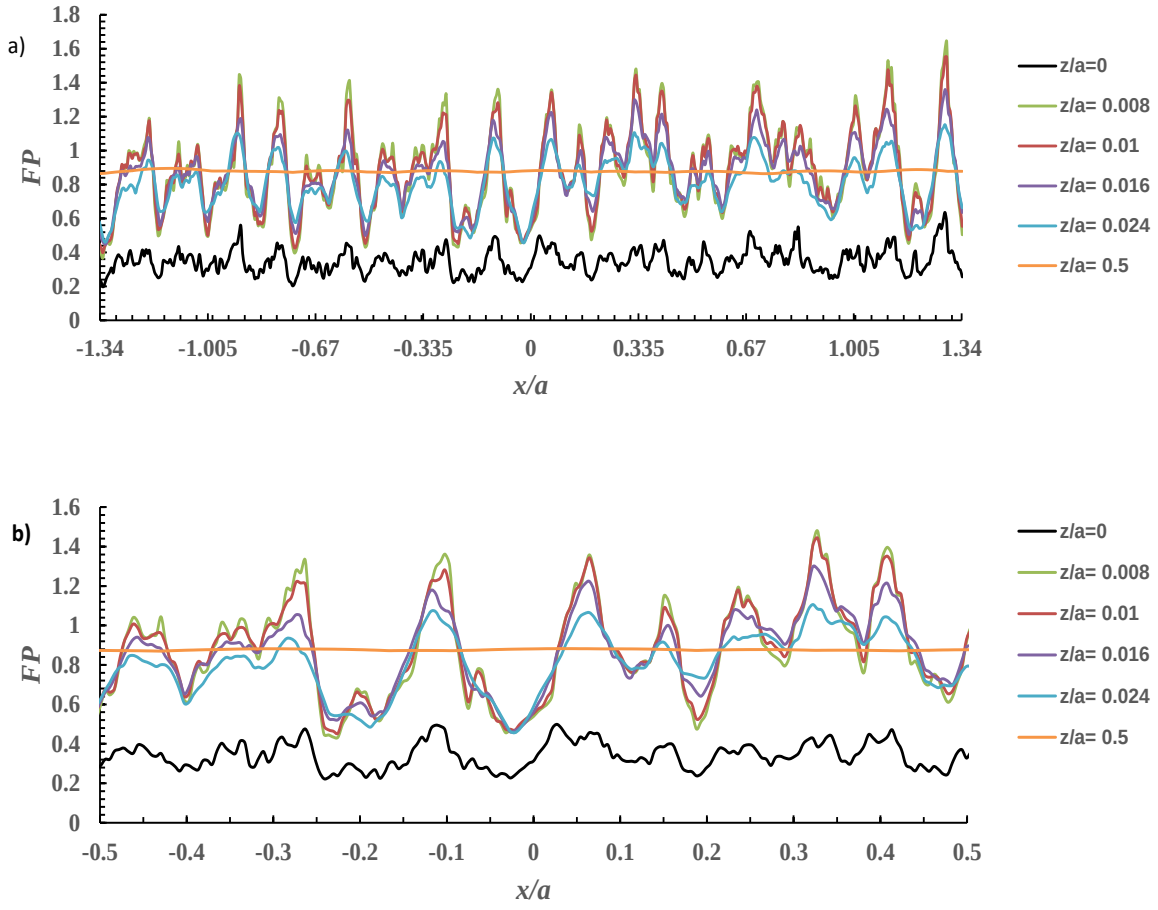


Figure 5-31 Distributions of fatigue parameter (FP) with residual stress at 10^7 cycles at different depth levels by using Findley criterion; a) full scale plot and b) large scale plot from *x/a* = - 0.5 to x/a = +0.5.

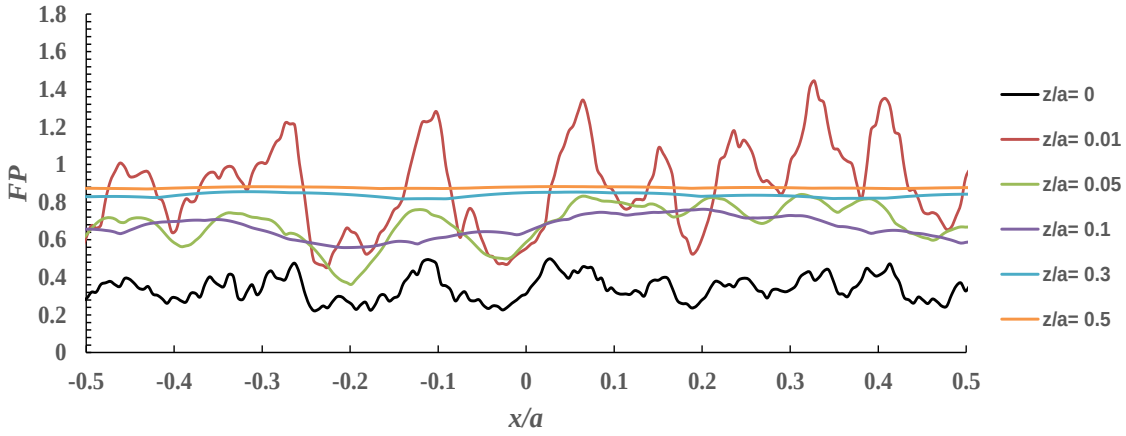


Figure 5-32 Distributions of fatigue parameter (*FP*) with residual stress at 10^7 cycles at different depth levels by using Findley criterion at larger scale plot from x/a = - 0.5 to x/a = +0.5.

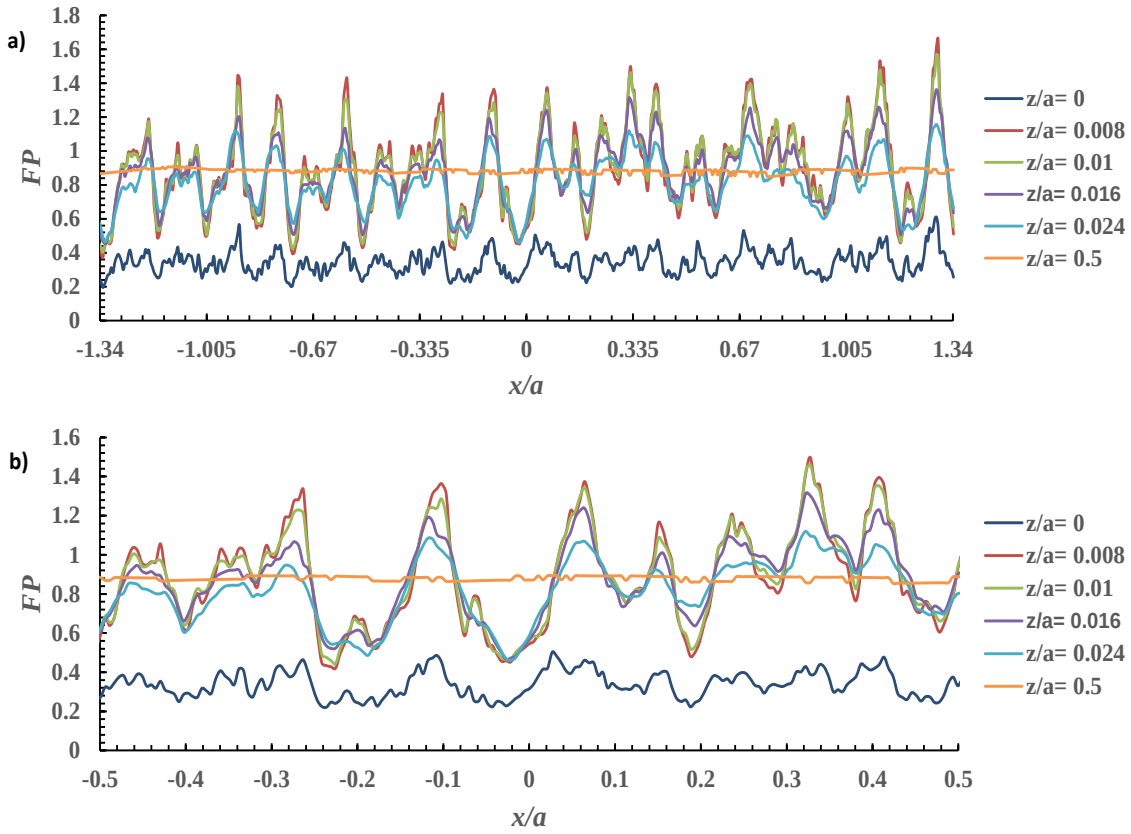


Figure 5-33 Distributions of fatigue parameter (*FP*) with residual stress at 10^7 cycles at different depth levels by using Matake criterion; a) full scale plot and b) large scale plot from x/a = - 0.5 to x/a = +0.5.

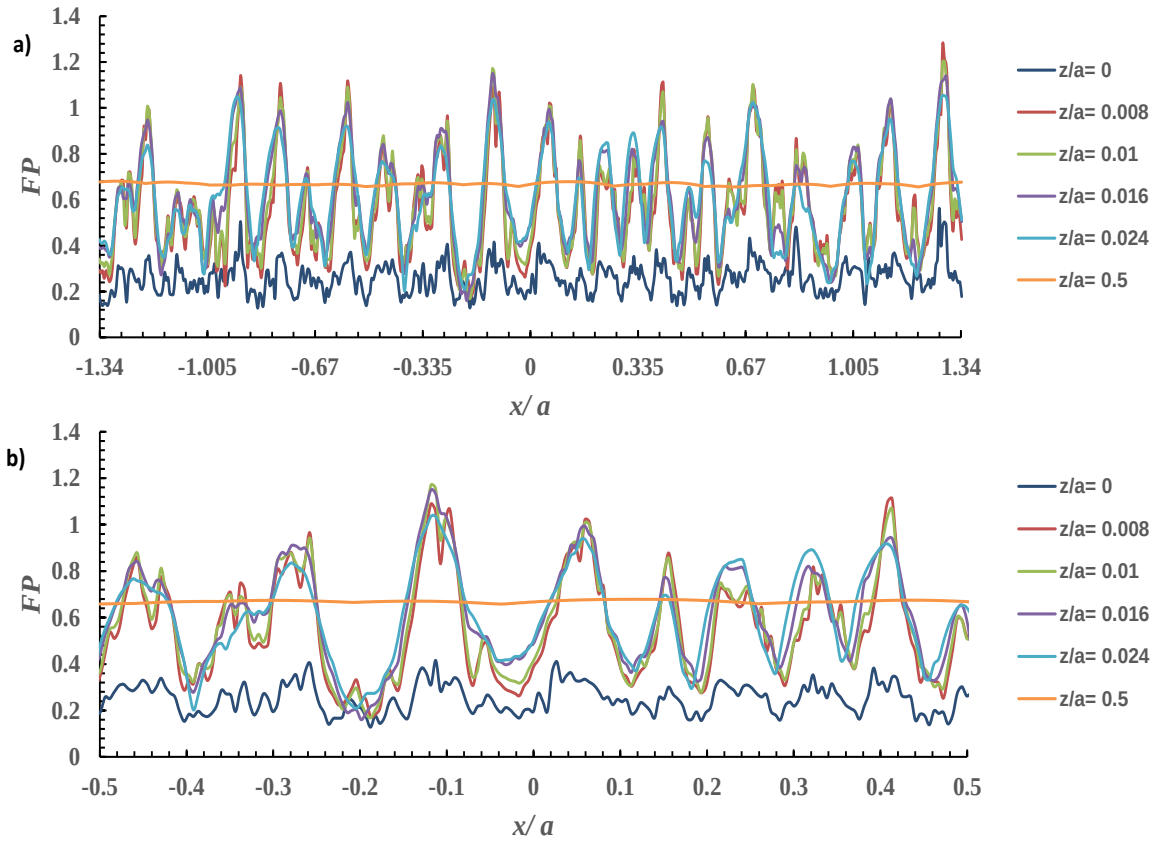


Figure 5-34 Distributions of fatigue parameter (FP) with residual stress at 10^7 cycles at different depth levels by using Dang Van criterion; a) full scale plot and b) large scale plot from x/a = - 0.5 to x/a = +0.5.

5.7 Conclusion

Three methods for calculating the fatigue parameter have been used, namely the Findley, the Matake, and the Dang Van criterion for EHL line contact of two rough surfaces. The results have been verified by comparison to numerical and experimentally observed micropitting data, where some micropits seen in LS18 (fatigue profile) were found to occur at positions that had high calculated fatigue parameter values. The calculated stresses were subsequently used for calculating fatigue damage according to the proposed multiaxial fatigue damage parameter. The fatigue threats for the three criteria are highest near the material surface due to the influence of roughness and descend rapidly deeper in the subsurface. It can be concluded that the residual stress due to running-in of asperities does cause changes in the damage prediction and calculated fatigue life of the surfaces. These changes can be significant at some positions in the material but are not necessarily detrimental.

Chapter 6 Fatigue analysis – applying variable amplitude multiaxial fatigue criteria

6.1 Introduction

A rough EHL contact simulation procedure to estimate the cumulative fatigue damage is established in this section. In the passage of the surface material through the EHL contact, different asperity interaction events will occur due to the relative rolling motion of the surfaces and as a result there will be a number of interactions so that the material is subjected to variable amplitude cyclic loading. These cycles are different and are potentially in different axial orientations. The model needs to relate to this loading pattern and does so by using a fatigue life model to determine the amount of damage caused by each cycle.

The critical plane approach is used in this procedure, which is built upon a strain-life model that is used to establish the damage associated with each effective loading cycle. The model used in this work is that of Fatemi and Socie, (1988). In such a model, the stresses and strains are calculated during cyclic loading for all of the planes in the material, and the maximum cyclic shear strain range with the maximum normal strains drive the fatigue crack initiation of the surface materials. It has generally been recognized that critical plane models often achieve better predictions for fatigue damage assessment. This chapter presents a variable amplitude multiaxial fatigue model calculation for the EHL line contact with rough surfaces based on a critical plane approach.

6.2 Fatigue models based on critical plane approach

Several models based on the critical plane approach have been developed for estimating fatigue life in the material, where the stresses and strains normal to the crack plane might have a major effect on the behaviour, accelerating the growth if they tend to open the crack. This situation has led to a number of proposals for critical plan approaches.

The model adopted in this work is Fatemi and Socie's fatigue model, (1988) which is based on the assumption that the damage will occur according to its parameters' effect that is shown in equation (6.1), which are a combination of a normal direct stress and shear strain.

$$\frac{\Delta\gamma_{\max}}{2}\left(1+k\frac{\sigma_n^{\max}}{\sigma_\gamma}\right)=\frac{\tau'_f}{G}\left(2N_f\right)^{b_o}+\gamma'_f\left(2N_f\right)^{c_o} \tag{6.1}$$

Where $\frac{\Delta\gamma_{max}}{2}$ refers to the amplitude of shear strain on the critical plane, and σ_n^{max} refers to the maximum tensile stress which is normal to the critical plane. Parameter $\boldsymbol{k}$ is a material constant with the range, $0.6 < k < 1.0$. The model is applied over a mesh of points within the block of material and at each of these points; it is applied to planes of all orientations in turn. In the simulation process the planes are rotated in one-degree steps and the orientation that gives rise to the largest accumulated damage value is the critical plane, and the value of *D* for the point is this maximum value. The value of damage obtained for each material point is based on all of the equivalent loading cycles identified during one pass of the material point through the EHL contact.

Chu, (1995) highlights the need to find the link between shear strain and normal stress components by developing a model that combines both shear and normal components. This model allows the number of loading cycles to failure N_f, to be calculated for the effective loading cycle.

This model was developed for the material with mixed cracking modes; cracks are initiated and grown on the maximum normal planes for tension loading, at high strain but on the maximum shear planes for torsion loading. Chu's model proposes that the fatigue is predicted by the strain life relation, which is described by equation (6.2);

$$\sigma_{\max}\varepsilon_a + J\Delta\tau\Delta\gamma = \frac{\left(\sigma'_f\right)^2}{E}\left(2N_f\right)^{2b} + \sigma'_f\varepsilon'_f\left(2N_f\right)^{b+c} \tag{6.2}$$

Where the maximum stress and the strain amplitude are defined by $\sigma_{\max}$ and ε_a , respectively, which are normal to the crack plane; the shear strain and the shear stress range are defined by $\Delta\gamma$ and $\Delta\tau$, respectively, on the same plane. *J* is a material constant, which can be obtained from tension/torsion tests. He determined the critical plane and the largest damage parameter from the transformation of strains and stresses onto planes spaced at equal increments. For fatigue life prediction of rolling contact fatigue creation, Johnson, (1989) strongly claimed that the cracks take place in shear under hydrostatic compression. Therefore, the Fatemi and Socie, (1988) shear failure model may be considered the more appropriate model for assessment of variable amplitude multiaxial fatigue damage in this research.

6.3 Fatigue damage evaluation procedures.

The evaluation of any particular fatigue model involves using the stress history for the block as it passes through the EHL contact zone. To evaluate the total-life the following steps are taken for each point in the fatigue calculation area:

1. Read the stress history for all the evaluation points in the material block considered for EHL analysis.

2. For each evaluation point in turn use the stress component history to calculate the strain history by using Hook's law as follows :

$$\varepsilon_x = \frac{1}{E}\left\{\left(1-\nu^2\right)\sigma_x - \nu\left(1+\nu\right)\sigma_z\right\} \tag{6.3}$$

$$\varepsilon_z = \frac{1}{E}\left\{\left(1-\nu^2\right)\sigma_z - \nu\left(1+\nu\right)\sigma_z\right\} \tag{6.4}$$

$$\gamma_{xz} = \frac{\tau_{xz}}{G} = \frac{2\left(1+\nu\right)}{E}\tau_{xz} \tag{6.5}$$

3. Consider each potential critical plane orientation in turn which is defined by θ, the angle between the normal of the plane and the *x*-axis. Obtain normal and shear stress history, and in the same way calculate normal and shear strain on the critical plane respectively according to Timoshenko and Goodier (1973) as:

$$\sigma = \sigma_x \cos^2\theta + \sigma_z \sin^2\theta + 2\tau_{xz}\sin\theta\cos\theta \tag{6.6}$$

$$\tau = \tau_{xz}\left(\cos^2\theta - \sin^2\theta\right) + \left(\sigma_z - \sigma_x\right)\sin\theta\cos\theta \tag{6.7}$$

$$\varepsilon = \varepsilon_x \cos^2\theta + \varepsilon_z \sin^2\theta + \gamma_{xz}\sin\theta\cos\theta \tag{6.8}$$

$$\frac{\gamma}{2} = \frac{\gamma_{xz}}{2}\left(\cos^2\theta - \sin^2\theta\right) + \left(\varepsilon_z - \varepsilon_x\right)\sin\theta\cos\theta \tag{6.9}$$

4. Determine the effective loading cycles based on the shear strain history. Finding the corresponding stress level for each effective loading cycle. These cycles are counted by using the rainflow method developed by Amzallag et al. (1994).
5. For each cycle identified in step 4 in turn, determine the shear strain amplitude γ_a , normal strain amplitude ε_a ,and the maximum normal stress σ_n .
6. Calculate the fatigue life, N_f as per the Fatemi and Socie shear model for the current cycle of strain by solving equation 6.10.

$$\frac{\Delta\gamma_{\max}}{2}\left(1+k\frac{\sigma_n^{\max}}{\sigma_\gamma}\right)=\frac{\tau'_f}{G}\left(2N_f\right)^{b_o}+\gamma'_f\left(2N_f\right)^{c_o} \tag{6.10}$$

Where $\frac{\Delta\gamma_{max}}{2}$ refers to the amplitude of shear strain on the plane considered, and σ_n^{max} refers to the maximum tensile stress which is normal to the plane. Parameter k is a material constant with the range, $0.6 < k < 1.0$. The fatigue damage associated with the current cycle will be determined on the candidate plane as:

$$damage_{cycle}=\frac{1}{N_f} \tag{6.11}$$

The total damage sustained from the cyclic loading is obtained using the Palmgren-Miner (1924) rule

$$D=\Sigma_{\substack{all\\effective\\loading\\cycles}}\frac{1}{N_f} \tag{6.12}$$

This is the damage value for the current candidate plane and a value of D equal to unity then corresponds to fatigue failure. The process is repeated from step 3 for each candidate critical plane orientation and the results used to identify the plane that has the highest value of D, D_{max}. This is the critical plane for the evaluation point being investigated, and $D_{\max}$ is the value of the accumulated damage at that point.

7. Repeat from step 1 for each evaluation point in the material in the fatigue evaluation area.
8. Finally determine the fatigue life of the component material and plot the value of D_{max} for each point in the fatigue evaluation area as:

$$N_f=\frac{1}{D_{max}} \tag{6.13}$$

6.4 Weibull cumulative density function (*CDF*)

The Weibull distribution is a theoretical model that is successfully used in statistical analysis to estimate the failure probability. The Weibull distribution model is a tool to develop the probabilistic analysis because of its ability to provide reasonably accurate failure analysis and failure calculations with extremely small sample data. The three-parameter Weibull distribution can be used to describe the probability density function *F* (D) of damage at any specified depth. These distributions are useful in comparing results for different conditions when presented in the cumulative form *F* (*D*). This function is skewed towards low damage, and the Weibull distribution is thus an appropriate stochastic model to use for this purpose. The cumulative density function for this distribution has the form:

$$F(D) = 1 - \exp\left[-\left(\frac{D - \mu}{\eta} \right)^{\beta} \right], \quad 0 \le \mu \le D, \quad \eta, \beta > 0 \tag{6.14}$$

Where η is the scale parameter, β is the shape parameter and μ is the location parameter. If $\mu = 0$, the 3-parameter Weibull distribution becomes the 2-parameter Weibull distribution.

6.5 Cracking mechanisms in multiaxial fatigue

It has been suggested by Brown and Miller (1973) that there are two different mechanisms for crack propagation at surfaces of materials subjected to multiaxial fatigue loading on the basis of the orientations of the crack planes of maximum shear strain amplitude with the free surface. This process is described in more detail by Suresh, (1998) based on the figure reproduced as Figure 6.1.

Figures (6.1 *a* and *e*) show a cubic element of material subjected to multiaxial strain cycles as well as the free surface plane. Figures (6.1 *b* and *c*) present the planes of maximum shear strain amplitude. There are two stages of crack growth in the materials, and they are called

stage I and stage I I cracks. Stage I crack is initiated by the single shear process at the free surface and progresses during subsequent cycling. In these plots, the shear stress acts parallel to the free surface, and there is no shear stress acting on the normal direction to that surface. The unit normal vectors to the planes of maximum shear strain amplitude are located on the material surface plane. Under these conditions, the cracks propagate more in a direction parallel to the surface than normal to the surface, thereby increasing the aspect ratio of the crack. This mechanism of crack propagation is called ‘case A' by Brown and Miller (1973).

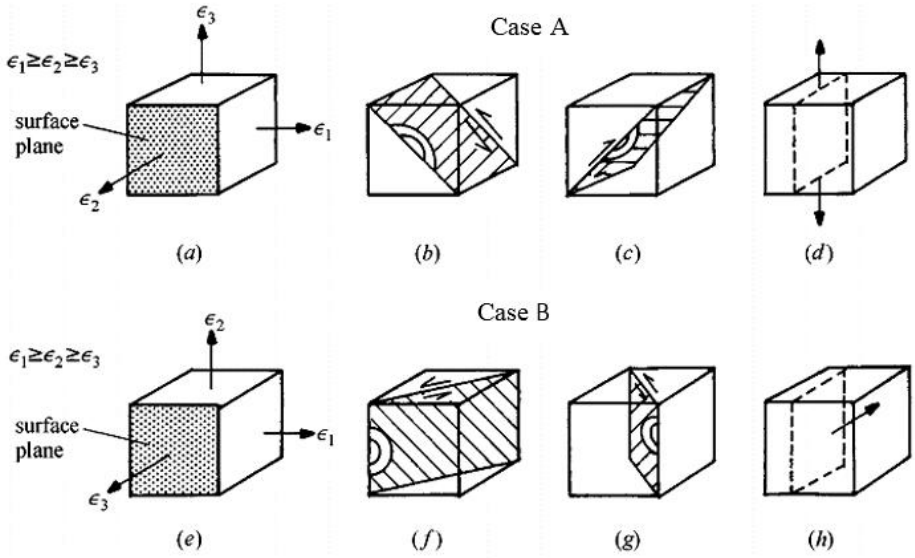


Figure 6-1 Schematic illustrations of case A and case B fatigue cracking in multiaxial fatigue: (a) and (e) show multiaxial strains, (b), (c), (f) and (g) denote planes of maximum shear strain amplitude and the planes and directions of stage I crack growth for case A and case B. (d) and (h) show the planes and directions of stage II crack growth for case A and case B, respectively(Suresh,1998).

For most materials, stage II crack growth occurs because of simultaneous or alternating slip including more than one processes slip system. At this time, the direction of crack advance and the plane on which it occurs are as shown in Figure (6.1 d). In the Figures (6.1 *f* and *g*) , the stage I cracks initiate at the surface and propagate at 45° angles into the material, and this mode of cracking has been termed ‘case B’. Fatemi and Socie (1988) determined that

the peak normal stress to the plane of maximum shear strain amplitude effects the propagation of stage I crack under a variety of multiaxial loading conditions.

6.6 Fatigue analysis results and discussion

In this section, the fatigue damage calculations with and without residual stress were carried out for the EHL line contact of rough surfaces of test disks from the experimental work. The effect of the FEA residual stress for the same profiles and used for the fatigue damage calculations were investigated. The procedure to estimate the fatigue damage of a material subjected to variable amplitude multiaxial loading is used which is based on the Fatemi and Socie (FS) shear strain based critical plane model. The parameters used to specify the EHL problem are the same as given in Table 3 in Chapter 5 unless otherwise specified. For damage modelling the material SAE4340 steel (BS970:En24) has been used where material properties can be found in Zahavi and Torbilo (1996) which are shown in Table 1. Dowling (1998) specifies the shear fatigue ductility coefficient and shear fatigue strength according to the following expression:

$$\tau'_f \approx \frac{\sigma'_f}{\sqrt{3}},\ \gamma'_f \approx \sqrt{3}\varepsilon'_f \tag{6.15}$$

Table 1 Material parameters of SAE4340 steel.

b	-0.091	ε'_f	0.48
c	-0.60	σ'_f	2.0 GPa
E_1,E_2	200 GPa	σ'_o	827 MPa
K	1.0	$\nu_1,\ \nu_2$	0.3

A sample roughness profile approximately 1*mm* long was taken from each disk and a mixed EHL transient analysis was carried out to determine the surface loading during the passage of the slow and fast surface profiles. This was used to determine the stress history at each

calculation point in the material. The stress/strain history at each point was broken down into a number of cycles as discussed in section 6.3. The damage for the trial was calculated according to equation 6.1and the accumulated damage for each calculation position a sequence of evaluation planes was considered where one degree increments of θ were used for each candidate plane orientation in the range $0 < \theta < 180^0$. For each orientation of the evaluation plane the damage for each cycle was calculated using equation 6.1 and the accumulated damage for the loading history was obtained using equation 6.13. This was repeated for each evaluation plane and the critical plane for the evaluation point was the plane, which lead to the maximum value of *D*. In this way, the value of damage and the critical plane orientation was established for all evaluation points in the material volume considered.

6.6.1 Damage distributions with and without residual stress

The fatigue analysis results for the slow disk are presented in Figures 6.2 to 6.5.

Figure 6.2 shows the overall results for the material volume considered for the slow surface loading contact zone. It presents contours of calculated damage for a single traverse of the contact zone. The roughness profile is shown together with damage contour plot for the cases where residual stress is and is not included in the calculation. The plots show areas of relatively high accumulated damage exceeding 10^{-5} in concentrated areas close to the surface within the approximate range $0.0 < z/a < 0.07$, indicating fatigue in 10^5 cycles. It can be seen that these concentrated areas are located beneath asperity features. A detailed comparison of the results obtained with and without including residual stresses is presented in figures 6.3 to 6.5. What is interesting about the data in these contour plots is that there is a significant increase in the assessed damage values compared to the Findley, the Matake, and the Dang Van models. For example, those prominent asperities at a larger scale located at x/a = -0.58, x/a =-0.8, x/a =-0.9, x/a =-0.1, x/a = -0.3, x/a = 1.28, where indicated fatigue

in fewer than 10^5 cycles is apparent at and below the surface whereas *FP* values of unity in the Findley, the Matake, and the Dang Van results correspond to fatigue in 10^7 cycles.

The larger scale plots in figures 6.3 to 6.5 show that including residual stress in the calculation of damage can change the peak values considerably. This is not detrimental in all cases but is clearly so for the asperity at x/a =1.34 in figure 6.5, for example. The damage results suggest that the mechanism of deterioration seems to be micropitting associated with the plastic deformation of the prominent asperities. One of these areas have been associated with areas of high residual tensile stress (Bryant et al., (2012), Alshahrany, (2015)). The current results have been verified by comparison between the numerical calculations and experimentally observed micropitting data, where micropitting, which has been observed in the tests, was found to occur at positions that had high-calculated fatigue damage. Therefore for practical purposes, these new valley features can be deemed very close to crack initiation. This can be seen in Figure (6.2 *a*) where the new valley features appeared to occur at the locations of previously observed asperities for the fatigue profile shown in red colour. An example of this can be seen at x/a =0.05, x/a=0.7, x/a=1.15, x/a= 1.28, x/a= -0.45 and x/a= -1.17. On observation of micropitted data presented in this work it is clear that fatigue damage occurs on certain asperity tips while other tips remain free of pits. The finding that emerges from this case study is that accumulate damage are localised near the surface of specific asperity features at shallow depths.

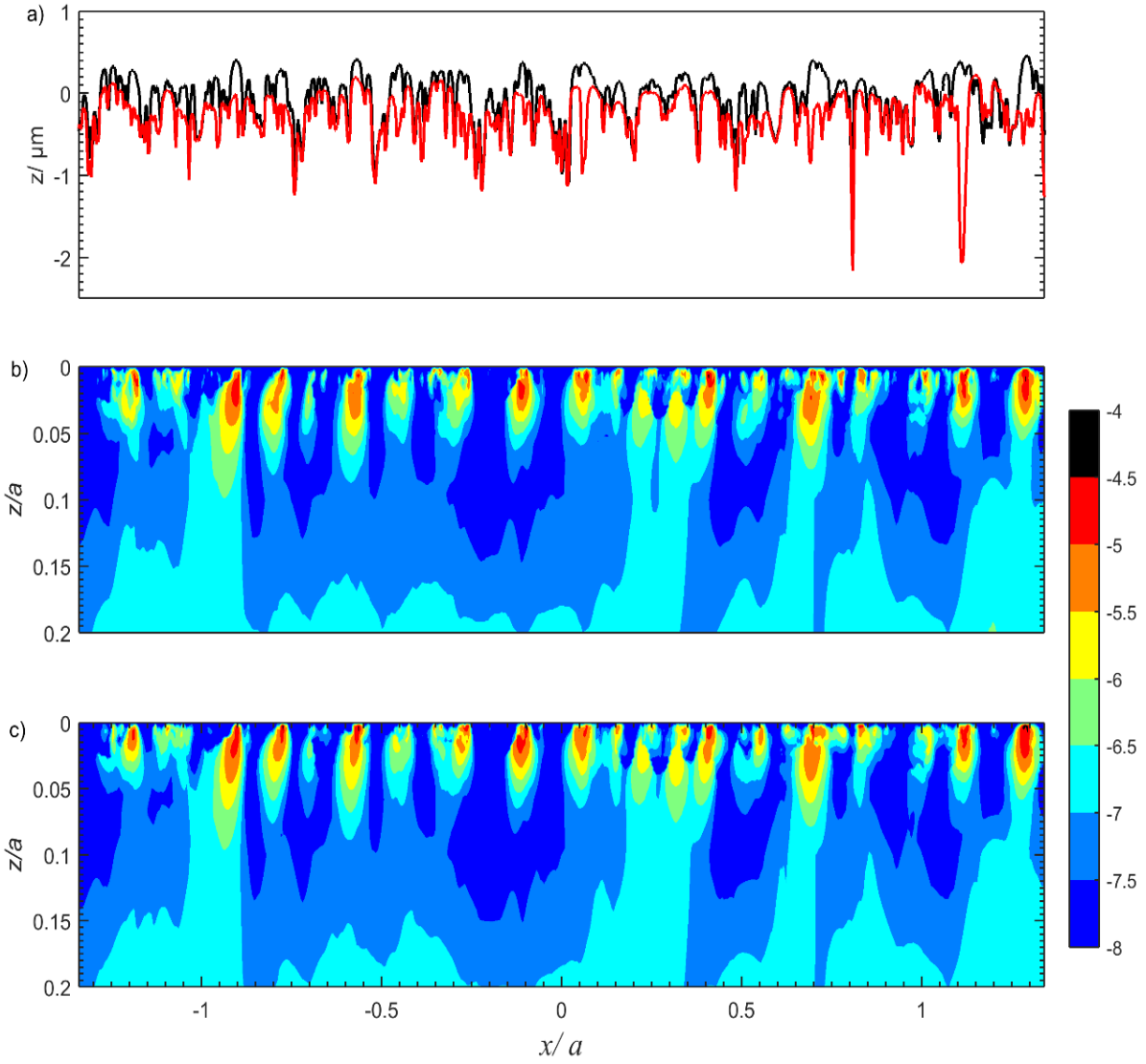


Figure 6-2 slow disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion, target load is 1.4GP: a) roughness profile; EHL- Black and Fatigue profile- Red; b) Damage without residual stress, c) Damage with residual stress, accumulated damage, 10^{-n}, indicating fatigue in 10^{n} cycles.

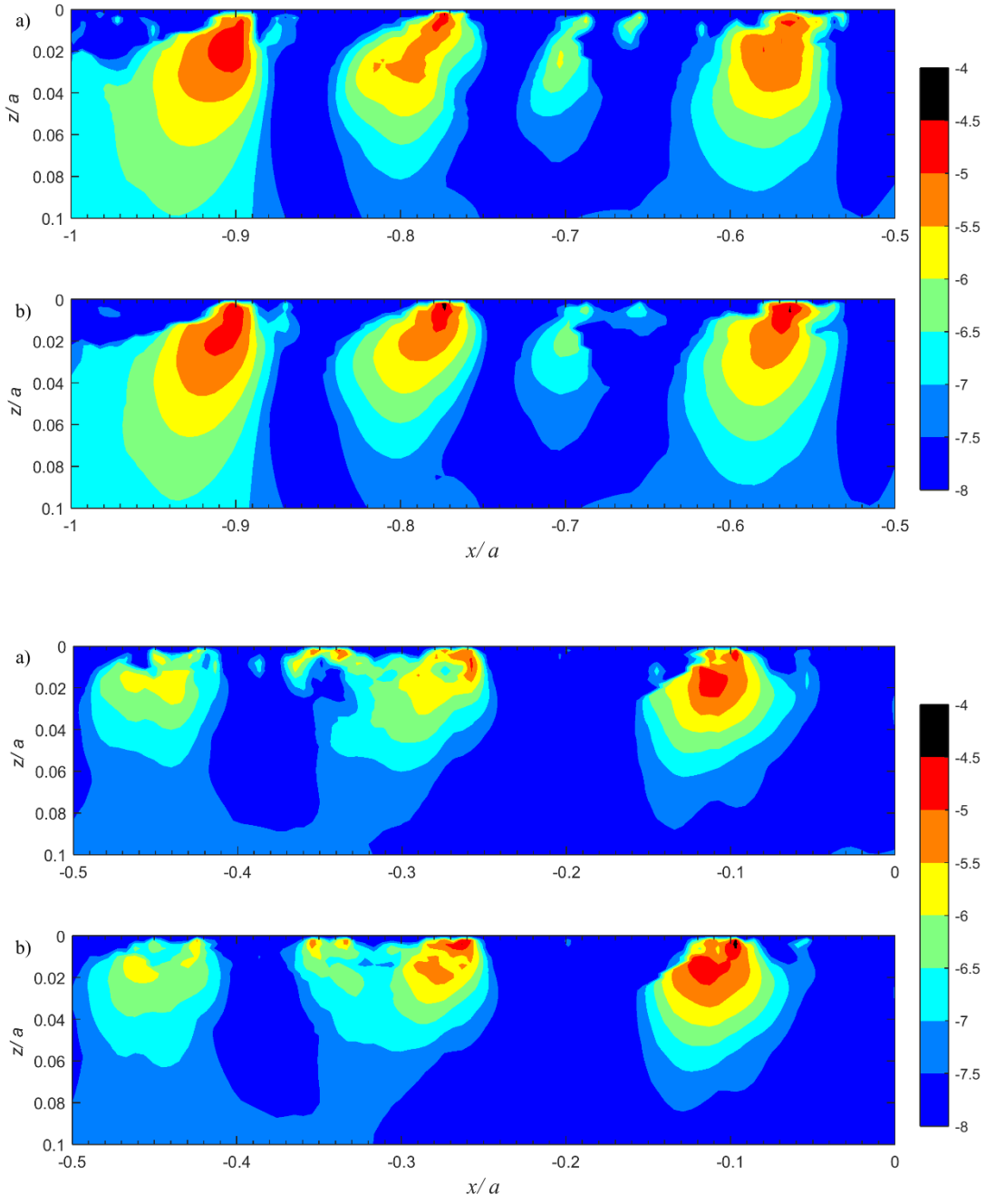


Figure 6-3 slow disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

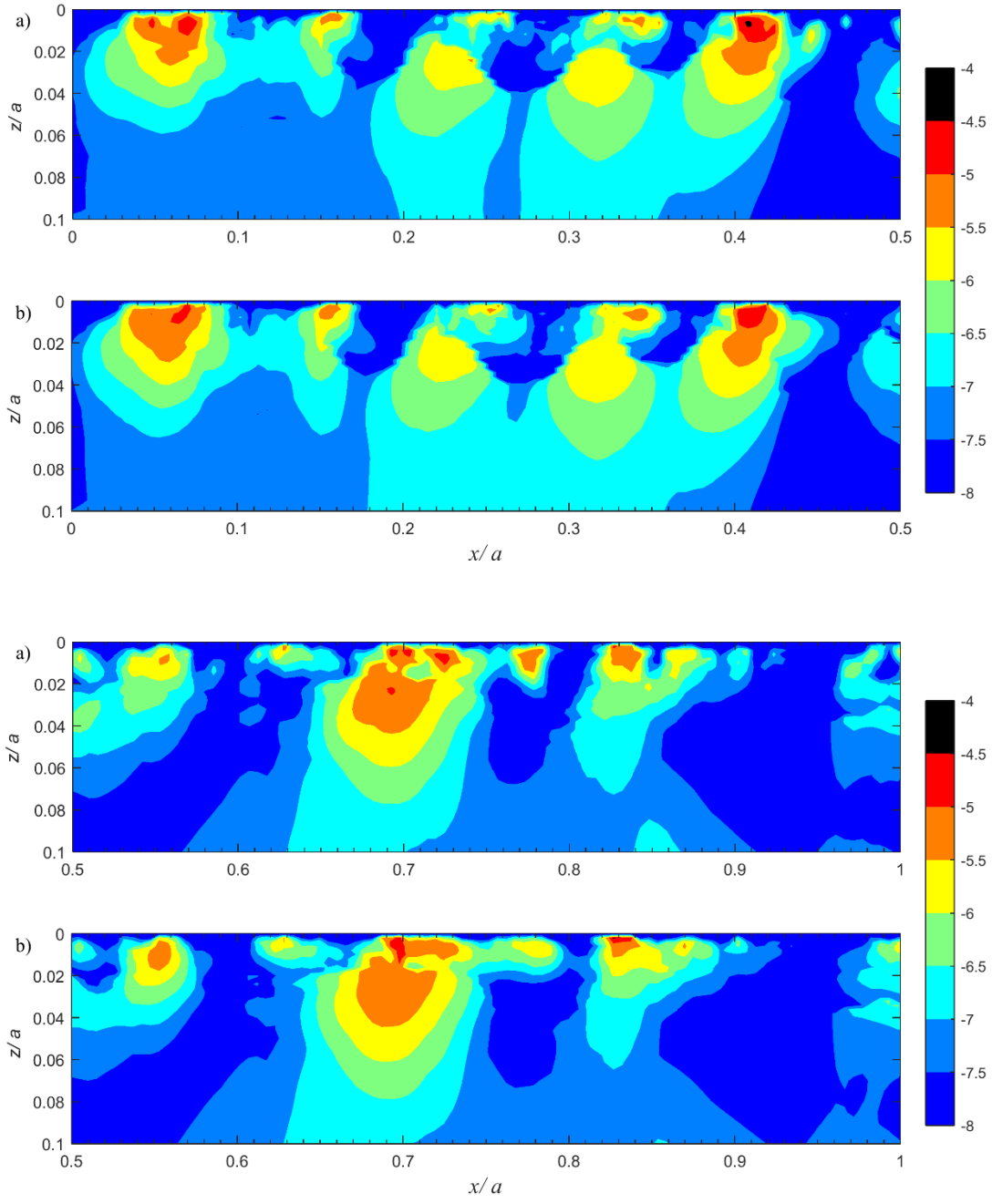


Figure 6-4 slow disk- contour of log10 (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

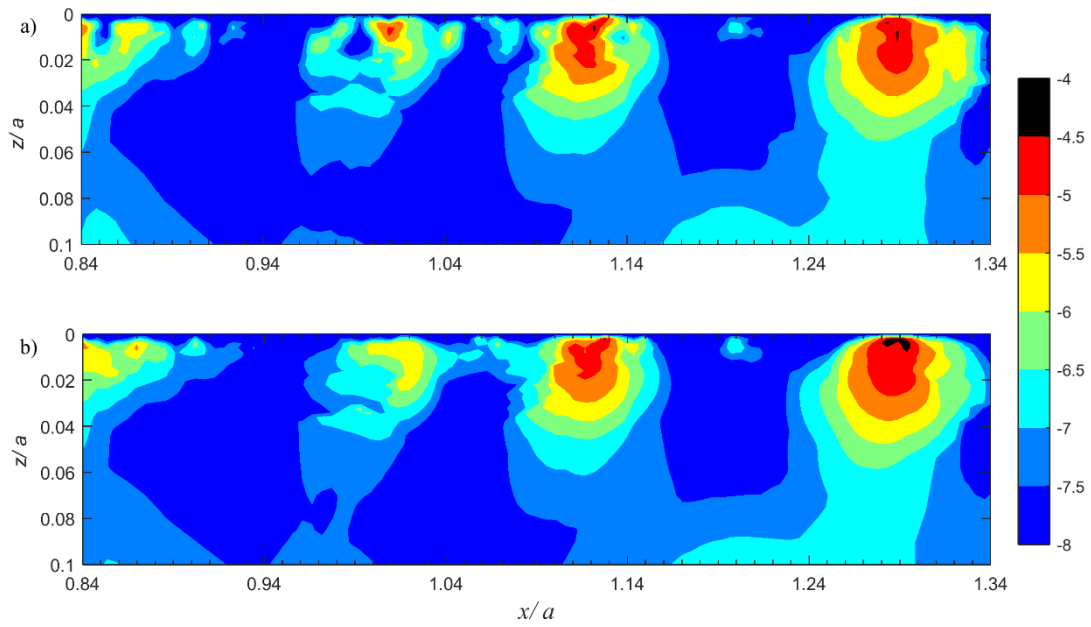


Figure 6-5 slow disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

Figures 6.6 to 6.9 compare the damage value that the points experience in the fatigue calculation with and without residual stress for the faster moving rough surface in passing through the contact zone without and with residual stresses. The length of profile considered is *2a* where *a* is equal to 0.469 mm. The contours level are chosen to vary logarithmically in value as *D* varies by up to five order of magnitude between the areas found have the most and least calculated fatigue damage. It can be seen that the accumulated damage calculated is concentrated near the surface of the prominent asperity features. It was found from the results that in some asperities more fatigue damage was apparent on the fast surface when the residual stress was included; an example of this is seen at x/a = -0.6, x/a = 0.1 and x/a = 0.25. It was found that after a relatively modest number of load cycles (approximately 10^5), a degree of fatigue had initiated near the surface of prominent asperities. Comparing the accumulated damage for the faster and slower surfaces, it is clear that the damage to the slower surface is greater than that on the faster surface. This is because asperities of the

slower moving surface are subject to higher numbers of loading cycles in EHL contact than are those of the faster moving surface.

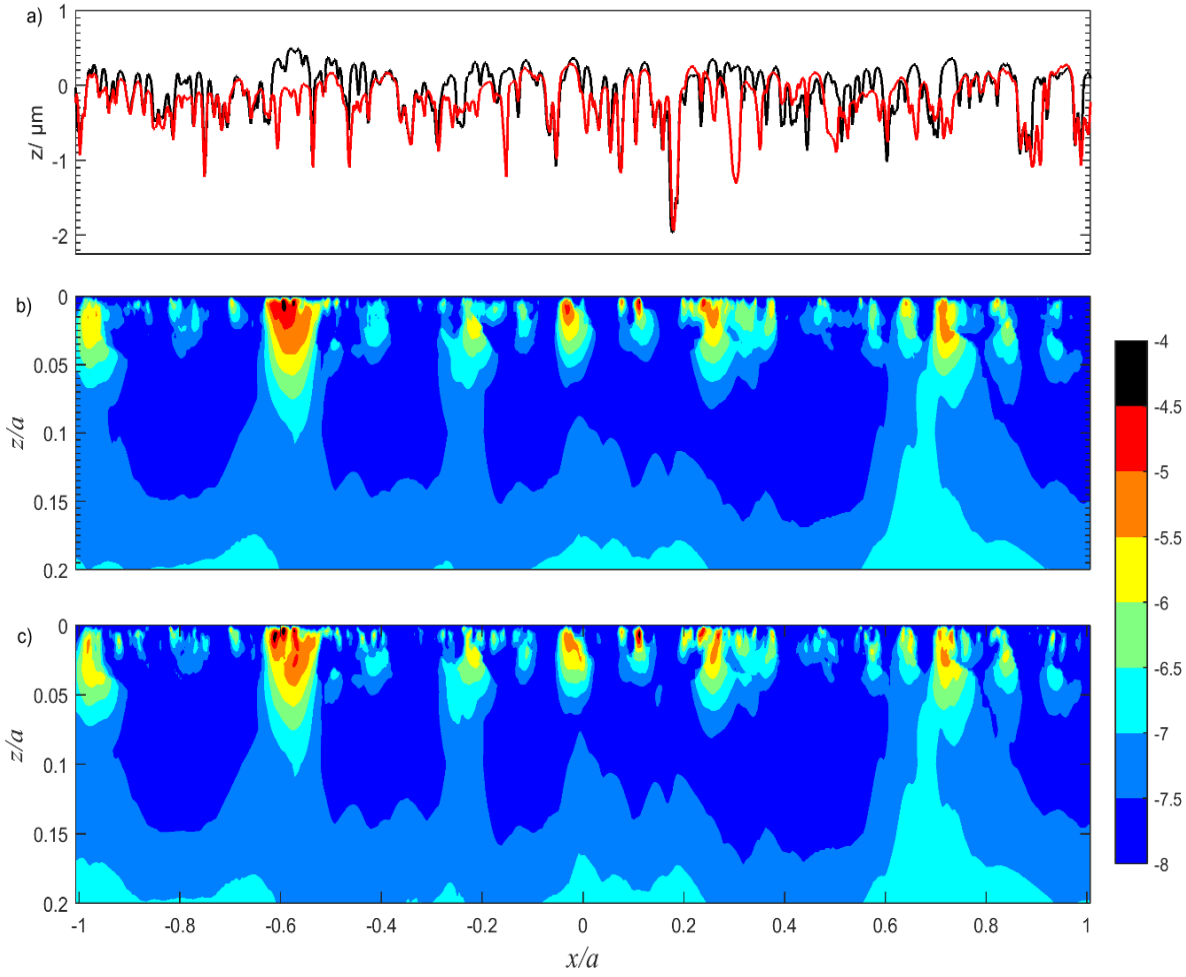


Figure 6-6 fast disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion, target load is 1.4GP: a) roughness profile; EHL- Black and Fatigue profile- Red; b) Damage without residual stress, c) Damage with residual stress, accumulated damage, 10^{-n}, indicating fatigue in 10^{n} cycles.

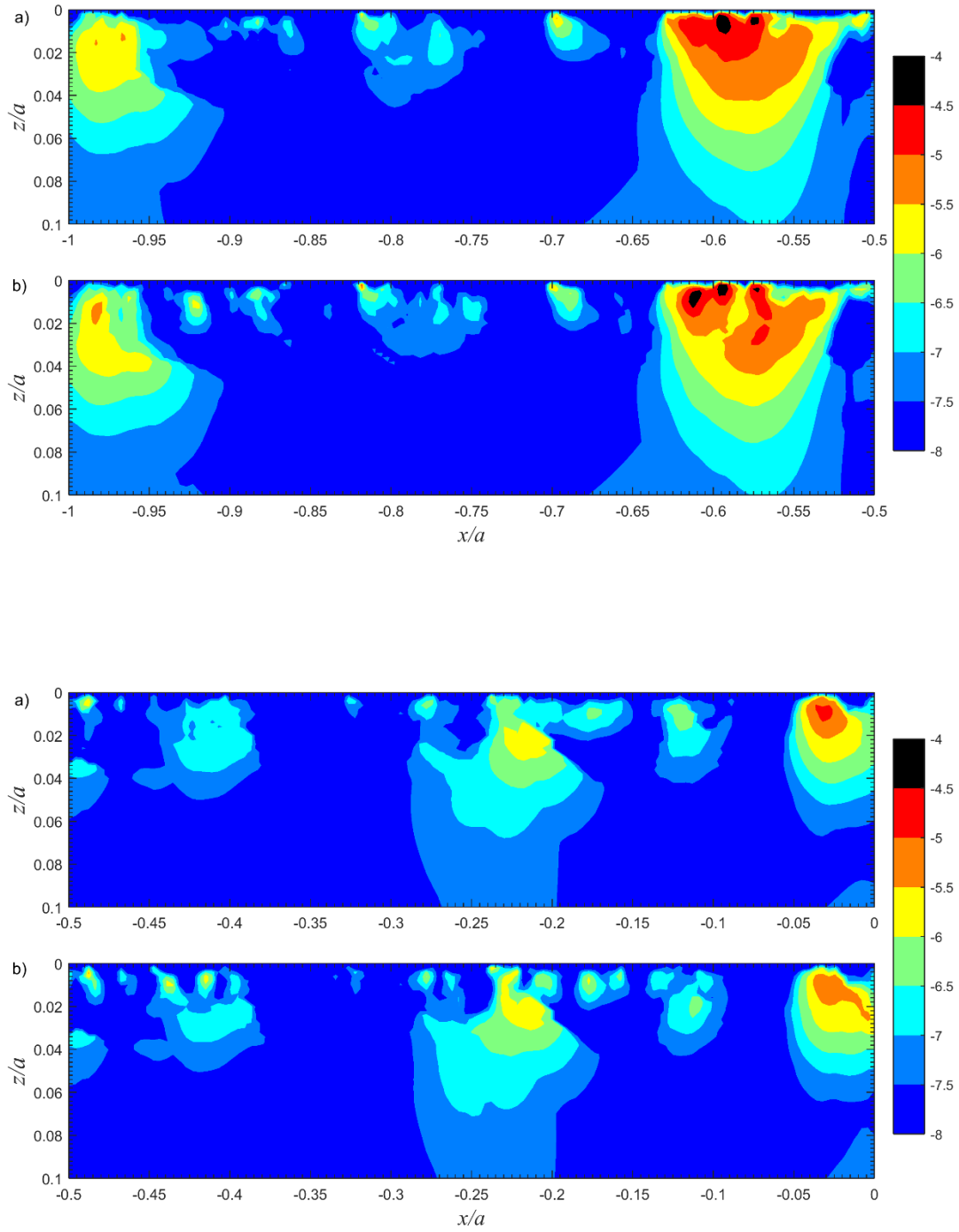


Figure 6-7 fast disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

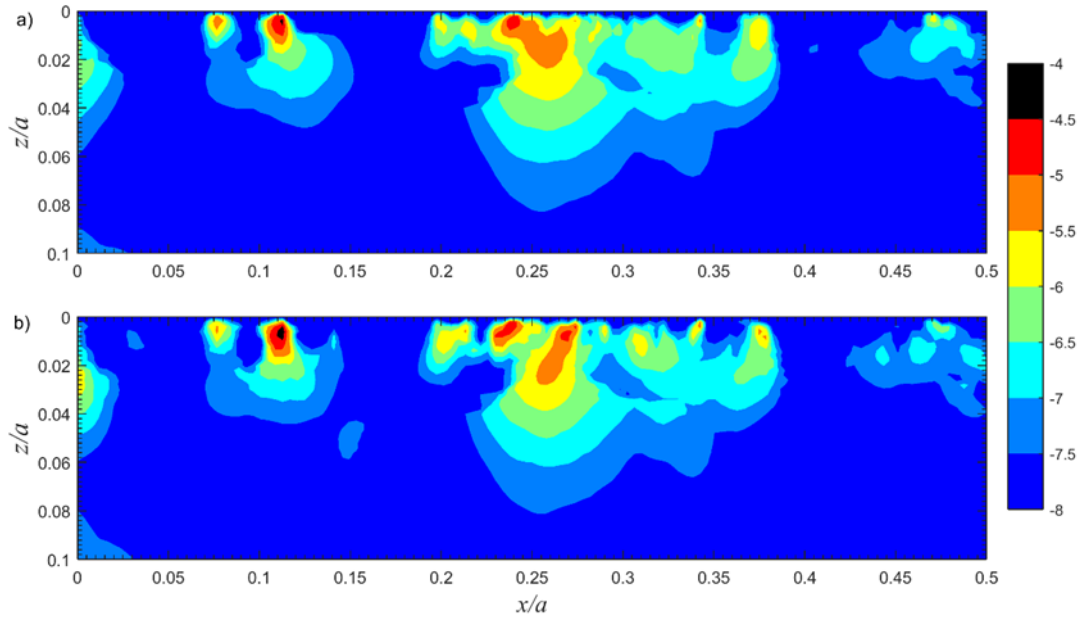


Figure 6-8 fast disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

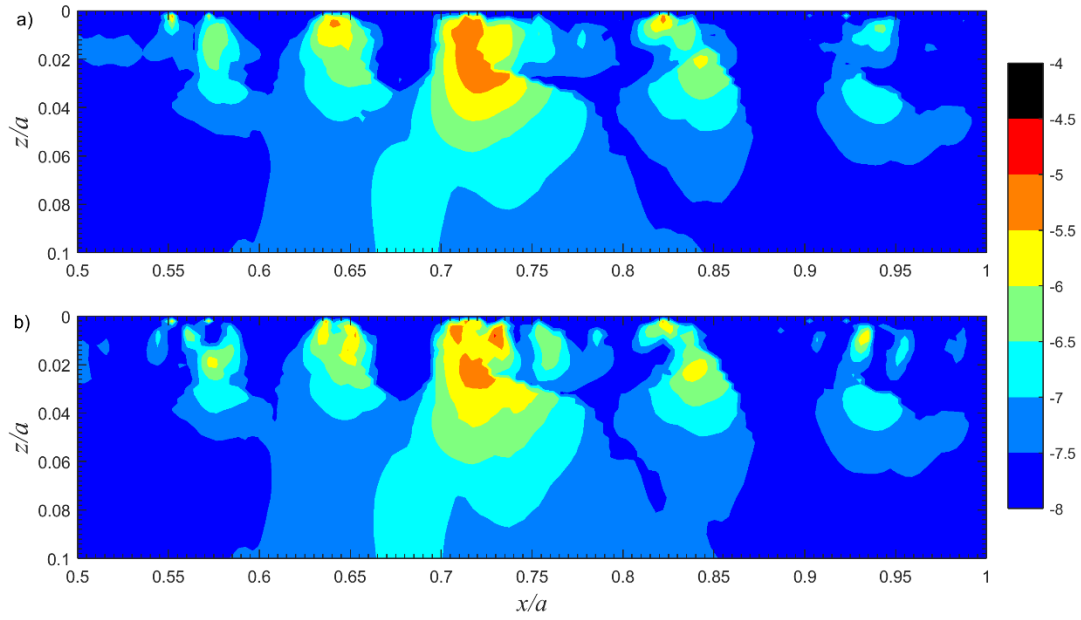


Figure 6-9 fast disk- contour of $\log_{10}$ (D) using Fatemi and Socie criterion at the close proximity to the contact surface; parts a) Damage without residual stress, parts b) Damage with residual stress.

Figure 6.10 shows the effect of different entrainment speeds on the fatigue damage distribution. It is evident from this figure that the fatigue damage increases as the speed decreases; this is because of the reduction in lubricant film thickness in the contact area. This is accompanied by a significant proportion of the fatigue damage associated with the prominent asperity features present on slow rough surface as shown in the Figure 6.10 b; an example of this can be seen at x/a =0.05, x/a =0.7 ,x/a = 1.15, x/a =1.28, x/a =-0.45 and x/a =-1.17. This is related to the pressure cycling that individual asperities undergo as they traverse the EHL contact zone.

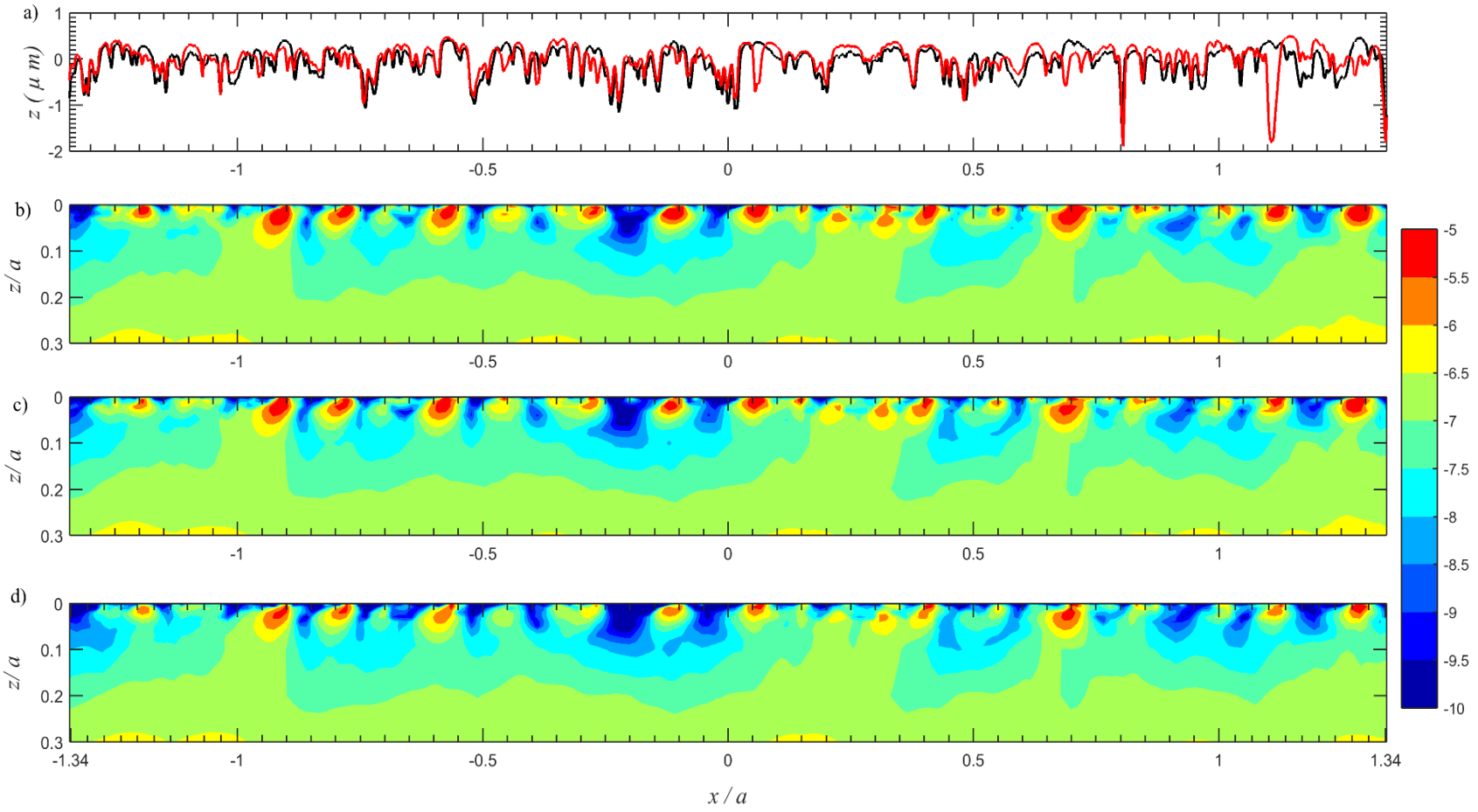


Figure 6-10 Contour of $\log_{10}(D)$ using Fatemi and Socie criterion at different entrainment speed : a) roughness profile; EHL- Black and Fatigue profile-Red, b) 0.638 m/s, c) 3.192 m/s and d) 6.384 m/s, , accumulated damage, 10^{-n}, indicating fatigue in 10^{n} cycles.

Figure 6.11 shows the cumulative Weibull distributions of damage with residual stresses in terms of *1-F* for the slower surface for a range of entrainment speeds at the maximum damage depth of $z = 0.02a$. The damage probability curves are plotted so as to give the probability that the damage level is greater than the abscissa value. The meaning of these curves can best be stated simply by reference to the probability distribution curve for entrainment speed= 0.638 m/s which passes through the accumulated damage $D = 10^{-6}$, $1-F = 0.25$. This means that about 25 % of material points at the depth $z = 0.02a$ have calculated damage equal or greater than $D = 10^{-6}$ during a single pass through the EHL contact zone. It is apparent from these curves that the calculated damage is greatest at the $z/a = 0.02$ level as might be expected from examination of Figure 6.11. However, at the greater speed, 6.384 m/s the level of damage has reduced by two or three orders of magnitude in comparison. This observation is in good agreement with the results presented in Figure 6.10.The most interesting aspect of these curves is that in each decrease in the value of entrainment speed the distribution curve shifts to the right corresponding to an increase in the expected value of *D*.

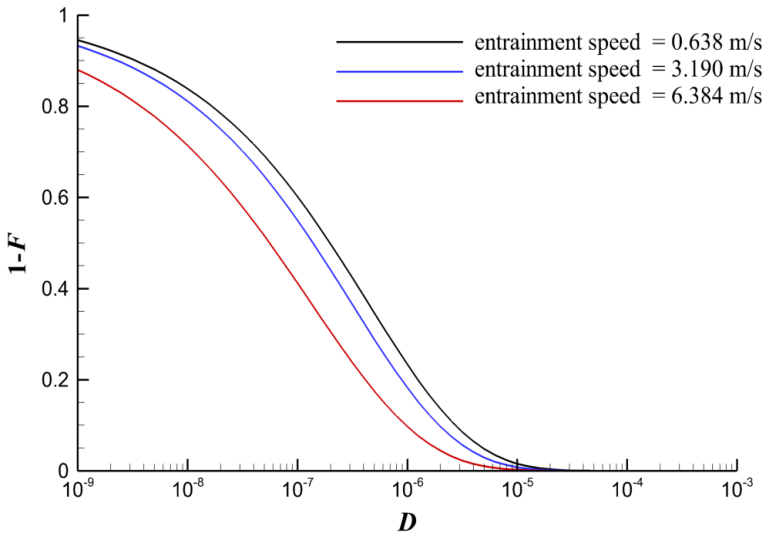


Figure 6-11 Cumulative damage distributions at a depth of *z/a* = 0.02 beneath the surface for a series of entrainment speed values.

Figure 6.12 shows comparisons of the Weibull cumulative damage distributions for the slower surface at six different depths corresponding to z/a = 0.002, 0.004, 0.01, 0.05, 0.065, 0.1 (this range includes the zone of highest accumulated damage identified in the contour maps results). These curves for cumulative damage distribution are plotted so as to give the probability that the damage level is greater than the abscissa value corresponds to a single meshing cycle (a single pass through the EHL contact). Thus for the z/a = 0.01 case, for example, about 15 % of material points at that depth experience calculated damage greater than $D = 10^{-6}$ during a single pass through the EHL contact zone, so that 15% of the material would experience predicted fatigue in 10^6 gear rotational cycles, or less at that depth.

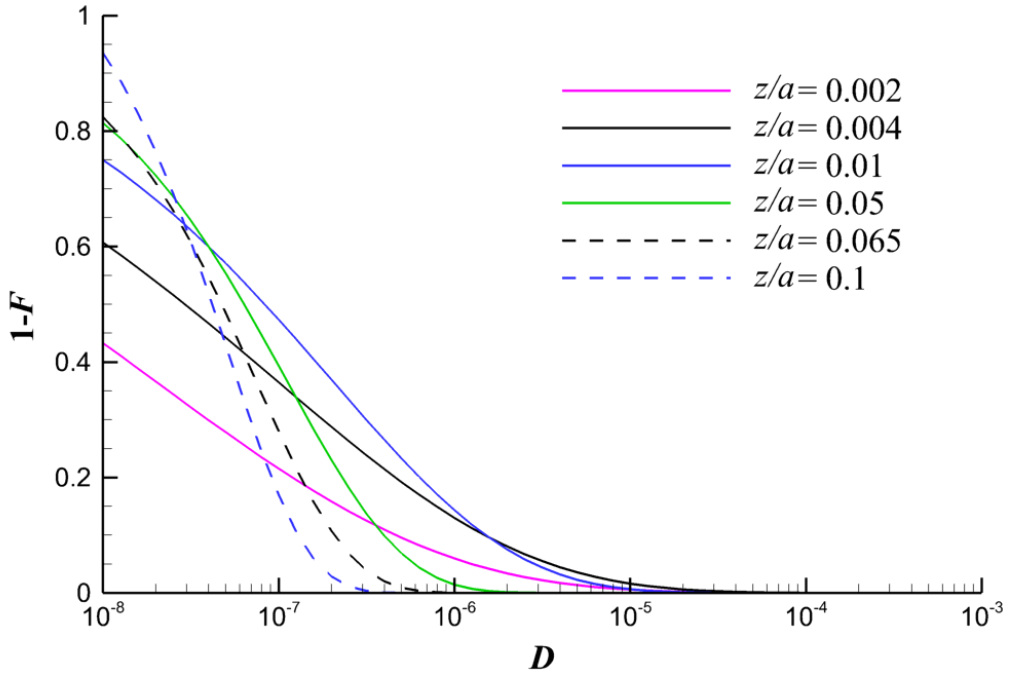


Figure 6-12 Cumulative damage distribution for the slow surface at a series of depths below the surface.

To study the effect of residual stress due to plastic deformation of rough surface, the Weibull cumulative damage distributions is used to quantify the damage at the asperity level. Figure 6.13 shows the Weibull cumulative damage distributions for the slow disk corresponding to

Figure 6.2 at two different depths corresponding to z/a = 0.02 and 0.04 (this range includes the zone of highest damage associated with the prominent asperities). Taking the curve for a depth of z/a =0.004 for damage without residual stresses (solid line in red colour), in this case, 10 % of the material at the depth considered has $D = 10^{-6}$. However, including residual stresses in the calculated damage indicates a decline in the predicted damage, for example, 3 % of the material at the depth considered has$D = 10^{-6}$. Thus, there is a clear reduction in the level of cumulative damage at that depth when the residual stresses are included by approximately 7 % thus corresponds to fatigue occurring in 10^6 repetitions of the loading experienced.

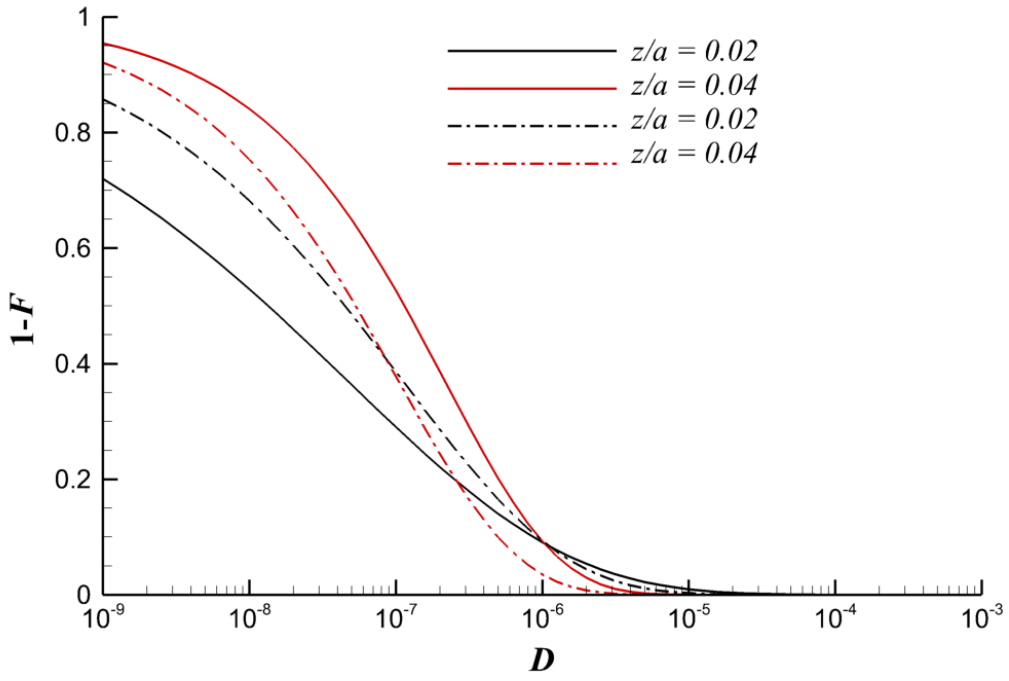


Figure 6-13 Cumulative damage distribution for the slower surface corresponding to damage with and without residual stresses at a series of depths below the surface. The solid lines are for damage without residual stresses, and the dashed lines are for damage with residual stresses.

Figure 6.14 shows the influence of changing the value of a material constant (k) in the Fatemi and Socie model for the slow disk with residual stresses. The surface roughness profile of the material volume is shown above the sequence of contour damage maps. The vertical (z) scale of the damage map is shown as a fraction of the corresponding Hertzian

contact semi-dimension *a* where the high damage zones are quite localized at that asperity level. Generally, fatigue failure zones for the most prominent asperity are coloured red when $D \geq 5.00\text{E}^{-6}$ correspond to a single meshing cycle (a single pass of the trial material through the EHL zone). It is clear that the volume of failure zones (red colour) is seen to decrease significantly with increased *k* values which are concentrated close to the surface within the approximate range $0 \leq z/a \leq 0.05$. Since there is a direct proportion between the material constant and the effect of the normal stress it is clear that the value of σ_n^{max} in equation (6.1) is negative (compressive stress) in the asperity level that has a high failure zones, so that the factor $(1+k\frac{\sigma_n^{\max}}{\sigma_y})$ is reduced as *k* is increased. Calculating the fatigue damage corresponding to $k = 0$ makes a significant difference to the value of calculated damage where the damage appears to be more extensive with the scale of surface roughness features. This is occurring because the normal stress has no effect and in this case, the Fatemi and Socie model responds only to the shear strain amplitude.

Figure 6.15 shows the cumulative Weibull distributions of damage in terms of *1-F* for the slower surface with residual stresses at a series of material constant (*k*) values and maximum damage depth of $z = 0.05\ a$. In general decreased *k* values lead to greater damage probabilities and shifts the damage probabilities curve to the right corresponding to an increase in the expected value of *D*. This observation is in agreement with the results presented in Figure 6.14.

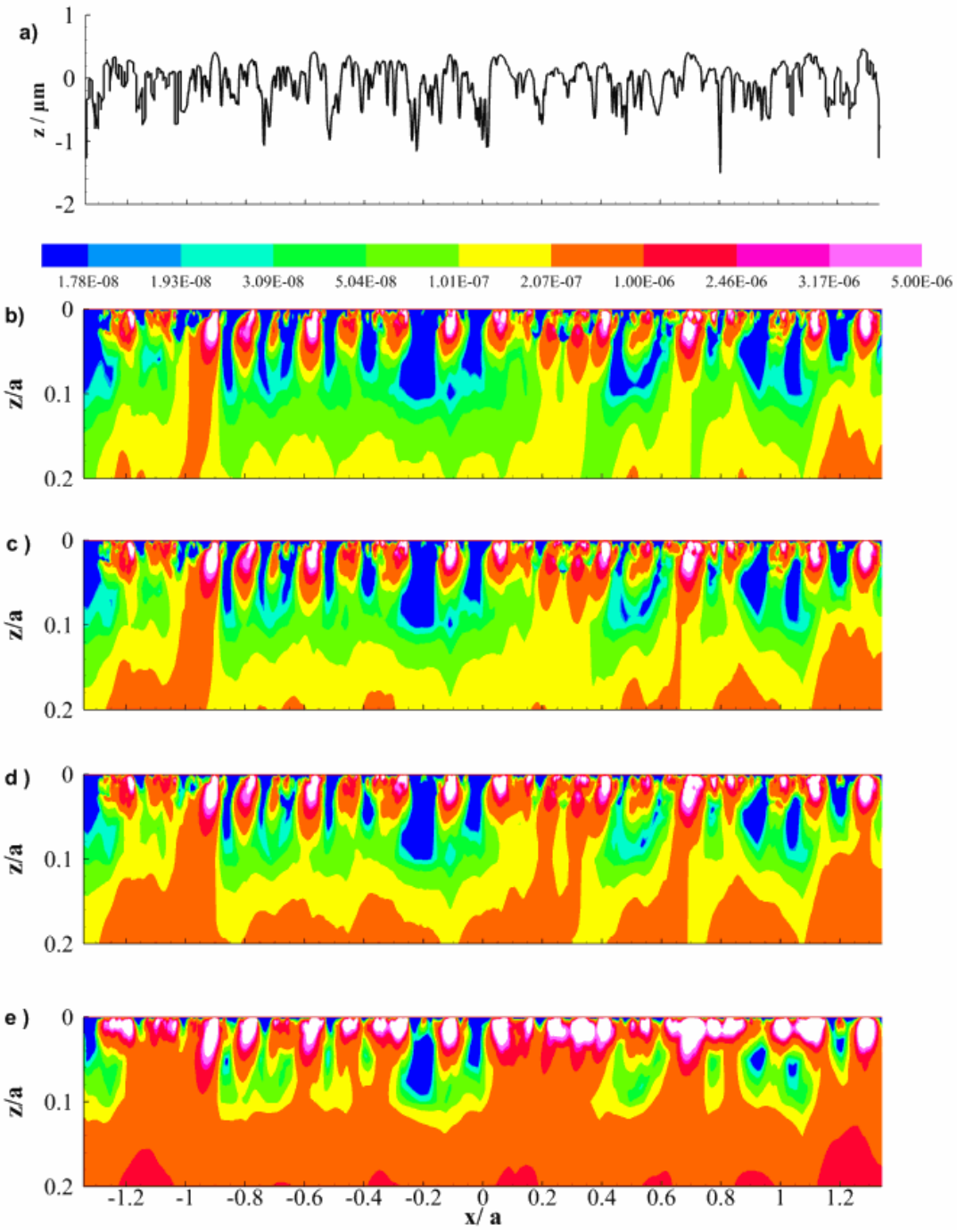


Figure 6-14 Contours of Accumulated Damage based on the Fatemi and Socie fatigue model at different material constant for slow disk with residual stresses ; a) EHL profile, b) *k*=1, c) *k*=0.8, d) *k*=0.6 and e) *k*= 0.

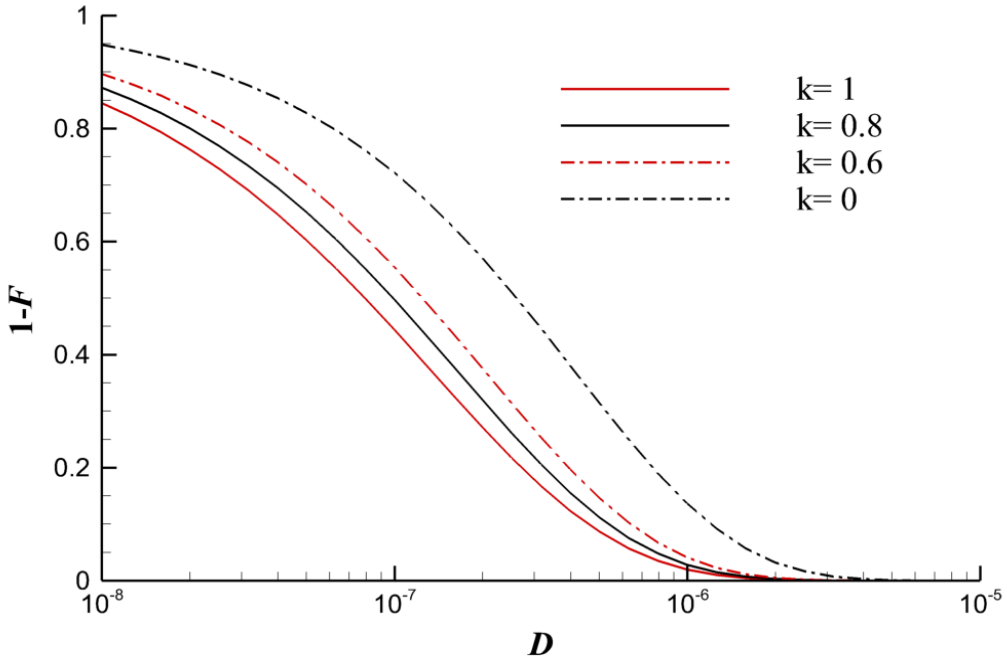


Figure 6-15 Cumulative damage distributions at a depth of *z/a* = 0.05 beneath the surface for the slower surface corresponding to damage with residual stresses for a series of material constant (*k*) values .

The effect of viscosity on fatigue damage is shown in Figure 6.16. It can be seen that the fatigue damage decreases with increased η_0.With lower viscosities, the highest damage occurs at the level of the surface roughness asperities. At the lower values of η_0 for example 0.005 regions of relatively high damage (red zones, $D > 5E^{-6}$) are concentrated close to the surface within the approximate range $z/a = 0 – 0.08$. This phenomenon is due to the reason that higher lubricant viscosity usually leads to an increase of the EHL film thickness; as a result, the interaction between the aggressive asperity decreases and thus local micro-EHL conditions, which occur at the contact interface, becomes weak. This conclusion is in good agreement with the significant work performed by Rico et al. (2003).

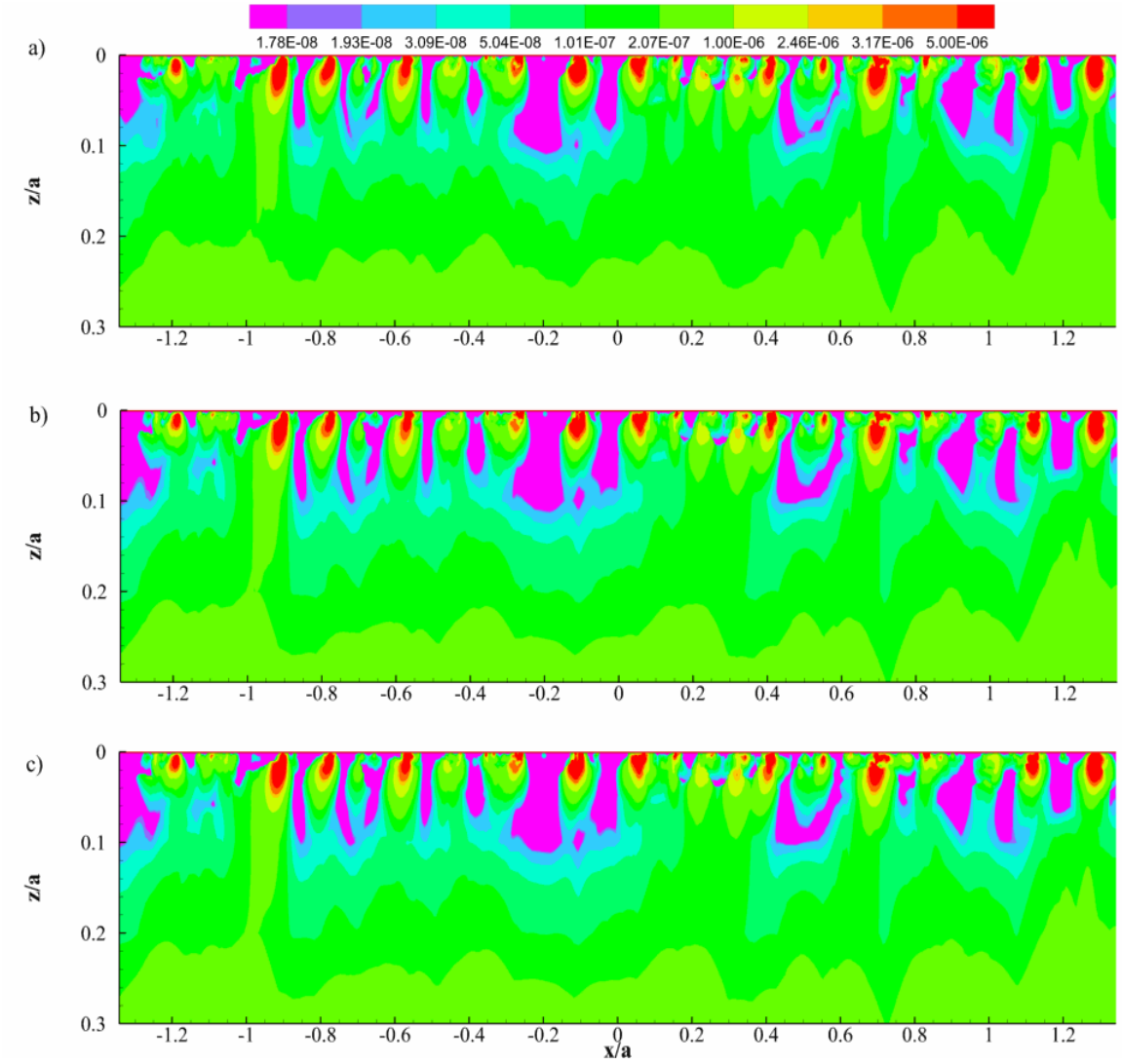


Figure 6-16 contours of subsurface damage calculated for the section of the slower moving surface for three η_0 values; a) 0.005 Pa.s, b) 0.015 Pa.s and c) 0.0277 Pa.s.

6.6.2 Results for a micropitting endurance test

The influence of roughness on the fatigue damage can be further investigated by the use of different rough profiles and different working conditions. This was achieved by modelling test disk profiles from an endurance test carried out by Hutt (2018) at steady state load and speed conditions. Profile data from the running in process was acquired from the fast and the slow test disks and used for the EHL analysis at the constant experimental conditions. The test was performed at a single speed to remove speed and temperature dependent variations from the predicted life.

Contours of damage obtained for a section of material of $2a \times a$ within the slower and faster surface are shown in Figures 6.17 to 6.18, respectively. The upper part of each figure is the corresponding shape of the rough profile considered which was measured after the first load stage and is run-in. Also shown is the measured profile after 100,000 cycles (LS4) and after 1,000000 cycles (Fatigue profile). It is clear that the asperities sustain calculated damage levels that are significantly higher than those experienced by the surrounding material. The profiles, which were taken from the micropitting test at a slide to roll ratio of 0.5, show the first formation of new valley features on both surfaces occurred after a relatively low number of load cycles of the order 30,000 on both surfaces (In the experiment the previous measurement stage was at 30,000 cycles). Examples of this can be seen in Figure 6.17 at x/a = -0.8, x/a = 0 and x/a = 0.65, and in Figure 6.18 at x/a = -0.25, x/a = -0.18 and x/a = 1. Closer inspection of the profiles in Figures 6.17 and 6.18 show that micro-pits were found across the entire width of the contact area and tended to be concentrated at positions of previously prominent asperities. In addition, it can be seen from these figures that the accumulated damage calculated is concentrated near the surface of the prominent asperity features within the approximate range $0.0 < z/a < 0.08$.

The micropit damage features are seen to be present at the 100,000 cycles measurement stage and have been initiated at a lower number of cycles. The results of the damage accumulation model indicate a life between 10,000 and 32,000 cycles for the brown contour area, and between 32,000 and 100,000 cycles for the red contour area. The model results are therefore of the same order as the observations. This is encouraging given that the model is for two run-in surface profiles that in all probability did not run against each other in the actual micropitting test.

Figure 6.19 compares the damage contours for the slower surface obtained for a series of η_0 values; that is, η_0= 0.002, 0.006, 0.015 and 0.027 Pas corresponding to slide roll ratio= 0.5. The regions of failure zones of relatively high damage (red zones, $\geq 10^{-5}$) are seen to decrease with increased η_0. It is clear that increasing viscosity will improve the fatigue performance. This happens because increasing viscosity usually leads to a significant increase in lubricant film thickness and as a result the interaction between the asperities will decrease leading to a reduction in the intensity of asperity interactions. A 14 fold increase in viscosity to 0.027 Pa s causes increasing in the calculated life of the points subjected to the highest damage by approximately two orders of magnitude.

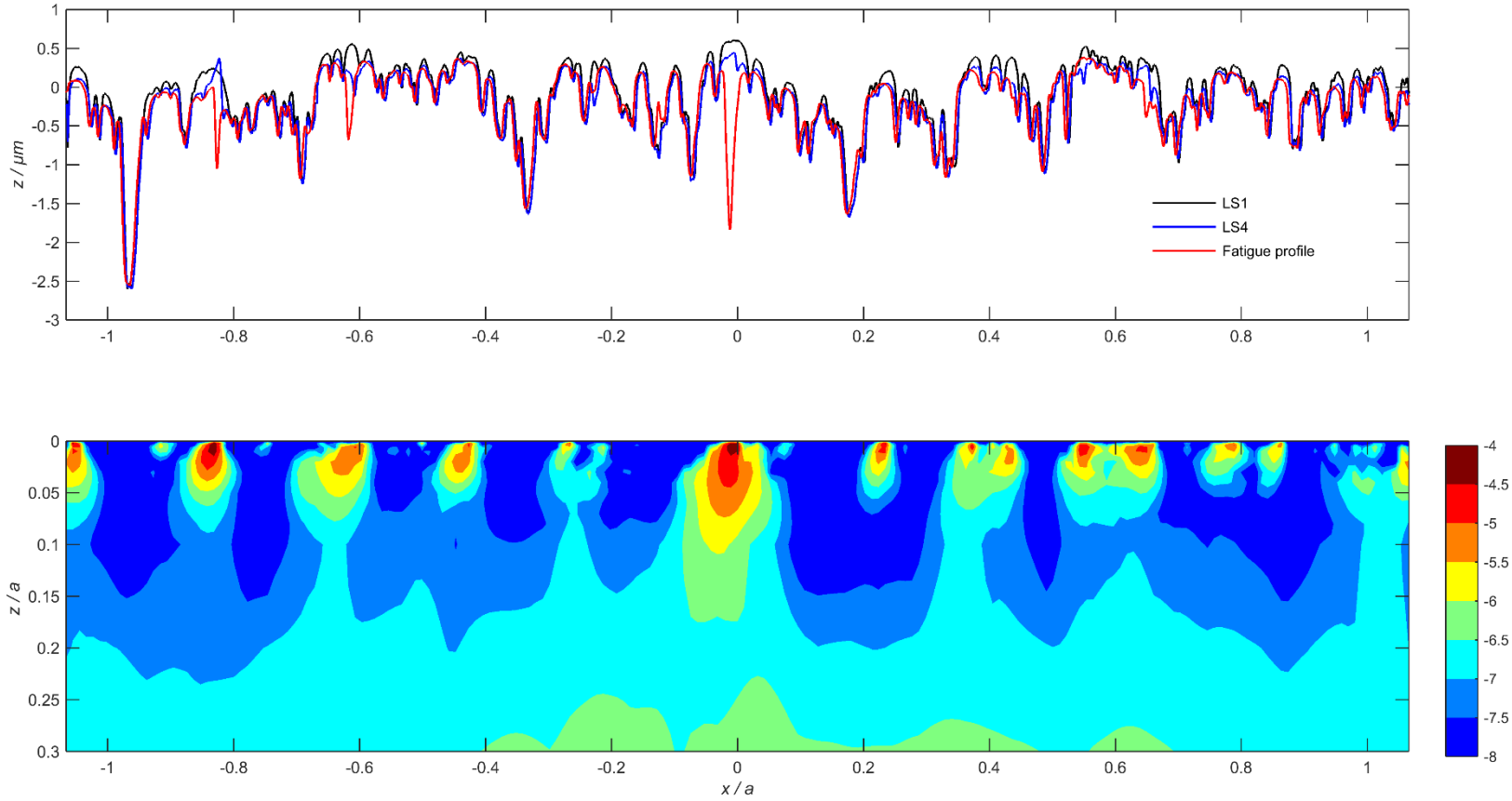


Figure 6-17 Slow disk—Contour of $\log_{10}$ (D) using Fatemi and Socie criterion, accumulated damage, 10^{-n}, indicating fatigue in 10^{n} cycles.

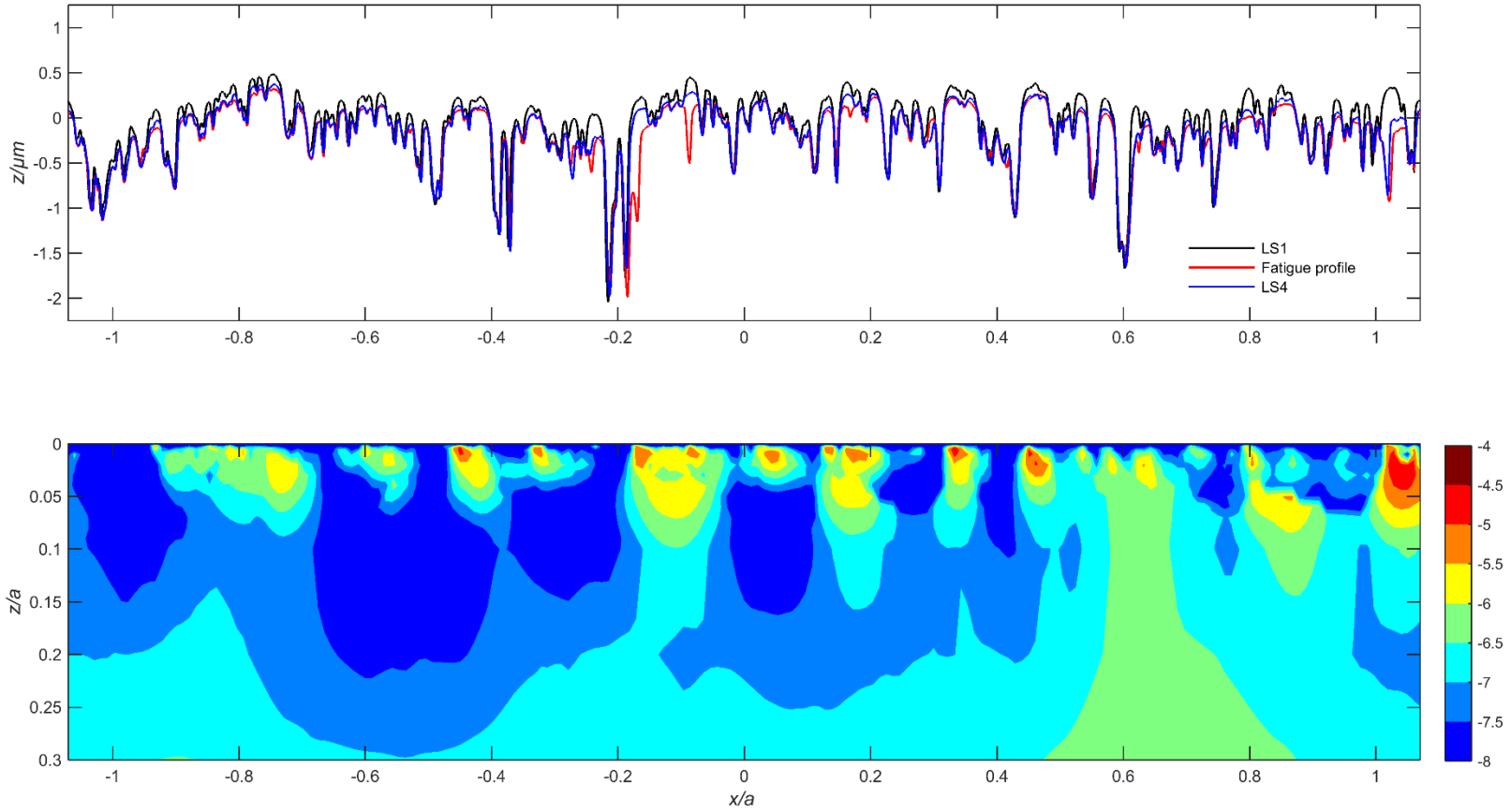


Figure 6-18 fast disk—Contour of $\log_{10}(D)$ using Fatemi and Socie criterion, accumulated damage, 10^{-n}, indicating fatigue in 10^{n} cycles.

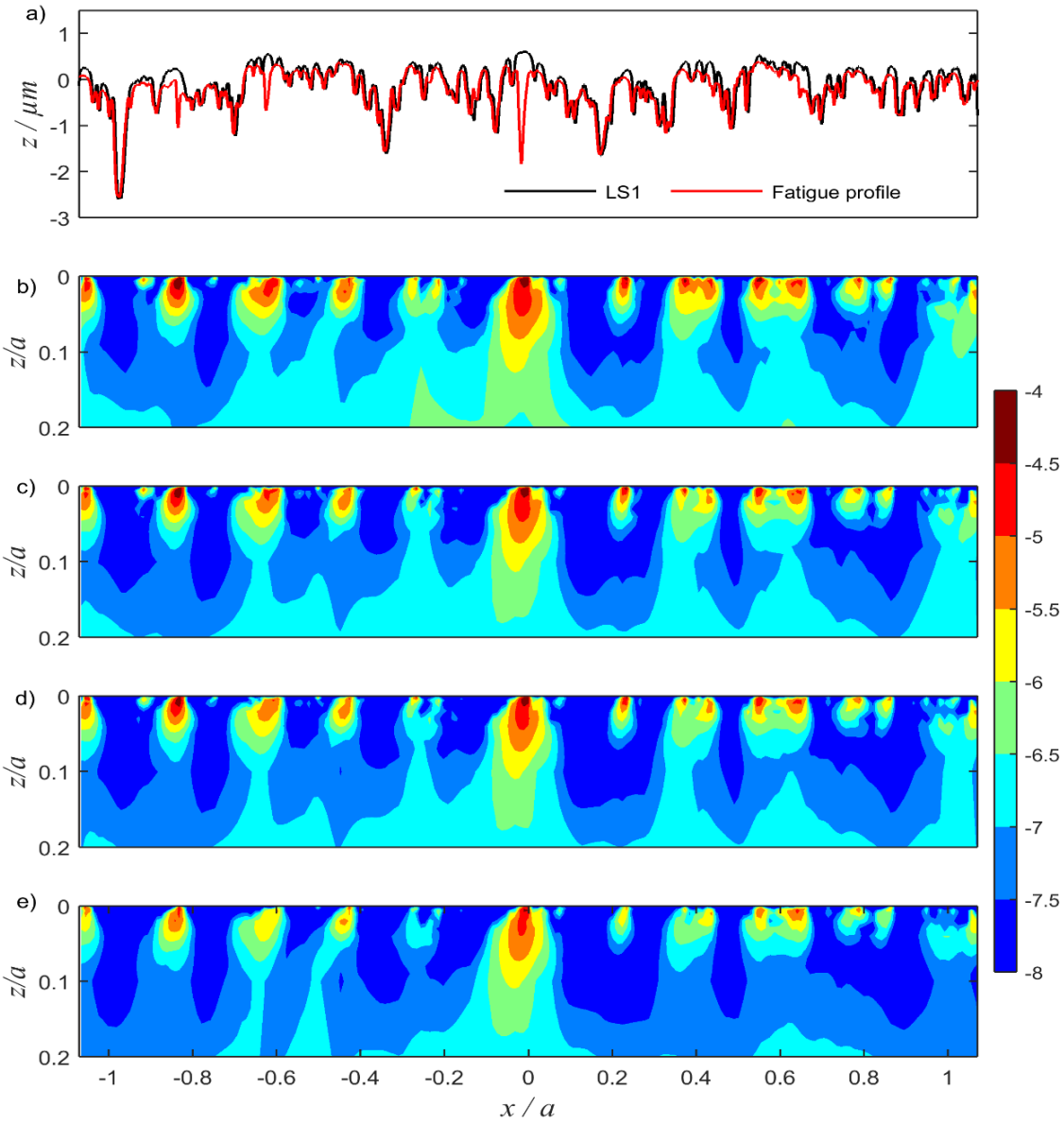


Figure 6-19 Contours of $\log_{10}$ (D) of sub-surface damage calculated for the section of the slower moving surface for different lubricant viscosity; b) η_0= 0.002 Pa.s, c) η_0= 0.006 Pa.s, d) η_0= 0.015 Pa.s and e) η_0= 0.027 Pa.s.

The effect of η_0 on damage probabilities is seen in Figure 6.20, which shows the cumulative damage distributions at a high damage depth level of z/a = 0.03 for each value of η_0 considered. It can be seen that the fatigue damage decreases with increased η_0 and shifts the damage probabilities curve to the right, corresponding to an increase in the expected value of D. This observation is in good agreement with the results presented in Figure 6.19. It is

clear that introducing the lowest viscosity lubricant considered to the contact causes a reduction in the calculated damage by at least two orders of magnitude.

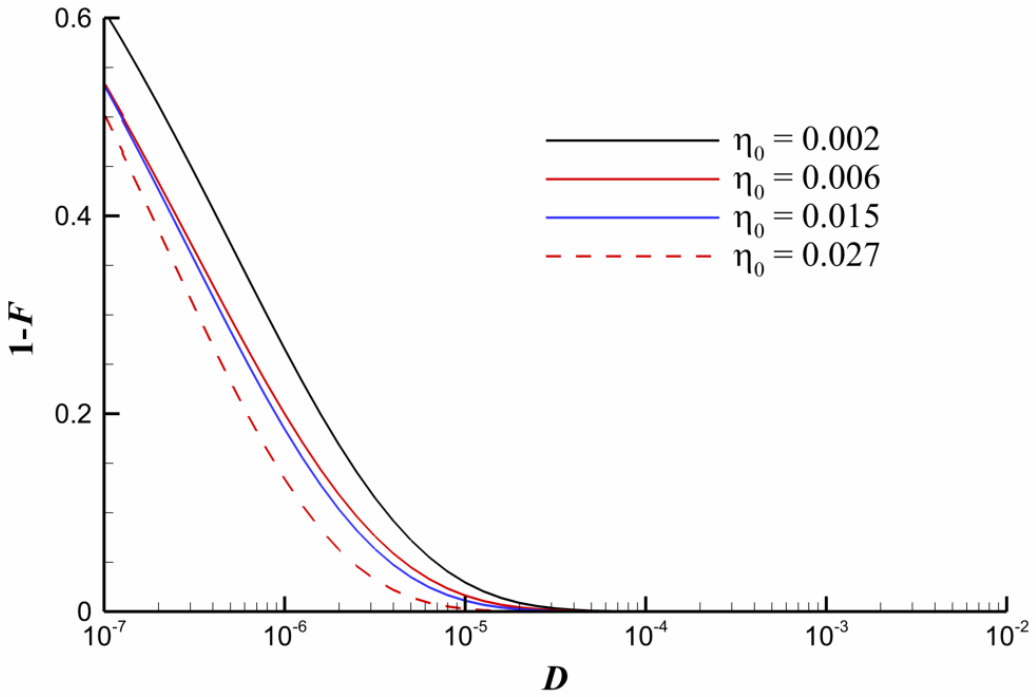


Figure 6-20 Cumulative damage distributions for slow surface at a depth of z/a = 0.03 beneath the surface for a series of η_0 values, the viscosity unit is Pa.s.

6.7 Conclusion

A numerical procedure to analyse the cumulative fatigue damage in a rough EHL contact has been developed. The procedure is based on the critical plane approach and the Fatemi-Socie shear strain model. The Palmgren – Miner linear damage accumulation rule is used to compute the cumulative damage for each effective loading cycle. The numerical results have shown that the residual stress due to running-in of asperities does cause changes in the damage prediction and calculated fatigue life of the rough surfaces. These changes can be significant at some asperities but are not necessarily detrimental. The numerical results have also shown that the volume of failure zones is seen to decrease significantly with increased *k* values, which are concentrated close to the rough surface. The slower moving surface has a greater tendency to become micropitted at the asperity level. This is because the slower rough surface is subject to higher numbers of stress cycles during contact analyses compared to the faster rough surface. This is in agreement with the results obtained by Qiao et al. (2008), and Alshahrany, (2015). The use of high-viscosity lubricant generates has a greater resistance to micropitting because of their thicker EHL films. Therefore, higher loads will require higher viscosity, while lower loads allow for lower viscosities. Speed can also have an effect on micropitting and surface fatigue. At lower speeds, the film thickness will decrease, whilst, at higher speeds, the film thickness can increase.

Chapter 7 conclusions and future work

7.1 Conclusions

The present study makes several contributions towards an understanding of the basic mechanism of micropitting initiation in gear tooth surfaces based on incorporating the asperity residual deformation and residual stress field into fatigue simulations under realistic operating conditions. Various aspects of this procedure are presented in detail in the previous chapters.

Numerical simulation of the fatigue model depends on the results of the transient micro-EHL solutions for fast and slow rough surface disks in rolling/sliding contact. Therefore, micro-EHL simulations were carried out using the coupled solution method incorporating a differential deflection technique. Real (or measured) roughness is included in a transient line contact analysis. The numerical results of the EHL model show that the time dependent pressure experienced at the encounters between run-in surface asperities are far higher (often by a factor of three or more) than estimated on the basis of the theory of smooth surfaces (Hertzian).

Elastic stress analysis was performed for the line contact problem (2*D* plane strain) to evaluate sub-surface elastic stress and strain distribution histories experienced by the contacting components. Numerical results have shown that the areas of high stress cycling located at a shallow depth are consistent with the observed depth of micropits.

A simulation of repeated loading at the same load and surface was performed to approximate the running-in process that takes place when freshly manufactured surfaces are first brought into contact. Simulation of further loading/unloading/loading cycles at the same load shows that the overwhelming majority of the changes in the residual asperity shape occur in the first loading cycle and there is an insignificant amount of additional plastic deformation of

the asperities due to the second application of the same load. The findings of this numerical investigation confirmed the findings of a recent experimental study described by Clarke et al. (2016).

Abaqus finite element analysis (FEA) was used to perform a dry contact elastic/plastic analysis based on real surface roughness profiles, which were taken from two steel disks that were used in experiments where the roughness profiles were acquired from profilometer traces. The numerical results have shown that regions of surface and subsurface residual tensile stress were found to occur close to heavily loaded asperity contacts. The effects of including such asperity residual stresses in fatigue calculations were examined by carrying out a range of fatigue analyses, both with and without the residual stress field.

Fatigue damage with and without the residual stress field was determined to establish the possible effects of asperity plastic deformation on subsequent high-cycle fatigue behaviour. This evaluation was performed by applying multiaxial fatigue criteria (the Findley criterion, the Matake criterion, and the Dang Van criterion), and the Fatemi and Socie (1988) shear strain model utilising the rainflow counting approach to break down the load history into effective loading cycles and the cumulative damage theory to determine the life of each part of the material. The main findings to emerge from this study are that:

- Including asperity residual stress in fatigue calculations does cause significant changes in the damage prediction and calculated fatigue life for the rough surfaces. These changes can be significant at some asperities, but they are not necessarily detrimental.
- The slower rough surface has a greater tendency to become micropitted at the asperity level when compared with the faster surface. This happens because the

asperities of the slower surface are subject to higher numbers of stress cycles during their passage through the EHL contact zone than are those of the faster surface.

- High-viscosity lubricant has a greater resistance to micropitting because of its thicker EHL films.
- Speed can have a significant influence on fatigue lives and damage accumulation where the results demonstrate that increased speed has a positive effect on the fatigue lives, as thicker lubricant films are generated leading to less severe asperity interactions.
- The line contact surface loading response can be significantly different for the two rough surfaces, which have nominally the same level of roughness.
- Plane strain fatigue and damage accumulation analysis shows that the predicted damage with and without residual stress fields is concentrated at asperities that are only a few microns deep. This supports the hypothesis that micropitting failure in gear teeth is mostly due to fatigue occurring at the level of the surface roughness asperities.
- Prominent asperities are subject to a high risk of damage because the damage value is affected by the asperity height and shape.
- Some of the experimentally observed micropits were found to occur at positions that had high calculated fatigue damage.
- The most obvious finding to emerge from this study is that the micro-EHL model results and Micropitting test results of the contacting disks were shown to agree well in terms of predicting the number of loading cycles that are required for the initial micropitting to occur.
- All the plastically deformed region features identified by Oila and Bull (2005) in their material metallurgical analysis can be seen to correspond to the circular band

area of the residual stress and the corresponding damage as shown previously in the chapters 5 and 6.

- Large amounts of plastic deformation of asperities resulted in surface and subsurface regions of high residual tensile stress at the asperity level. The position of these residual tensile stresses aligns well with the position at which cracks initiate in the detailed experimental studies of Oila and Bull (2005).

7.2 Future work

- To allow simple modelling of a plane strain problem by using Abaqus software, the asperity features were considered to be two-dimensional. However, the asperity features are usually three-dimensional and modelling them as such will provide the most accurate representation of modelling the asperity contact problem.
- It may be necessary to model higher asperity overlaps and also incorporate the effect of sliding asperity interaction to be representative of an experimental test condition that is based on the physical mechanisms of material plasticity and surface traction effects.
- Further research should focus on measuring residual stresses for contacting surfaces asperities in-depth and in both longitudinal and transversal directions, especially in the early contact stages of the running-in process. Modelling residual stresses is considered to be a significant step for understanding how the residual stress distribution evolves during the plastic deformation process.
- In the EHL model, asperity features were considered to be two-dimensional for simplicity. Although this is not an unreasonable assumption, more research using three-dimensional roughness state is needed to model the EHL and rolling contact fatigue and, thereby, improve our ability to efficiently predict micropitting damage.
- The effects of surface chemistry and tribofilms during asperity/asperity contacts have not been considered in the current lubricated contact model. Where they act as a 'third body' they can have a remarkable influence on both the friction behaviour and the micropitting performance. These findings suggest that it will be necessary to combine the solid and fluid mechanics of rolling/sliding contacts with the behaviour of interacting surface asperities at the molecular level.

- Much uncertainty still exists about the relationship between the mechanical effects, such as the influence of internal fluid pressure on crack growth, and the effect of material discontinuities at grain boundaries on the EHL problem. Therefore, it would be interesting to assess the effects of such issues to enhance our understanding of micropitting initiation. These damages could be related to the effect of lubricating oil seepage inside cracks as discussed by Omidvar and Ghorbanpoor, (1998).
- A thermal analysis of the EHL solution should be included, especially at high rolling speed and high slide roll ratio, where such thermal influences may become significant.

References

Abaqus, 2010 a. Abaqus analysis: user's manual. Providence, RI, Abaqus Inc.

Abaqus, 2013. Abaqus analysis: *user's manual. Providence*, RI, Abaqus Inc.

Ahmed, R., 2002. Rolling contact fatigue. *ASM handbook*, *11*, pp.941-956.

Ai, X. and Cheng, H.S., 1994. A Transient EHL Analysis for Line Contacts with Measured Surface Roughness Using Multigrid. *Journal of tribology*, *116(3)*, pp.549-556

Akbarzadeh, S. and Khonsari, M.M., 2011. Experimental and theoretical investigation of running-in. *Tribology International*, *44*(2), pp.92-100.

AL-Mayali. M. F., Evans, H.P. and Sharif, K.J., 2016, October. Assessment of the effects of residual stresses on fatigue life of real rough surfaces in lubricated contact. In Students on Applied Engineering (ISCAE), International Conference for (pp. 123-128). IEEE.

Almuramady N. and Borodich F. M., 2016. Adhesive Contact between Silicon-Based Mems Tooth Surfaces Modelled by the Multiscale Multi-Block Model. *Proceedings of the 1th International Conference on Advances in Automotive Technologies 2016, 11-14 October 2016, Yildiz Technical University, Istanbul, Turkey*, ISBN: 978-605-9546-01-0, pp 129–135.

Almuramady N. and Borodich F. M. 2017. Adhesive Contact between Silicon-Based Mems Tooth Surfaces Modelled by the Multiscale Multi-Block Model. *International Journal of Advances on Automotive and Technology*, 1, No. 2, pp. 59-66. doi: 10.15659/ijaat.17.04.523.

Alshahrany, S., 2015. *Rolling contact fatigue in heavily loaded gear transmission contacts* (PhD dissertation, Cardiff University).

Amzallag, C., Gerey, J.P., Robert, J.L. and Bahuaud, J., 1994. Standardization of the rainflow counting method for fatigue analysis. *International journal of fatigue*, *16*(4), pp.287-293.

Anthes, R.J., 1997. Modified rainflow counting keeping the load sequence. *International journal of fatigue*, *19*(7), pp.529-535.

Aslantaş, K. and Taşgetiren, S., 2004. A study of spur gear pitting formation and life prediction. *Wear*, *257*(11), pp.1167-1175.

ASTM E1049, 2011. *Standard practices for cycle counting in fatigue analysis*.

ASTM STP 1122, 1992. *Advances in Fatigue Lifetime Predictive Techniques.*

Bannantine, J., 1990. *Fundamentals of metal fatigue analysis*. Prentice Hall.

Bannantine, J.A. and Socie, D.F., 1992. A multiaxial fatigue life estimation technique. In *Advances in fatigue lifetime predictive techniques*. ASTM International.

Barber, G.C., Lee, J.C. and Ludema, K.C., 1987. Materials and surface finish effects in the breaking-in process of engines. *ASME J. Eng. Gas Turbines Power*, *109*, pp.380-387.

Barus, C., 1893. Isothermals, isopiestics and isometrics relative to viscosity. *American journal of science*, 266, pp.87-96.

Batdorf, S. B., and Budiansky, B., 1949. *A Mathematical theory of plasticity based on the concept of slip*. Washington, D.C., National Advisory Committee for Aeronautic.

Beheshti, A. and Khonsari, M.M., 2012. Asperity micro-contact models as applied to the deformation of rough line contact. *Tribology International, 52*, pp.61-74.

Bhushan, B. ed., 1998. *Handbook of micro/nano tribology*. CRC press.

Bhushan, B., 2000. *Modern tribology handbook, two volume set*. CRC press.

Bhushan, B., 2013. *Introduction to tribology*. John Wiley and Sons.

Bishop, I. F. and Snidle, R. W. 1982. Paper III(iii) - Some experimental aspects of running-in and scuffing failure of steel disks operating under elastohydrodynamic conditions. *In:* Dowson, D., Taylor, C. M., Godet, M. and Berthe, D. (eds.) *The Running-in Process in Tribology.* Butterworth-Heinemann.

Blau, P.J., 2006. On the nature of running-in. *Tribology International*, *38*(11), pp.1007-1012.

Boardman, B., 1990. *Fatigue resistance of steels*. *In: Metals Handbook, Tenth Edition, Volume 1Properties and Selection: Irons, Steels, and High-Performance Alloys*. Prepared under the direction of the ASM International Handbook Committee, pp 673-688.

Bosman, R., Hol, J. and Schipper, D.J., 2011. Running-in of metallic surfaces in the boundary lubrication regime. *Wear*, *271*(7), pp.1134-1146.

Brandao, J.A., Martins, R., Seabra, J.H. and Castro, M.J., 2012. Surface damage prediction during an FZG gear micropitting test. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 226(12), pp.1051-1073.

Brandao, J.A., Martins, R., Seabra, J.H. and Castrol, M.J., 2015. An approach to the simulation of concurrent gear micropitting and mild wear. Wear, 324, pp.64-73.

Brown, M.W. and Miller, K.J., 1973. A theory for fatigue failure under multiaxial stress-strain conditions. *Proceedings of the Institution of Mechanical Engineers*, *187*(1), pp.745-755.

Bryant, M.J, 2013. *Running-in and residual stress: finite element contact analysis of as measured rough surfaces and comparison with experiment* (PhD dissertation, Cardiff University).

Bryant, M.J., Evans, H.P. and Snidle, R.W., 2012. Plastic deformation in rough surface line contacts—a finite element study. *Tribology International*, *46*(1), pp.269-278.

Bull, S.J., Evans, J.T., Shaw, B.A. and Hofmann, D.A., 1999. The effect of the white layer on micro-pitting and surface contact fatigue failure of nitrided gears. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 213(4), pp.305-313.

BS 3518-1, 1993.*Methods of fatigue testing. Guide to general principles.*

BS 3518-3, 1963. Methods of fatigue testing. Direct stress fatigue tests.

Carpinteri, A., De Freitas, M. and Spagnoli, A. eds., 2003. *Biaxial/multiaxial fatigue and fracture*, 31, Elsevier.

Chang, W.R., Etsion, I. and Bogy, D.B. 1987. An Elastic-Plastic Model for the Contact of Rough Surfaces. *Journal of Tribology* 109(2), p. 257.

Chu C. C., Conle, F. A And Bonnen, J. F. 1993. Multiaxial stress-strain modeling and fatigue life prediction of SAE axle shafts. *In: Advances in Multiaxial Fatigue, ASTM STP 1191 (Edited by D. L. McDowell and R. Ellis), American Society for Testing and Materials, Philadelphia, PA,* pp. 37-54.

Chu, C.C., 1995. Fatigue damage calculation using the critical plane approach. *Journal of Engineering Materials and Technology*, *117*(1), pp.41-49.

Chue, C.H. and Chung, H.H., 2000. Pitting formation under rolling contact. *Theoretical and applied fracture mechanics*, *34*(1), pp.1-9.

Ciavarella, M. and Maitournam, H., 2004. On the Ekberg, Kabo and Andersson calculation of the Dang Van high cycle fatigue limit for rolling contact fatigue. *Fatigue & Fracture of Engineering Materials & Structures*, *27*(6), pp.523-526.

Clarke, A., Evans, H.P. and Snidle, R.W., 2016. Mixed lubrication, friction and running-in of ground surfaces under gear contact conditions. Proceedings of the International Conference on Power Transmissions, pp. 975 -980.

Clarke, A., Weeks, I.J.J., Snidle, R.W. and Evans, H.P., 2016. Running-in and micropitting behaviour of steel surfaces under mixed lubrication conditions. *Tribology International*, *101*, pp.59-68.

Clarke, A., Weeks, I.J.J., Evans, H.P. and Snidle, R.W., 2016. An investigation into mixed lubrication conditions using electrical contact resistance techniques. *Tribology International*, *93*, pp.709-716.

Conry, T. F., Wang, S., and Cusano, C., 1987. A Reynolds-Eyring Equation for Elastohydrodynamic Lubrication in Line Contacts. Journal of Tribology. 109, 648.

Cook, R. D., Malkus, D. S., and Plesha, M. E., 1989. *Concepts and applications of finite element analysis*. New York, Wiley.

Crossland, B., 1956. Effect of large hydrostatic pressures on the torsional fatigue strength of an alloy steel. In Proceedings of the international conference on fatigue of metals (pp. 138-149). Lon-don: Institution of Mechanical Engineers.

Dong, Q., Yang, J., Wang, X., M Keer, L. and Zhou, K., 2016. Heterogeneous structures with inhomogeneous inclusions under elastohydrodynamic lubrication contact with consideration of surface roughness. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *230*(5), pp.571-582.

Dowling, N.E., 1998. *Mechanical Behaviour of Materials-Engineering Methods for Deformation,Fracture and Fatigue*, second edition. Prentice Hall, Englewood Cliffs, NJ.

Dowling, N.E., 2013. *Mechanical Behaviour of Materials-Engineering Methods for Deformation,Fracture and Fatigue, Fouth edition,Pearson education,Inc*.

Downing, S.D. and Socie, D.F., 1982. Simple rainflow counting algorithms. *International Journal of Fatigue*, *4*(1), pp.31-40.

Dowson, D. and Higginson, G.R., 1966, *Elastohydrodynamic Lubrication*, Pergamon, Oxford.

Dowson, D., 1998. *History of Tribology*. 2nd ed. Professional Engineering Publishing Limited, London, UK.

Dowson, D., and Higginson, G. R., 1966. *Elastohydrodynamic lubrication*. London, Pergamon Press.

Dowson, D., and Higginson, G. R., 1977. Elasto-hydrodynamic lubrication. Oxford [England], Pergamon Press.

Ekberg, A., Kabo, E., Lundén, R. and Matsui, M., 2016. Stress gradient effects in surface initiated rolling contact fatigue of rails and wheels. *Wear*, *366*, pp.188-193.

Elcoate, C.D., Evans, H.P., Hughes, T.G. and Snidle, R.W., 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*, *215*(4), pp.319-337.

Evans, H.P. and 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*, *214*(4), pp.563-584.

Evans, H.P. and Snidle, R.W., 1993. Wildhaber-Novikov circular arc gears: elastohydrodynamics. *Transactions-American Society Of Mechanical Engineers Journal Of Tribology*, 115(3), pp.487-487.

Evans, H.P. and Snidle, R.W., 2009. The future of engineering tribology in concentrated contacts. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *223*(12), pp.2939-2948.

Evans, H.P., Snidle, R.W., Sharif, K.J., Shaw, B.A. and Zhang, J., 2013. Analysis of micro-elastohydrodynamic lubrication and prediction of surface fatigue damage in micropitting tests on helical gears. *Journal of Tribology*, *135*(1), p.011501.

Fatemi A., Stephens R.I, Stephens R.R. and Fuchs H.O. 2000. *Metal Fatigue in Engineering*, Wiley Interscience.

Fatemi, A. and Socie, D.F., 1988. A Critical Plane Approach to Multiaxial Fatigue Damage Including out-of-Phase Loading. *Fatigue and Fracture of Engineering Materials and Structures*, *11*(3), pp.149-165.

Feng, D., Shen, M.X., Peng, X.D. and Meng, X.K., 2017. Surface Roughness Effect on the Friction and Wear Behaviour of Acrylonitrile–Butadiene Rubber (NBR) Under Oil Lubrication. *Tribology Letters*, *65*(1), p.10.

Findley, W.N., 1956. Fatigue of metals under combinations of stresses. Division of Engineering, Brown University.

Fouvry, S., Kapsa, P. and Vincent, L., 2001. An elastic–plastic shakedown analysis of fretting wear. *Wear*, *247*(1), pp.41-54.

Fujita, K. and Yoshida, A., 1981. Effect of hardness difference on the surface durability and surface failure of steel rollers. *Wear*, *67*(2), pp.187-200.

Gadelmawla, E.S., Koura, M.M., Maksoud, T.M.A., Elewa, I.M. and Soliman, H.H., 2002. Roughness parameters. *Journal of Materials Processing Technology*, *123*(1), pp.133-145.

Garud, Y.S., 1981. Multiaxial fatigue: a survey of the state of the art. *Journal of Testing and Evaluation*, *9*(3), pp.165-178.

Gelinck E.R.M., and Schipper D.J.,1999. Deformation of rough line contacts. *Journal of Tribology—Transactions of the ASME*, 121, pp. 449-54.

Gohar, R., and Rahnejat, H., 2008. *Fundamentals of tribology*. London, Imperial College Press.

Greenwood, J.A. and Tripp, J.H., 1967. The elastic contact of rough spheres. *ASME J. Appl. Mech*, *34*(1), pp.153-159.

Greenwood, J.A. and Williamson, J.B.P., 1966, December. Contact of nominally flat surfaces. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (Vol. 295, No. 1442, pp. 300-319).

Hamrock, B.J., Schmid, S.R. and Jacobson, B.O., 2004. *Fundamentals of fluid film lubrication*. CRC press.

Harris, T.A. and Yu, W.K., 1999. Lundberg-Palmgren fatigue theory: Considerations of failure stress and stressed volume. *Journal of Tribology*, *121*(1), pp.85-89.

Hassan, T. and Kyriakides, S., 1994. Ratcheting of cyclically hardening and softening materials: I. Uniaxial behavior. *International Journal of Plasticity*, *10*(2), pp.149-184.

Hertz, H., 1881. On the contact of elastic solids. *J. reine angew. Math*, *92*,*pp.* 156-171.

Hertzberg, R.W.1976. *Deformation and Fracture Mechanics of Engineering Materials*. John Wiley and Sons, New York.

Holmes, M.J.A., Evans, H.P. and Snidle, R.W., 2004, January. Analysis of mixed lubrication effects in simulated gear tooth contacts. In *ASME/STLE 2004 International Joint Tribology Conference*, pp. 447-457.

Holmes, M.J.A., Evans, H.P., Hughes, T.G. and Snidle, R.W., 2003a. Transient elastohydrodynamic point contact analysis using a new coupled differential deflection method Part 1: Theory and validation. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *217*(4), pp.289-304.

Holmes, M.J.A., Evans, H.P., Hughes, T.G. and Snidle, R.W., 2003b. 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*, *217*(4), pp.305-322.

Holmes, M.J.A., Qiao, H., Evans, H.P. and Snidle, R.W., 2005. Surface contact and damage in micro-EHL. *Tribology and interface engineering series*, *48*, pp.605-616.

Horng, J.H., Len, M.L. and Lee, J.S., 2002. The contact characteristics of rough surfaces in line contact during running-in process. *Wear*, *253*(9), pp.899-913.

Hu, Y.Z. and Zhu, D., 2000. A full numerical solution to the mixed lubrication in point contacts. *ASME J. Tribol*, *122*(1), pp.1-9.

Hughes, T.G., Elcoate, C.D. and 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*, *214*(4), pp.585-598.

Hutchings, I.M., and Shipway 1992. *Tribology: friction and wear of engineering materials*. *Butterworth-Heinemann.*

Hutt, S., 2018. *Understanding the Mechanism of Acoustic Emission Generation Due to Surface Asperity Interaction in Mixed Lubrication Conditions* (PhD dissertation, Cardiff University).

Jackson, R.L. and Green, I., 2005. A finite element study of elasto-plastic hemispherical contact against a rigid flat. *Transactions of the ASME-F-Journal of Tribology*, *127*(2), pp.343-354.

Jackson, R.L. and Streator, J.L., 2006. A multi-scale model for contact between rough surfaces. *Wear*, *261*(11), pp.1337-1347.

Jamari, J. and Schipper, D.J., 2006a. An elastic–plastic contact model of ellipsoid bodies. *Tribology Letters*, *21*(3), pp.262-271.

Jamari, J. and Schipper, D.J., 2007a. Deformation due to contact between a rough surface and a smooth ball. *Wear*, *262*(1), pp.138-145.

Jamari, J. and Schipper, D.J., 2007b. Plastic deformation and contact area of an elastic–plastic contact of ellipsoid bodies after unloading. *Tribology International*, *40*(8), pp.1311-1318.

Jamari, J. and Schipper, D.J., 2008. Deterministic repeated contact of rough surfaces. *Wear*, *264*(3), pp.349-358.

Jamari, J., De Rooij, M.B. and Schipper, D.J., 2007. Plastic deterministic contact of rough surfaces. *Journal of Tribology*, *129*(4), pp.957-962.

Jiang, X., Hua, D. Y., Cheng, H. S., Ai, X., & Lee, S. C. (1999). A Mixed Elastohydrodynamic Lubrication Model With Asperity Contact. *Journal of Tribology*. 121,pp. 481-491.

Johnson, K.L. 1995. Contact mechanics and the wear of metals. *Wear* 190, pp. 162–170.

Johnson, K.L. and Johnson, K.L., 1985. *Contact mechanics*. Cambridge university press.

Johnson, K.L. and Johnson, K.L., 1987. *Contact mechanics*. Cambridge university press.

Johnson, K.L., 1989. The strength of surfaces in rolling contact. *Proceedings of the Institution of Mechanical Engineers, Part C: Mechanical Engineering Science*, *203*(3), pp.151-163.

Kadin, Y., Kligerman, Y. and Etsion, I., 2006. Multiple loading–unloading of an elastic–plastic spherical contact. *International journal of solids and structures*, *43*(22), pp.7119-7127.

Keer, L. M., and Bryant, M. D.1983. A Pitting Model for Rolling Contact Fatigue. *Journal of Lubrication Technology,* 105(2), pp.198-205.

Kim, H.J., Ehret, P., Dowson, D. and Taylor, C.M., 2001. Thermal elastohydrodynamic analysis of circular contacts part 2: Non-Newtonian model. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *215*(4), pp.353-362.

Kogut, L. and Etsion, I., 2002. Elastic–plastic contact analysis of a sphere and a rigid flat. *ASME J. Appl. Mech*, *69*(5), pp.657-662.

Komvopoulos, K. and Bogy, D.B., 1993. Elastic-plastic finite element analysis of repeated indentation of a half-space by a rigid sphere. Journal of Applied Mechanics, 60, p.829.

Kong, S., 2001. *Contact, kinematics and film formation in worm gears*. (PhD dissertation, University of Wales).

Kragelsky, I. V., Alisin, V. V., Myshkin, N. K., And Petrokovets, M. I., 2001. *Tribology-lubrication, friction, and wear*. London, Professional Engineering Pub.

Kreyszig, E., 1999. *Advanced engineering mathematics*. 8th edition. John Wiley and Son, Inc.

Kweh, C. C., Patching, M. J., Evans, H. P., and Snidle, R. W., 1992. Simulation of Elastohydrodynamic Contacts Between Rough Surfaces. *Journal of Tribology.* 114(3), pp. 412-419.

Landgraf, R.W., 1970. The resistance of metals to cyclic deformation. In *Achievement of High Fatigue Resistance in Metals and Alloys*. ASTM International.

Lee, K.M. and Cheng, H.S., 1973, Effects of Surface Asperity on Elastohydrodynamic Lubrication, *NASA CR-2195, Washington, D.C.*

Li, S. and Kahraman, A., 2011. A fatigue model for contacts under mixed elastohydrodynamic lubrication condition. International Journal of Fatigue, 33(3), pp.427-436.

Li, S. and Kahraman, A., 2013. A physics-based model to predict micro-pitting lives of lubricated point contacts. International Journal of Fatigue, 47, pp.205-215.

Li, S.and Kahraman, A., 2013. Micro-pitting fatigue lives of lubricated point contacts: Experiments and model validation. International Journal of Fatigue, 48, pp.9-18.

Li, S. and Kahraman, A., 2014. A micro-pitting model for spur gear contacts. International Journal of Fatigue, 59, pp.224-233.

Li, S. and Wagner, J.J., 2016. An Approach for the Gear Rolling Contact Fatigue Acceleration. Journal of Mechanical Design, 138(3), p.034501.

Loewenthal, S.H., 1984. *Factors that affect the fatigue strength of power transmission shafting and their impact on design*. Lewis Research Center.

Lubrecht, A. A. and Ioannides, E., 1991. A fast solution of the dry contact problem and the associated sub-surface stress field, using multilevel techniques. *ASME J. Tribol, 113*(1), pp.128-133.

Lubrecht, A. A., ten Napel, W.E. and Bosma, R., 1986. Multigrid, an Alternative Method of Solution for Two-Dimensional Elastohydrodynamically Lubricated Point Contact Calculations. *Trans. ASME, Journal o f Tribology*, 108(3), pp.551 -556.

Masjedi, M. and Khonsari, M.M., 2015. On the effect of surface roughness in point-contact EHL: Formulas for film thickness and asperity load. *Tribology International, 82,* pp.228-244.

Matsuishi, M. and Endo, T., 1968. Fatigue of metals subjected to varying stress. Japan Society of Mechanical Engineers, Fukuoka, Japan, 68(2), pp.37-40.

Miller, G.R., Keer, L.M. and Cheng, H.S. 1985. On the Mechanics of Fatigue Crack Growth Due to Contact Loading. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 397(1813), pp. 197–209.

Morales, G.E., Brizmer, V. and Piras, E., 2015. Roughness evolution in mixed lubrication condition due to mild wear. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 229*(11), pp.1330-1346

Musallam, M. and Johnson, C.M., 2012. An efficient implementation of the rainflow counting algorithm for life consumption estimation. *IEEE Transactions on Reliability*, *61*(4), pp.978-986.

Nogueira, I., Dias, A.M., Gras, R. and Progri, R., 2002. An experimental model for mixed friction during running-in. *Wear*, *253*(5), pp.541-549.

Oila, A. and Bull, S.J., 2005. Assessment of the factors influencing micropitting in rolling/sliding contacts. *Wear*, *258*(10), pp.1510-1524.

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*, *219*(5), pp.313-330.

Omidvar, B. and Ghorbanpoor, A., 1998. The role of oil seepage in fatigue crack growth of lubricated wearing systems. *Engineering fracture mechanics*, *60*(2), pp.239-250.

Ostvik, R. and Christensen, H., 1968, September. Changes in surface topography with running-in. In *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, 183, (16), pp. 57-65.

Papadopoulos, I.V., Davoli, P., Gorla, C., Filippini, M. and Bernasconi, A., 1997. A comparative study of multiaxial high-cycle fatigue criteria for metals. *International Journal of Fatigue*, 19(3), pp.219-235.

Papuga, J., Vargas, M. and Hronek, M., 2012. Evaluation of uniaxial fatigue criteria applied to multiaxially loaded unnotched samples. *Engineering Mechanics*, *19*(2), pp. 99-111.

Patching, M. J. 1994. *The effect of surface roughness on the micro-elastohydrodynamic lubrication and scuffing performance of aerospace gear tooth contacts* (PhD dissertation, Cardiff University).

Paulson, N.R., Sadeghi, F. and Habchi, W., 2017. A coupled finite element EHL and continuum damage mechanics model for rolling contact fatigue. *Tribology International*, *107*, pp.173-183.

Peng, H., Liu, Z., Huang, F. and Ma, R., 2013. A study of elastic–plastic contact of statistical rough surfaces. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *227*(10), pp.1076-1089.

Pletcher, R.H., Tannehill, J.C. and Anderson, D., 2012. *Computational fluid mechanics and heat transfer*. CRC Press.

Pu, W., Zhu, D., Wang, J. and Wang, Q.J., 2016. Rolling–sliding contact fatigue of surfaces with sinusoidal roughness. *International Journal of Fatigue*, *90*, pp.57-68.

Pullen, J. and Williamson, J.B.P., 1972, March. On the plastic contact of rough surfaces. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 327, (1569), pp. 159-173.

Qiao, H., Evans, H. P., and Snidle, R. W., 2008. Comparison of fatigue model results for rough surface elastohydrodynamic lubrication. Proceedings of the Institution of Mechanical Engineers, Part J: *Journal of Engineering Tribology*. 222, 381-393.

Raja, J., Muralikrishnan, B. and Fu, S., 2002. Recent advances in separation of roughness, waviness and form. *Precision Engineering*, *26*(2), pp.222-235.

Rico, J.F., Battez, A.H. and Cuervo, D.G., 2003. Rolling contact fatigue in lubricated contacts. *Tribology International*, *36* (1), pp.35-40.

Robinson, J.W., Zhou, Y., Bhattacharya, P., Erck, R., Qu, J., Bays, J.T. and Cosimbescu, L., 2016. Probing the molecular design of hyper-branched aryl polyesters towards lubricant applications. *Scientific reports*, *6*.

Rycerz, P., Olver, A. and Kadiric, A., 2017. Propagation of surface initiated rolling contact fatigue cracks in bearing steel. *International Journal of Fatigue*, *97*, pp.29-38.

Shankar, S. and Mayuram, M.M., 2008. Effect of strain hardening in elastic–plastic transition behavior in a hemisphere in contact with a rigid flat. *International Journal of Solids and Structures*, *45*(10), pp.3009-3020.

Sharif, K.J., Evans, H.P. and Snidle, R.W., 2012. Modelling of elastohydrodynamic lubrication and fatigue of rough surfaces: The effect of lambda ratio. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *226*(12), pp.1039-1050.

Sherrington, I., University of Central Lancashire, 2009. *Lubrication control system*. U.S. Patent 7,552,799.

Sherrington, I. and Hayhurst, P., 2001. Simultaneous observation of the evolution of debris density and friction coefficient in dry sliding steel contacts. *Wear*, *249*(3), pp.182-187.

Sines, G. and Ohgi, G., 1981. Fatigue criteria under combined stresses or strains. *J. Engng. Mater. Technol., Trans. ASME,* 103(2), pp. 82-90.

Sines, G., 1959. Behavior of metals under complex static and alternating stresses. Metal fatigue, 1, pp.145-169.

Snidle, R.W. and Evans, H.P., 1994. A simple method of elastic contact simulation. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *208*(4), pp.291-293.

Stachowiak, G. and Batchelor, A.W., 2005. *Engineering tribology*. Amsterdam, Elsevier Butterworth-Heinemann.

Stachowiak, G. and Batchelor, A.W., 2013. *Engineering tribology*. Butterworth-Heinemann.

Stolarski, T.A., Tobe, S. and Kalousek, J., 2002. Rolling Contacts. Tribology in Practice Series. *Applied Mechanics Reviews*, *55*, p.B53.

Suresh, S., 1998. *Fatigue of materials*. Cambridge university press.

Tabor, D. 1951. *The Hardness of Metals*. Oxford: Oxford University Press Inc.

Tao, J., Hughes, T.G., Evans, H.P., Snidle, R.W., Hopkinson, N.A., Talks, M. and Starbuck, J.M. 2003.Elastohydrodynamic Lubrication Analysis of Gear Tooth Surfaces From Micropitting Tests. *Journal of Tribology* 125(2), p. 267.

Teer, D.G. and Arnell, R.D. 1975a. Friction Theories. In: Halling, J. ed. *Principles of Tribology*. London and Basingstoke: The Macmillan Press Ltd., pp. 72–93.

Terrin, A., Dengo, C. and Meneghetti, G., 2017. Experimental analysis of contact fatigue damage in case hardened gears for off-highway axles. *Engineering Failure Analysis*, *76*, pp.10-26.

Timoshenko, S., and Goodier, J.N. 1973. Theory of Elasticity, *3rd Edn. New York, London et al.:* McGraw-Hill.

Van, K.D., Cailletaud, G., Flavenot, J.F., Le Douaron, A. and Lieurade, H.P., 2013, October. Criterion for high-cycle fatigue failure under multiaxial loading. In *ICBMFF2*.

Venner, C.H., 1991. *Multilevel solution of the EHL line and point contact problems*. (Ph.D. dissertation, University of Twente).

Verdu, C., Adrien, J. and Buffière, J.Y., 2008. Three-dimensional shape of the early stages of fatigue cracks nucleated in nodular cast iron. *Materials Science and Engineering: A*, *483*, pp.402-405.

Vincent, A., Lormand, G., Lamagnere, P., Gosset, L., Girodin, D., Dudragne, G. and Fougeres, R., 1998. From white etching areas formed around inclusions to crack nucleation in bearing steels under rolling contact fatigue. In *Bearing Steels: Into the 21st Century*. ASTM International.

Walvekar, A.A., Paulson, N., Sadeghi, F., Weinzapfel, N., Correns, M. and Dinkel, M., 2017. A New Approach for Fatigue Damage Modeling of Subsurface-Initiated Spalling in Large Rolling Contacts. *Journal of Tribology*, *139*(1), p.011101.

Wang, W., Wong, P.L. and Zhang, Z., 2000. Experimental study of the real time change in surface roughness during running-in for PEHL contacts. *Wear*, *244*(1), pp.140-146.

Webster, M.N. and Sayles, R.S., 1986. A numerical model for the elastic frictionless contact of real rough surfaces. *Journal of Tribology*, *108*(3), pp.314-320.

Weeks, I., 2015. *An experimental investigation into the mixed lubrication of steel surfaces* (PhD dissertation, Cardiff University).

Williams, J.A., Dyson, I.N. and Kapoor, A., 1999. Repeated loading, residual stresses, shakedown, and tribology. *Journal of materials research*, *14*(04), pp.1548-1559.

Williams, J.E., 1994. *Engineering Tribology*. Oxford: Oxford University Press Inc.

Yastrebov, V.A., Anciaux, G. and Molinari, J.F., 2015. From infinitesimal to full contact between rough surfaces: evolution of the contact area. *International Journal of Solids and Structures*, *52*, pp.83-102.

Yastrebov, V.A., Durand, J., Proudhon, H. and Cailletaud, G., 2011. Rough surface contact analysis by means of the finite element method and of a new reduced model. *Comptes Rendus Mécanique*, *339*(7-8), pp.473-490.

Yoshida, A., and Konishi, D., 1995. Study on Influences of Machining Method and Surface Roughness on Pitting of Thermally Refined Steel Rollers (Part 4): In the Case of Rollers with Axial Tool Marks Mating with Cylindrically Ground Rollers. *Japanese Journal of Tribology*. 40, 425-436.

Yoshida, A., Ohue, Y. and Karasuno, I., 1994. Surface failure and durability of induction-hardened sintered powder metal rollers and gears with various hardened depths. *Journal of Mechanical Design*, *116*(3), pp.730-737.

You, B.R. and Lee, S.B., 1996. A critical review on multiaxial fatigue assessments of metals. *International Journal of Fatigue*, *18*(4), pp.235-244.

Zahavi, E., Torbilo, V. and Press, S., 1996. *Fatigue design: life expectancy of machine parts*. CRC press.

Zhou, R.S., Cheng, H.S. and Mura, T., 1989. Micropitting in rolling and sliding contact under mixed lubrication. *J. Tribology(Trans. ASME)*, *111*(4), pp.605-613.

Zhu, D. and Ai, X., 1997. Point contact EHL based on optically measured three-dimensional rough surfaces. *Transactions-American Society Of Mechanical Engineers Journal Of Tribology*, *119*, pp.375-384.

Zhu, D. and Hu, Y.Z., 2001a. Effects of rough surface topography and orientation on the characteristics of EHD and mixed lubrication in both circular and elliptical contacts. *Tribology transactions*, *44*(3), pp.391-398.

Zhu, D. and Hu, Y.Z., 2001b. 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*, *44*(3), pp.383-390.

Zhu, H., Ge, S., Cao, X. and Tang, W., 2007. The changes of fractal dimensions of frictional signals in the running-in wear process. *Wear*, *263*(7), pp.1502-1507.