Prediction of Contact Fatigue for the Rough Surface Elastohydrodynamic Lubrication Line Contact Problem under Rolling and Sliding Conditions

by

Hua Qiao

Thesis submitted to Cardiff University
for the degree of
Doctor of Philosophy

School of Engineering
Cardiff University

October 2005

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

Abstract

The study of elastohydrodynamic lubrication (EHL) with rough surfaces results in a transient problem due to time varying geometry. In conditions where the Λ ratio (the ratio of smooth surface film thickness to composite roughness) is low, significant pressure deviations from the corresponding smooth surface results occur. The relative motion of asperity features on the two surfaces leads to cycling of these pressure deviations as surface features move through the nominal contact area. Micropitting, which is a form of surface distress seen in gears, may well be viewed as rolling contact fatigue at the roughness asperity scale as a result of this load cycling.

In order to understand the failure mechanism associated with rough surface EHL, a full theoretical model of lubrication of gear contacts under low Λ conditions is presented in this thesis. The model takes account of the real gear operating conditions in terms of loads, speeds, surface roughness and lubricant properties to predict pressures, film thickness, temperatures, and friction between the teeth. Subsequent contact stress analysis was performed to determine the stress history developed in the contacting solids.

Conventional pitting is usually associated with failure beneath the surface of a rolling contact and in the past has been linked with the occurrence of the maximum shear stress that occurs in the classical Hertzian solution for smooth surfaces. In this thesis, a numerical procedure for predicting contact fatigue damage in an EHL line contact between two rough surfaces is developed to take the features of variable amplitude multiaxial fatigue into account due to the roughness effect. The way in which the calculated fatigue damage varies with changes in factors such as slide roll ratio, speed, viscosity, rough surface profiles, and Λ ratio adopted for the analysis is demonstrated in a series of numerical examples using roughness profiles taken from gears in micropitting experiments.

Acknowledgements

I gratefully acknowledge the receipt of an ORS award and a Cardiff School of Engineering scholarship that have helped to make this study possible.

The work presented in this thesis could not have been finished without the great help of my supervisors, Prof. H.P. Evans and Prof. R.W. Snidle. They have always been there when I needed most. I sincerely thank them for their encouragement, thoughtful advice and guidance and invaluable support during my study.

I would like to thank Dr K.J. Sharif for his great help, especially for helping us to settle down in Cardiff when we first came here. Thanks are also due to him for establishing the thermal solution method and never-ending advice and discussions. I would also like to express my thanks to Dr M.J.A. Holmes for providing the line contact solver, and to Mr C. Davies for providing his dry contact solutions and thermal solutions.

The three years at Cardiff will never be forgotten in my life because of nice people, beautiful environment, mild weather and because my son was born here. I would like to thank all my friends and colleagues who have been with me during this period.

Finally I would like to thank my wife for helping me through every effort. Thanks to her for spending all her time to look after our baby so that I could concentrate on my study.

Contents

Chapter 1

Introduction

1.1 Background

Elastohydrodynamic lubrication (EHL) is the lubrication mechanism that protects highly loaded concentrated contacts between machine elements from surface damage and wear. Due to the very limited contact area and the high pressure involved, elastic deformation of the surfaces will occur and the viscosity of the lubricant can be much higher than its value at atmospheric pressure. Both these effects play a vital role in lubricant film formation.

Typical examples of highly loaded EHL contacts occur in rolling element bearings, cams and tappets and gear tooth contacts. The latter case presents particular difficulties as the surfaces produced by present day gear manufacturing methods have roughness features that are significantly greater than the predicted oil film. Consequently gears tend to operate in a regime described as "mixed" or "micro" EHL in which there is significant interaction of roughness asperities on the two surfaces. Almost all failures of surfaces in rolling/sliding contact with an EHL lubricant film occur in this regime. Therefore, surface roughness plays an important role in determining gear lubrication and the occurrence of surface failures. The work reported in this thesis addresses the fundamental problem of EHL of rough surfaces.

This chapter introduces some fundamentals related to EHL such as Hertzian theory, lubricant properties, rough surfaces and modes of surface failure. The research objective of this work is then presented. The organisation of the thesis is outlined at the end of this chapter.

1.2 Hertzian Contact

The Hertz theory deals with determining the size of the elastic contact zone for given geometry, elastic properties and load, and the pressure distribution developed within that contact zone. This theory assumes that contact between two surfaces is nonconforming and frictionless, involving small strains, and that each body can be considered as a semi-infinite elastic solid (Johnson 1985).

When two convex bodies are pressed together, contact is initially made at a point or a line. As the load increases, the contact patch will grow, forming a circular, elliptical or rectangular shape. The contact between a pair of spur gear involute teeth is a line contact problem which is shown schematically in Figure 1.1, where two equivalent cylinders are brought into contact. If the total load per unit length is w', the Hertzian semi-contact width for a line contact is given by (Williams 1994)

$$a = \sqrt{\frac{8w'R'}{\pi E'}} \tag{1.1}$$

where R' is the radius of the cylinder on an equivalent cylinder/plane contact, and E' is the reduced Young's modulus and effectively makes one of the surfaces rigid. If the solids have radii R_1 and R_2, respectively, Young's moduli E_1 and E_2 with Poisson's ratios ν_1 and ν_2, R' and E' can be obtained from

$$\frac{1}{R'} = \frac{1}{R_1} + \frac{1}{R_2} \tag{1.2}$$

$$\frac{1}{E'} = \frac{1}{2}\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right) \tag{1.3}$$

The contact pressure distribution within the contact area is semi-elliptical:

$$p(x) = p_0\sqrt{1-\left(\frac{x}{a}\right)^2} \tag{1.4}$$

The pressure rises from zero at the edges of the contact ($x = \pm a$) to a maximum pressure p_0 at the centre. The peak Hertz pressure is

$$p_0 = \frac{2w'}{\pi a} = \sqrt{\frac{w'E'}{2\pi R'}} \tag{1.5}$$

and the mean pressure is

$$p_m = \frac{\pi}{4} p_0 \tag{1.6}$$

The corresponding stresses for a point (x, z) within the material under this surface loading are given by (Johnson 1985)

$$\sigma_x = -\frac{p_0}{a}\left[m\left(1+\frac{z^2+n^2}{m^2+n^2}\right) - 2z \right] \tag{1.7a}$$

$$\sigma_x = -\frac{p_0}{a} m\left(1-\frac{z^2+n^2}{m^2+n^2}\right) \tag{1.7b}$$

$$\tau_{xz} = -\frac{p_0}{a} n\left(\frac{m^2-z^2}{m^2+n^2}\right) \tag{1.7c}$$

where m and n have the same signs ('+' or '–') as z and x respectively, and are defined by

$$m^2 = \frac{1}{2}\left[\sqrt{\left(a^2-x^2+z^2\right)^2+4x^2z^2} + \left(a^2-x^2+z^2\right) \right] \tag{1.8a}$$

$$n^2 = \frac{1}{2}\left[\sqrt{\left(a^2-x^2+z^2\right)^2+4x^2z^2} - \left(a^2-x^2+z^2\right) \right] \tag{1.8b}$$

The stress state beneath the contact is compressive in character. The maximum principal shear stress $\tau_{1,\max} = 0.3p_0$ occurs beneath the contact interface at $x = 0$ and $z = 0.78a$.

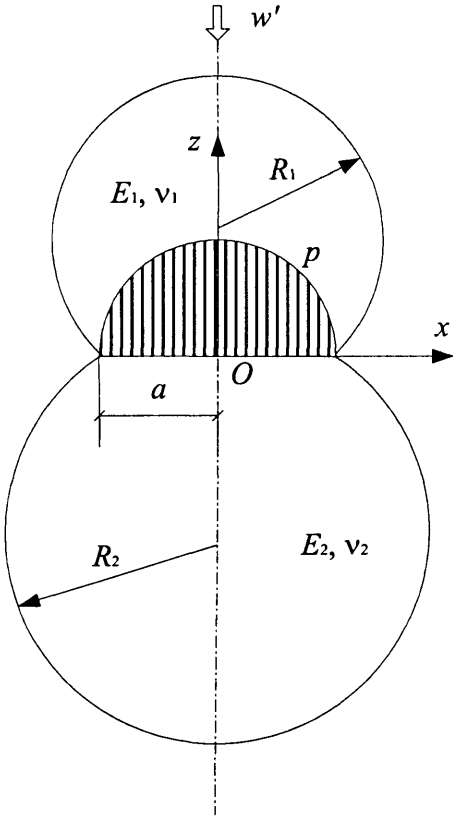


Figure 1.1: Hertz contact between two cylinders.

1.3 Lubricant Properties

In EHL, the properties of the lubricant play a significant role in forming the lubricating film between the contacting surfaces. The most important lubricant property is its viscosity (Williams 1994). The viscosity η is a measure of resistance to relative shearing motion and is defined by Newton's equation:

$$\eta = \frac{\tau}{\dot{\gamma}} \tag{1.9}$$

where τ is the shear stress acting on an element of the lubricant in the direction of flow. $\dot{\gamma}$ is the shear strain rate (also called shear rate), which is equivalent to the velocity gradient normal to the shear stress of the element (Gohar 1988), i.e.

$$\dot{\gamma} = \frac{\partial u}{\partial z} \tag{1.10}$$

1.3.1 Temperature and Pressure Effects

Both temperature and pressure have influences on the viscosity of the lubricant. With increasing temperature, the viscosity falls rapidly. This viscosity-temperature relationship can be approximated by a simple equation suggested by Reynolds (Gohar 1988):

$$\eta = \eta_0 \exp[-\beta(T - T_0)] \tag{1.11}$$

where η_0 is the viscosity at reference temperature T_0, η is the viscosity at temperature T, and β is a constant. With increasing pressure, the viscosity increases rapidly and this viscosity-pressure relationship can be approximated by the Barus equation (Gohar 1988):

$$\eta = \eta_0 \exp(\alpha p) \tag{1.12}$$

where η_0 is the viscosity at reference (atmospheric) pressure, α is the pressure-viscosity coefficient.

Equations (1.11) and (1.12) are only accurate for a limited temperature range and a relatively low pressure. Because of these limitations, several other equations have been proposed. One of the most used in EHL is the Roelands equation (Stachowiak & Batchelor 1993), which takes into account the effects of both temperature and pressure simultaneously:

$$\eta_R = \eta_0 \exp(\alpha^* p) \tag{1.13}$$

The Roelands pressure-viscosity coefficient, α^*, is a function of both T and p according to

$$\alpha^* p = \ln(\eta_0/\kappa)\left[\left(\frac{T-138}{T_0-138}\right)^{-S_0}(1+\chi p)^Z - 1\right] \tag{1.14}$$

where κ and χ are constants, Z and S_0 are constant properties for a specific oil, independent of temperature and pressure. The constant Z can be obtained from α by comparing the Roelands equation and Barus equation as p tends to zero as follows (Hoglund 1999):

$$Z = \frac{\alpha}{\chi \ln(\eta_0/\kappa)} \tag{1.15}$$

Temperature and pressure effects on density can also be quite significant due to the compressibility or the thermal expansion of the lubricant. Based on the density-pressure relationship used by Dowson and Higginson (1966), Sui and Sadeghi (1991) gave the following density-pressure-temperature relationship:

$$\rho = \rho_0 \frac{1+\gamma p}{1+\lambda p}\left[1-\varepsilon(T-T_0)\right] \tag{1.16}$$

where γ and λ are constants, ρ_0 is the reference density (at atmospheric pressure) and ε is the coefficient of lubricant thermal expansivity.

1.3.2 Non-Newtonian Behaviour

The relationship between the shear rate and shear stress can be described using a simple function:

$$\dot{\gamma} = f(\tau) \tag{1.17}$$

If this function is linear, as described by Equation (1.9), then the lubricant is said to be Newtonian. In practice as the shear rate increases the relationship between shear stress and shear strain rate is no longer linear and non-Newtonian lubricant effects come into play. Many non-Newtonian models have been proposed in the literature and typical examples of models used in EHL analysis are given in Table 1.1 (Mihailidis et al. 1999, Hoglund 1999), where the characteristic or limiting shear stress, τ_0 or τ_1, as well as the exponent n are derived

from experiments. Some of these models use the concept of a limiting shear stress, which cannot be exceeded; that is, at limiting shear stress the strain rate becomes arbitrarily large. In the presence of a pressure gradient this proposed property imposes a maximum possible speed variation across the lubricant film and it is hypothesised that if this sliding speed demand is exceeded gross slip between the lubricant and the surface will occur. More details concerning lubricant behaviour and the models developed in elastohydrodynamic contacts have been discussed by Moore (1997).

Table 1.1: Functions $f(\tau)$ describing the relationship between the shear rate and shear stress for various non-Newtonian models.

Ree-Eyring (sinh model)	$\frac{\tau_0}{\eta}\sinh\left(\frac{\tau}{\tau_0}\right)$
Bair-Winer (ln model)	$\frac{\tau_1}{\eta}\ln\left(1-\frac{\tau}{\tau_1}\right)^{-1}$
Gecim-Winer ($\tanh^{-1}$ model)	$\frac{\tau_1}{\eta\tanh(\tau/\tau_1)}$
Livonen-Hamrock (straight-line model)	$\frac{\tau_1}{\eta}\left[(1-\tau/\tau_1)^{-1}-1\right]$
Elsharkawy-Hamrock (general model)	$\frac{\tau}{\eta\sqrt[n]{1-(\tau/\tau_1)^n}}$
Lee-Hamrock (circular model)	$\frac{\tau}{\eta\sqrt{1-(\tau/\tau_1)^2}}$
Power Law	$\frac{\tau}{\eta}\left\|\frac{\partial u}{\partial z}\right\|^{1-n}$

1.4 Rough Surface EHL

Real engineering surfaces formed by machining or equivalent processes are always flawed with peaks and valleys. Statistical parameters are widely used to describe the surface

roughness. Such parameters include height parameters (describing the variation in height) and spatial parameters (describing how height varies in the surface plane). Some of these parameters are defined in Table 1.2, where ϕ is the height of the surface measured above the mean level and L is the length of measurement. More details about rough surfaces and descriptors can be found in references (Moore 1975, Dagnall 1986, Stachowiak & Batchelor 1993, Williams 1994).

Table 1.2: Commonly used parameters describing surface roughness.

Center line average (CLA) roughness, R_a	$R_a = \frac{1}{L}\int_0^L \lvert\phi(x)\rvert dx$
Root mean square (RMS) roughness, R_q	$R_q = \sqrt{\frac{1}{L}\int_0^L \phi^2(x)dx}$
Autocovariance function (ACVF), $R(\delta)$	$R(\delta) = \lim_{L\to\infty}\frac{1}{L}\int_0^L \phi(x)\phi(x+\delta)dx$
Autocorrelation function (ACF), P(δ),	$\mathrm{P}(\delta) = \frac{R(\delta)}{R_q^2}$

Due to the influence of surface roughness, there are three regimes associated with rough surface EHL (Olver 2002):

1. Micro-EHL.
2. Mixed lubrication.
3. Boundary lubrication.

In Regime 1, there is a continuous fluid film, but the pressure and film thickness are subject to local fluctuations due to the surface roughness. In Regime 2, the fluid film is discontinuous. Part of the load is carried by pressure in the hydrodynamic film, and part of the load by contact pressure between parts of the surfaces that make direct contact without the benefit of a hydrodynamic film. In Regime 3, there is some lubricant, which is confined to the valleys and carries a negligible proportion of the total load. Essentially full solid-solid contact is obtained in the equivalent Hertzian contact area. A schematic representation of

these regimes is shown in Figure 1.2 (Olver 2002). This thesis will focus on micro-EHL and mixed lubrication.

A useful measurement of the importance of surface roughness in EHL is specific film thickness Λ, which is the ratio of the minimum film thickness for smooth surfaces to the composite surface roughness (Gohar 1988):

$$\Lambda = \frac{h_{\min}}{\sqrt{R_{q1}^2 + R_{q2}^2}} \tag{1.18}$$

where R_{q1} and R_{q2} are RMS roughness values for two surfaces. At lower values of Λ (i.e. $\Lambda < 1$), it is conventionally supposed that asperity interaction becomes more severe (Williams 1994). The minimum film thickness may be calculated from the dimensionless load parameter W, speed parameter U, and material parameter G, with reasonable accuracy if the contact surfaces are smooth, using the well-known formula of Dowson and Higginson (1966):

$$\begin{aligned} \frac{h_{\min}}{R'} &= 1.6 W^{-0.13} G^{0.6} U^{0.7} \\ &= 1.6 \left(\frac{w'}{E'R'} \right)^{-0.13} (\alpha E')^{0.6} \left(\frac{\eta_0 u}{E'R'} \right)^{0.7} \end{aligned} \tag{1.19}$$

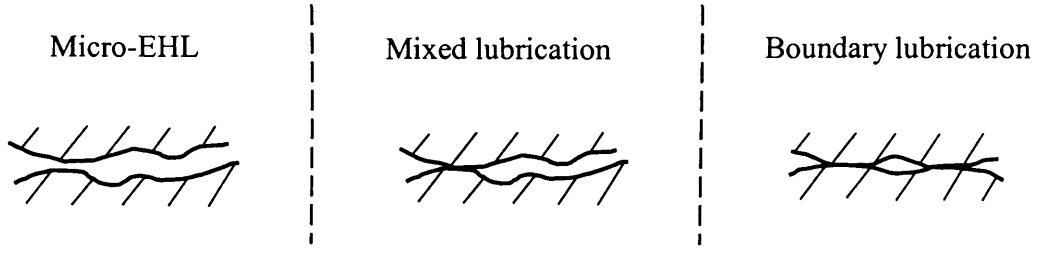


Figure 1.2: Schematic representation of different regimes of lubrication of rough surfaces.

1.5 Surface Fatigue Failures

The most important modes of failure of surfaces in rolling contact, as reviewed by Johnson (1989), include plastic deformation, contact fatigue, wear, scuffing and corrugation. The complex nature of rolling contact makes it difficult to separate individual failure modes. This work focuses on surface failure of gears caused by rolling contact fatigue operating in the micro-EHL or mixed lubrication regime.

Contact fatigue is the surface damage process which occurs when the surfaces of two bodies repeatedly contact each other. Although the fatigue process is influenced by a number of factors such as the material and the operating conditions, this mode of failure leads to crack formation at or near the contact surface. Major rolling contact fatigue failures include:

(a) *Pitting.* Pitting is the formation of microscopic surface notches and cavities caused by rolling and sliding contact fatigue. If full EHL film conditions are achieved and friction at the surface is comparably low, pitting may develop below the surface, primarily in the zone of highest subsurface stress due to the Hertzian contact pressure.

(b) *Micropitting.* Micropitting can be regarded as fatigue failure with a net of cracks close to the surface. Micropitting is more commonly observed in hardened gears under mixed lubrication conditions.

(c) *Spalling.* Spalling is macro-scale Hertzian contact fatigue leading to the formation of macroscopic craters in the contact region. Spalling is similar to subsurface pitting, except that the spalled areas are larger.

(d) *Surface distress.* Surface distress is micro-scale spalling fatigue of rolling contact surfaces due mainly to asperity-level normal contact loads.

Some prominent discussions of fatigue failure with regard to mechanism, modes and predictive methods can be found in Colangelo & Heiser (1987), Johnson (1989), Fernandes & McDuling (1997) and Suresh (1998).

1.6 Research Objective

Surface contact fatigue is the most common cause of gear failure (Fernandes & McDuling 1997). Fatigue life can be determined by experiments. For example, Gao at al. (2005) conducted fatigue tests on a commercially available contact fatigue test rig to investigate the variation of fatigue lives with Λ ratios. However, performing such tests is time consuming, expensive and requires a large number of test samples to produce meaningful results. Also the results are affected by the test rig and test conditions and may be difficult to correlate to actual service gears.

To understand the failure mechanism a full theoretical model of lubrication of gear contacts under low Λ conditions is needed. The model must take account of the real gear operating conditions in terms of loads, speeds, surface roughness and lubricant properties, and be able to predict pressures, local film thickness, temperatures, friction between the teeth and subsequent stress and fatigue damage. The objective of the work reported in this thesis is to develop such a theoretical model to assist design against surface contact fatigue, such as micropitting.

1.7 Thesis Organisation

The basic elements of the fatigue prediction procedure presented in this thesis are illustrated in Figure 1.3. Each of the chapters in this thesis addresses a specific element of the procedure as shown in this figure.

Chapters 2 and 3 address EHL analysis taking the effects of temperature, non-Newtonian behaviour and surface roughness into consideration. Chapter 2 presents numerical formulations for isothermal problems, and Chapter 3 focuses on thermal effects in rough surface EHL contacts.

Chapter 4 deals with stress analysis. Distributions and histories of surface and subsurface stress are determined numerically.

Chapters 5-7 focus on fatigue analysis. Chapter 5 introduces some theories of fatigue and reviews models developed for rolling contact fatigue. Chapters 6 and 7 deal with the numerical evaluation of fatigue performance using critical plane approaches by examining different affecting factors. Chapter 6 applies multiaxial fatigue criteria, whereas Chapter 7 uses a shear strain based model incorporating cumulative damage analysis.

Chapter 8 draws conclusions based on the previous chapters. Suggestions for future work are also presented in this chapter.

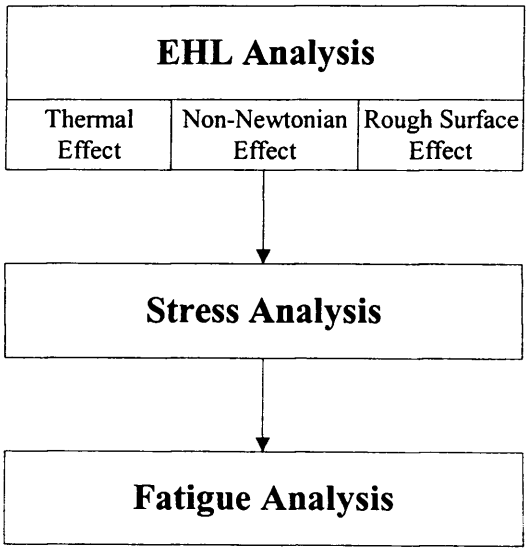


Figure 1.3: Basic elements of the fatigue prediction procedure for EHL problems.

Chapter 2

Rough Surface EHL in Line Contact

2.1 Introduction

The main features of EHL for smooth surfaces are well known from the pioneering work of Dowson and Higginson (1966). However, most engineering surfaces, particular those of the teeth of gears, are rough compared to the film thickness. Consequently they operate under conditions in which asperity-asperity interactions significantly affect both the deformation and the pressures within the nominal contact region. This regime of lubrication is described as micro-EHL.

The theoretical study of rough surface EHL has attracted many researchers over the past twenty years or so. Initial efforts involved stationary idealized roughness features, such as single irregularities (Goglia et al. 1984a) and sinusoidal surfaces (Goglia et al. 1984b, Lubrecht et al. 1988, Kweh et al. 1989), and real roughness features (Kweh et al. 1992). These analyses are limited to simple sliding conditions and the interactions of asperities on the opposing surfaces cannot be simulated. To study the asperity interactions, transient ("moving roughness") effects have to be considered to examine the evolution and perturbation of pressure and film thickness within the EHL contact. Chang and Webster (1991) simulated the interactions of two rough surfaces in line contact and roughness features are modeled by sinusoidal functions. Venner & Lubrecht (1994) investigated indentation and waviness in transient line contact and results showed the effect of slide roll ratio on the pressure and film thickness profile. Chang and Zhao (1995) studied the important differences between Newtonian and non-Newtonian lubricant treatments in transient micro-EHL and results suggested that the differences would be most pronounced with surface roughness of

short wavelengths. Ai and Cheng (1994) used measured surface roughness and a Newtonian fluid model in a line contact micro-EHL analysis. The micro-EHL point contact with measured three-dimensional surface roughness was investigated by Xu and Sadeghi (1996) and Zhu and Ai (1997). Where real roughness data were incorporated in transient analyses, however, they are generally of relatively low amplitude compared with the minimum film thickness in the numerical solution because of the numerical difficulties.

Hughes et al. (2000) and Elcoate et al. (2001) made further contributions to the study of micro-EHL by developing a new fast solution method suitable for severe rough-surface thin-film conditions. The essential feature of the method is the full mathematical coupling of the elastic and hydrodynamic equations in the numerical solution scheme. Film thickness and pressure are treated as simultaneously active variables throughout the solution process. This leads to a highly robust and rapidly convergent numerical scheme. Using this technique, Elcoate et al. (2001) examined the behaviour of real gear lubrication conditions in which surface roughness is at least an order of magnitude greater than the minimum film thickness. Later the method was extended to the geometry of point contacts by Holmes et al. (2003a).

The study of micro-EHL leads naturally to mixed lubrication where part of the load is carried by pressurised lubricant that separates the surfaces completely, and part of the load is carried by dry or boundary lubricated contact between asperity features. Mixed lubrication is a particularly challenging and important regime to understand as the micro-EHL action, asperity-contact mechanics and lubricant-surface tribo-chemistry all play significant roles (Chang 1995a). In recent years analyses have attempted to address this situation and different numerical approaches have been developed. As described by Wang et al. (2004a), three approaches have been used in mixed lubrication studies: (i) deterministic modeling, (ii) stochastic modeling, and (iii) interactive deterministic-stochastic modeling. Stochastic models use statistic parameters to deal with the influence of roughness on lubrication. Deterministic models directly examine realistic localised asperity contacts. The interactive deterministic-stochastic modeling, or macro-micro approach, combines the macro-scale average lubrication analysis with the micro-scale deterministic asperity contact simulation.

Deterministic models are fundamental to mixed lubrication and provide a solid base for the development, refinement, and verification for other models (Wang et al. 2004a). There have been a number of reports on deterministic modeling of mixed lubrication. Chang (1995b) presented a deterministic model for line-contact mixed lubrication problems. The governing equations of the model are composed of the asperity contact equation and the equations that govern basic transient micro-EHL problems. Similarly, Jiang et al. (1999) and Zhao et al. (2001) solved point-contact mixed lubrication problems by partitioning the contact area into sub areas where lubricant is and is not present. Hu & Zhu (2000) developed a unified approach for a full coverage of all regions including full film, mixed lubrication and boundary lubrication. They reduced the Reynolds equation for conditions of very low film thickness (asperity contact areas). Wang et al. (2004b) extended this model to take thermal effects into account. However, some of the results (Zhu 2002a, 2002b) produced by this approach have been shown to be questionable (Glovnea et al. 2003, Holmes et al. 2003b). More recently, Holmes et al. (2005a, 2005b) have developed a unified approach for mixed lubrication without any simplifying assumptions. In their method, whenever negative film thicknesses occurred at individual mesh points during a time step calculation, the Reynolds equation for that mesh point was deleted from the overall solution scheme and replaced with the boundary condition of zero film thickness at that point. The values of pressure and film thickness are still required to satisfy the elastic deflection equation subject to that boundary condition.

This chapter presents some numerical formulations for EHL line contact with rough surfaces. The coupled numerical solution method for determining pressure and film thickness in micro-EHL and mixed lubrication conditions is briefly explained. This solution method will be used as the EHL analysis tool in this thesis. In this chapter, more efforts are paid to the representation of measured rough profiles using a Fourier analysis procedure. The numerical results are presented to examine the performance of this procedure, which is incorporated into the line contact EHL solver.

2.2 Governing Equations

The three following effects play a major role in the formation of lubrication films in EHL (Stachowiak & Batchelor 1993):

- The hydrodynamic film formation,
- The modification of the film geometry by elastic deformation,
- The transformation of the lubricant's viscosity and rheology under pressure.

These effects act simultaneously and mathematical modeling of all these effects with some assumptions leads to the derivation of the governing equations of EHL.

2.2.1 Reynolds Equation

The fundamental equations of fluid dynamics are based on the following universal laws of conservation (Tannehill et al. 1997):

- Conservation of Mass (continuity equation),
- Conservation of Momentum (momentum equation),
- Conservation of Energy (energy equation).

Based on the continuity equation and momentum equation, hydrodynamic lubrication can be expressed mathematically in the form of an equation commonly known as the Reynolds equation, which is a simplification of the famous Navier-Stokes equation. The energy equation is the basis of thermal analysis and will be discussed in the next chapter.

There are several ways to derive the Reynolds equation. It can be derived from the full Navior-Stokes equation. It is more often derived by simply applying a typical engineering approach and considering the equilibrium of an element of liquid subjected to viscous shear and applying the continuity of flow principle. A steady-state Reynolds equation, including non-Newtonian behaviour based on the Eyring model, was developed by Conry et al. (1987) using this approach. With the introduction of rough surfaces in the two contact solids, transient effects must be considered due to the time varying geometries. In the

following the transient Reynolds equation is derived after Conry et al. (1987) incorporating the time-dependent squeeze-film term.

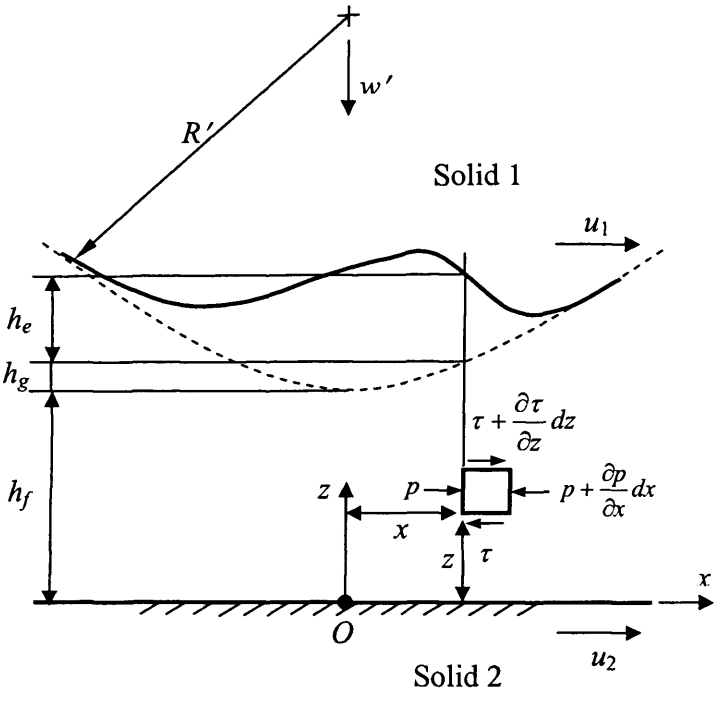


Figure 2.1: Illustration of EHL line contact.

Consider the case shown in Figure 2.1 where the two solids are separated by a thin lubricating film. By considering the elemental volume of the fluid, the equation of equilibrium in the x-direction can be written as

$$\frac{\partial \tau}{\partial z} = \frac{\partial p}{\partial x} \tag{2.1}$$

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 does not vary across the film, integration of Equation (2.1) results in

$$\tau = \tau_2 + \frac{\partial p}{\partial x} z \tag{2.2}$$

where τ_2 is the shear stress at surface 2. The Eyring rheological model that relates shear stress to shear strain rate, ignoring the elastic shear term, is

$$\dot{\gamma} = \frac{\partial u}{\partial z} = \frac{\tau_0}{\eta} \sinh\left(\frac{\tau}{\tau_0}\right) \tag{2.3}$$

Substituting Equation (2.2) into Equation (2.3) and integrating gives the velocity of the fluid at point (x, z) as

$$\begin{aligned} u &= u_2 + \int_0^z \frac{\tau_0}{\eta} \sinh\left(\frac{\tau_2 + \frac{\partial p}{\partial x} s}{\tau_0}\right) ds \\ &= u_2 + \frac{\tau_0^2}{\eta \frac{\partial p}{\partial x}} \left[\cosh\left(\frac{\tau_2 + \frac{\partial p}{\partial x} z}{\tau_0}\right) - \cosh\left(\frac{\tau_2}{\tau_0}\right) \right] \end{aligned} \tag{2.4}$$

Defining the mid-plane ($z = h/2$) shear stress

$$\tau_m = \tau_2 + \frac{\partial p}{\partial x} \frac{h}{2} \tag{2.5}$$

and a dimensionless function

$$\Sigma = \frac{h}{2\tau_0} \frac{\partial p}{\partial x} \tag{2.6}$$

where h is the film thickness. Equation (2.4) can be rewritten as

$$u = u_2 + \frac{\tau_0 h}{2\eta\Sigma} \left\{ \cosh\left[\frac{\tau_m}{\tau_0} - \Sigma + \frac{2z}{h}\Sigma\right] - \cosh\left(\frac{\tau_m}{\tau_0} - \Sigma\right) \right\} \tag{2.7}$$

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

$$u_1 - u_2 = \frac{\tau_0 h}{\eta \Sigma} \sinh\left(\frac{\tau_m}{\tau_0}\right) \sinh(\Sigma) \tag{2.8}$$

Therefore, the mid-plane shear stress can be determined from

$$\sinh\left(\frac{\tau_m}{\tau_0}\right) = \frac{\eta(u_1 - u_2)}{\tau_0 h} \frac{\Sigma}{\sinh \Sigma} \tag{2.9}$$

The mass flow is defined as

$$M = \int_0^h \rho u dz \tag{2.10}$$

Substitution of Equation (2.7) into Equation (2.10) and integrating yields

$$\begin{aligned} M &= \rho u_2 h + \rho \frac{\tau_0 h}{2\eta\Sigma} \left\{ \frac{h}{2\Sigma} \left[\sinh\left(\frac{\tau_m}{\tau_0} + \Sigma\right) - \sinh\left(\frac{\tau_m}{\tau_0} - \Sigma\right) \right] - h \cosh\left(\frac{\tau_m}{\tau_0} - \Sigma\right) \right\} \\ &= \rho u_2 h + \rho \frac{\tau_0 h^2}{2\eta} \left[\frac{\sinh \Sigma}{\Sigma} \sinh\left(\frac{\tau_m}{\tau_0}\right) + \frac{\sinh \Sigma - \Sigma \cosh \Sigma}{\Sigma^2} \cosh\left(\frac{\tau_m}{\tau_0}\right) \right] \end{aligned} \tag{2.11}$$

Substitution of Equation (2.9) into Equation (2.11) leads to

$$\begin{aligned} M &= \rho h \frac{u_1 + u_2}{2} + \frac{\rho h^3}{12\eta} \frac{\partial p}{\partial x} \left[\frac{3(\sinh \Sigma - \Sigma \cosh \Sigma)}{\Sigma^3} \cosh\left(\frac{\tau_m}{\tau_0}\right) \right] \\ &= M_C + M_P \end{aligned} \tag{2.12}$$

The total flow M comprises of two parts, the Couette flow M_C, which is related to the entrainment of the fluid and the Poiseuille flow M_P, which is the pressure driven flow. By defining a mean velocity

$$\bar{u} = \frac{u_1 + u_2}{2} \tag{2.13}$$

and a dimensionless factor

$$S = -\frac{3(\sinh\Sigma - \Sigma\cosh\Sigma)}{\Sigma^3}\cosh\left(\frac{\tau_m}{\tau_0}\right) \tag{2.14}$$

Equation (2.12) can be simplified as

$$M = \rho\bar{u}h - \frac{\rho h^3}{12\eta}S\frac{\partial p}{\partial x} \tag{2.15}$$

Conservation of net mass flow yields

$$\frac{\partial M}{\partial x} + \frac{\partial}{\partial t}(\rho h) = 0 \tag{2.16}$$

where t is time. Substituting equation (2.15) into equation (2.16) and rearranging gives

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{12\eta}S\frac{\partial p}{\partial x}\right) - \frac{\partial}{\partial x}(\rho\bar{u}h) - \frac{\partial}{\partial t}(\rho h) = 0 \tag{2.17}$$

Equation (2.17) is the transient Reynolds equation based on the Eyring equation for the fluid rheology. Removing the last term related to time t results in the steady-state Reynolds equation. Using hyperbolic identities, the non-Newtonian factor S becomes (Hughes et al. 2000)

$$S = \frac{3(\Sigma\cosh\Sigma - \sinh\Sigma)}{\Sigma^3}\sqrt{1 + \frac{\eta^2(u_2 - u_1)^2}{\tau_0^2 h^2}\frac{\Sigma^2}{\sinh^2\Sigma}} \tag{2.18}$$

with Σ defined by Equation (2.6). This demonstrates that S depends on the pressure, film thickness, siding speed and pressure gradient. Taking a fixed value of unity for factor S results in the Newtonian fluid formulation.

2.2.2 Film Thickness Equation

As shown in Figure 2.1, the film thickness profile is given by

$$h(x,t) = \phi(x,t) + h_g(x) + h_e(x,t) + h_f$$
$$= \phi(x,t) + \frac{x^2}{2R'} - \frac{4}{\pi E'}\int p(s,t)\ln|x-s|ds + h_f \tag{2.19}$$

where the expression $\phi(x, t)$ is the composite surface roughness. h_g is the separation due to the geometry of the undeformed solids, h_e is the combined elastic deformation of the solids due to pressure distribution p assumed to be that experienced by semi-infinite bodies in a state of plane strain, and h_f is a constant.

2.2.3 Other Equations

The non-conformal geometry of the contacting surfaces causes an intense concentration of load over a very small area and extreme pressures (typically >1GPa) in the lubricant are inevitable. Lubricant viscosity and density may change dramatically under such a high pressure. For isothermal situations, the viscosity and density dependence on pressure are determined by the Roelands equation and the Dowson and Higginson relation, respectively, as described in Chapter 1. These equations can be written in the following general forms:

$$\eta = \eta_0 \exp\left\{\ln(\eta_0/\kappa)[(1+\chi p)^Z - 1]\right\} \tag{2.20}$$

$$\rho = \rho_0 \frac{1+\gamma p}{1+\lambda p} \tag{2.21}$$

The load carried by the EHL film is the integral of the pressure distribution and can be expressed as

$$w' = \int p(x)dx \tag{2.22}$$

2.3 Numerical Solutions

Numerical solutions can be obtained by disretization of the governing equations. The study reported in this thesis is based on numerical analysis of EHL contacts using a novel coupled differential deflection solution method developed by Cardiff Tribology Group (Hughes et al. 2000, Elcoate et al. 2001, Holmes et al. 2003a, 2005a, 2005b). Both the accuracy and the efficiency of the method have been verified by Hughes et al. (2000) and Holmes et al. (2003a). The essence of the method is that the two fundamental equations (Reynolds equation and elastic deflection equation) are fully coupled in the solution scheme.

2.3.1 Differential Deflection Technique

The elastic deflection given in Equation (2.19) is the convolution integration of pressure with the appropriate weighting function:

$$u(x) = -\frac{4}{\pi E'} \int p(s) \ln|x - s| ds \tag{2.23}$$

The numerical calculation of deflection at a point can be reduced to a simple quadrature:

$$u(x_i) = \sum_k g_{k-i} p_k \tag{2.24}$$

where the influence coefficients g_{k-i} depend on the geometry and material properties.

In the differential deflection technique developed by Evans and Hughes (2000) Equations (2.23) and (2.24) are expressed in a differential form:

$$\frac{\partial^2 u(x_i)}{\partial x^2} = \sum_k f_{k-i} p_k \tag{2.25}$$

The purpose of such a modification is based on the observation that the normalized f_j coefficients decay significantly faster than the g_j coefficients as the indices j increase from zero, confirming that the effect of pressure on the second order derivative of the deflection is extremely localised.

The summation of pressure multiplied by the influence coefficients in Equation (2.25) can be a time consuming process for solving fine meshed problems as it requires $O(N^2)$ operations, where N is the number of nodes. To reduce CPU time, the fast Fourier transform (FFT) technique has been used to evaluate the convolution integral in Equation (2.25). This approach reduces the number of operations in the order $O(N\log_2 N)$. Therefore, for very fine meshes the potential time saving can be enormous.

With the differential deflection technique, the film thickness equation (2.19) can be written in the following differential form:

$$\frac{\partial^2 h}{\partial x^2} = \frac{\partial^2 \phi}{\partial x^2} + \frac{1}{R'} + \sum_k f_{k-i} p_k \tag{2.26}$$

2.3.2 Coupled Solution Method

In the coupled method, the Reynolds equation and the film thickness equation are solved as a set of simultaneous equations. Discretised equations (2.17) and (2.26) are expressed as

$$\sum_{k=0}^{n_c} A_k p_k + \sum_{k=0}^{n_c} B_k h_k = R_i \tag{2.27a}$$

$$\sum_{k=0}^{n_c} C_k p_k + \sum_{k=0}^{n_c} D_k h_k = E_i \tag{2.27b}$$

For each point i in the computing mesh where p_0 and h_0 are the values of p and h at the mesh point, and p_k and h_k are the values at the neighbouring mesh points $k = 1,2,..n_c$ that make a contribution to the discretised equations. For the line contact situation these equations have a narrow bandwidth (n_c) and may be solved by elimination. For a point contact application the bandwidth increases so that an iterative approach is appropriate. An iterative approach is also advantageous in mixed lubrication situations for point and line contacts as it allows the existence or otherwise of a full film at any mesh point to be determined as the equations are solved.

When the discretised equations are to be solved iteratively they are organised into a pair of simultaneous equations in the variables p_0 and h_0:

$$A_0 p_0 + B_0 h_0 = \hat{R}_i \tag{2.28a}$$

$$C_0 p_0 + D_0 h_0 = \hat{E}_i \tag{2.28b}$$

where

$$\hat{R}_i = R_i - \sum_{k=1}^{n_c} A_k p_k - \sum_{k=1}^{n_c} B_k h_k \tag{2.29a}$$

$$\hat{E}_i = E_i - \sum_{k=1}^{n_c} C_k p_k - \sum_{k=1}^{n_c} D_k h_k \tag{2.29b}$$

Equations (2.28) are solved using a coupled iterative scheme to update the values of all the unknown nodal values of p and h by solving Equations (2.29) as a pair of simultaneous equations in p_0 and h_0:

$$p_0^{new} = \frac{\hat{R}_i D_0 - \hat{E}_i B_0}{A_0 D_0 - B_0 C_0} \tag{2.30a}$$

$$h_0^{new} = \frac{\hat{E}_i A_0 - \hat{R}_i C_0}{A_0 D_0 - B_0 C_0} \tag{2.30b}$$

Equations (2.28) are solved in this way within each timestep. The rheological behaviour ensures that coefficients A_k and B_k are dependent on p and h in a highly non-linear way. Coefficients E_i are also pressure dependent in that they contain the pressure terms in Equation (2.26) for mesh points that are not included in the region $k < n_c$. These factors require Equations (2.28) to be solved repeatedly within each timestep. However, the coupled nature of the formulations and the differential deflection technique allow timestep convergence to be achieved rapidly.

The strategy described by Holmes et al. (2005a, 2005b) is used to deal with direct contacts between components having significant surface roughness. If during iteration within a timestep the value of h_0^{new} obtained from Equation (2.30b) is negative then Equation (2.28a) is changed and replaced by the boundary condition $h_0 = 0$ resulting in

$$h_0^{new} = 0 \tag{2.31a}$$

$$p_0^{new} = \hat{E}_i / C_0 \tag{2.31b}$$

This effectively replaces the Reynolds equation with the boundary condition $h = 0$, and applies the deflection equation subject to that boundary condition.

Equations (2.30) are thus used at each mesh point during each iterative sweep and are replaced by Equations (2.31) at mesh points where the current evaluation of Equations (2.30) yield a negative value for h_0^{new}. The ease with which contact conditions can be incorporated using this approach is a further advantage of the coupled differential deflection technique. Holmes et al. (2005a) have shown that this technique determines that contact occurs at a mesh point in a way that is entirely consistent with that inferred from a gradual reduction in viscosity.

2.4 Fourier Representation for Rough Profiles

In this thesis, the profiles of rough surfaces used for the analyses are taken from Holmes et al. (2004) and shown in Figure 2.2, where in each case the solid metal is below the profile. These profiles are respectively denoted as (A), (B) and (C) for consistency with that work. In the figure, profiles (B) and (C) are shown with a negative offset by 2.5 and 5μm, respectively, for clarity. Trace (A) is a profile taken from a well run-in transverse ground disk used in the scuffing experiments by Patching et al. (1995). Traces (B) and (C) are taken from FZG micropitting tests at QinetiQ and have been run for several load stages, and as a result have become run-in to some extent, but close examination shows that they have sharper asperities than profile (A).

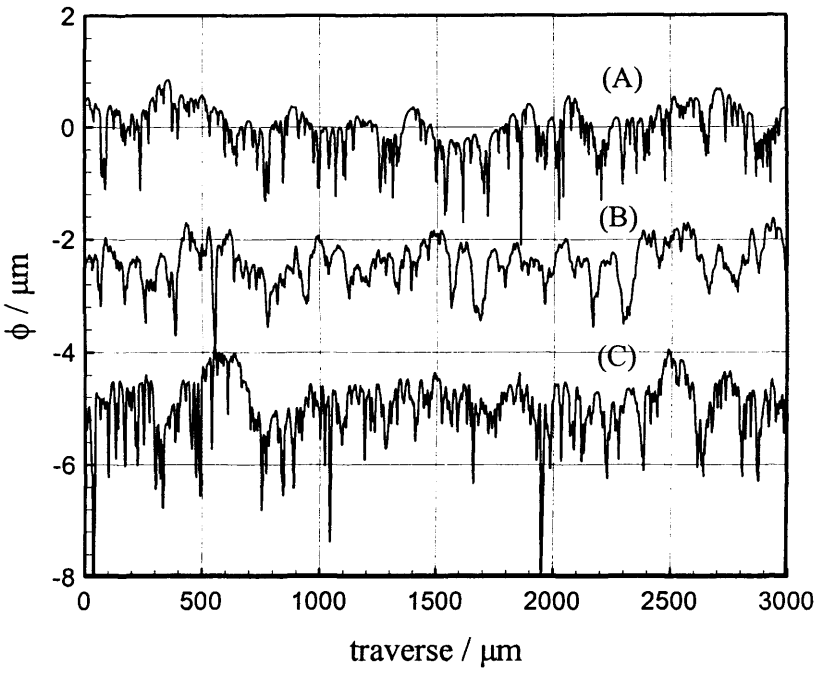


Figure 2.2: Surface profiles used for analyses. R_a values are (A) 0.32μm, (B) 0.22μm, and (C) 0.31μ m.

The roughness features are measured by surface profilometer. Interpolation is required between measured points in the trace(s) in order to determine ϕ (x, t) in Equation (2.26) at all time steps. It is also necessary to interpolate between measured points if the computing mesh is different from the resolution at which the profilometer trace was acquired. In previous work (Hughes et al. 2000, Elcoate et al. 2001, Holmes et al. 2004, 2005a), each surface was represented by a family of cubic splines so that it was piecewise continuous with continuous slope. This property is important as discontinuity in slope with a moving rough surface leads to a fictitious pumping action due to the consequent variation in the squeeze film effect. The cubic spline approach avoids this difficulty and ensures that the mesh point values of the rough surface profile obtained through interpolation are a consistent representation of the component's surface finish.

In the following, an alternative approach is developed where Fourier series methods are used to represent rough profiles. In this way, the rough surface has not only continuous slops but also continuous second (and higher) order derivatives. This might be important for the application of the differential deflection technique, which uses second order derivatives. Another advantage of using Fourier series is that in order to calculate using a large number of timesteps, the rough profile can be used repeatedly due to the periodic nature of Fourier series.

2.4.1 Analysis Procedure

With the use of Fourier series, the real rough surface can be expressed over a length L in its Fourier polynomial of degree m (Tolstov 1976):

$$\phi(x) = \frac{a_0}{2} + \sum_{k=1}^{m} \left\{ a_k \cos\left(\frac{2k\pi x}{L}\right) + b_k \sin\left(\frac{2k\pi x}{L}\right) \right\} \tag{2.32}$$

Fourier coefficients a_k (k = 0, 1, 2, ..., m) and b_k (k = 1, 2, ..., m) can be determined by Fourier transform (Brigham 1988) or numerical integration (Tolstov 1976). Suppose that there are N subintervals of equal spacing $\delta x = L / N$ and ϕ_i is the rough value measured at $x_i = i\delta x$. Then the coefficients are

$$a_k = \frac{2}{N} \sum_{i=0}^{N-1} \cos\left(\frac{2k\pi}{N} i\right) \phi_i \tag{2.33a}$$

$$b_k = \frac{2}{N} \sum_{i=0}^{N-1} \sin\left(\frac{2k\pi}{N} i\right) \phi_i \tag{2.33b}$$

The construction is possible provided $2m + 1 \leq N$.

Another way to determine Fourier coefficients is an iterative algorithm described by Xu (1994) and is employed in this thesis. If there are $2m+1$ sample points in the interval [0, L] with equal width $\delta x = L / (2m+1)$, then the first $2m+1$ coefficients are determined as follows:

(1) Calculate w_1 and w_2:

$$\begin{cases} w_{2m+2} = w_{2m+1} = 0 \\ w_i = \phi_i + 2w_{i+1}\cos k\theta - w_{i+2}, \quad i = 2m, 2m-1, \cdots, 2, 1 \end{cases} \tag{2.34}$$

where $\theta = \dfrac{2\pi}{2m+1}$.

(2) Calculate a_k and b_k:

$$\begin{cases} a_k = \dfrac{2}{2m+1}(\phi_0 + w_1\cos k\theta - w_2) \\ b_k = \dfrac{2}{2m+1} w_1 \sin k\theta \end{cases} \tag{2.35}$$

In the above calculations, $\cos(k\theta)$ and $\sin(k\theta)$ are calculated by the following recursive formulas

$$\begin{cases} \cos(k\theta) = \cos\theta\cos(k-1)\theta - \sin\theta\sin(k-1)\theta \\ \sin(k\theta) = \sin\theta\cos(k-1)\theta + \cos\theta\sin(k-1)\theta \end{cases} \tag{2.36}$$

Once Fourier coefficients are determined, like the cubic spline approach, the mesh point values of the rough surface profile are obtained through interpolation by Equation (2.32). In this thesis, the time step used is $\Delta t = \Delta x/2u_1$, where u_1 is the velocity of the faster moving surface and Δx is the mesh size applied in the contact area. That is, the faster surface moves a distance of $\Delta x/2$ for every timestep. Timesteps of this magnitude (or less) were shown by Holmes et al. (2003a) to be necessary to achieve consistent solutions to the transient micro-EHL problem. Therefore, the interpolation is only needed once for the faster moving surface at a mesh spacing of $\Delta x/2$. For the slower moving surface, however, the interpolation is needed at each timestep during the time-dependent solution to accommodate any values of slide roll ratio, which is defined as

$$\xi = \frac{u_s}{\bar{u}} = \frac{u_1 - u_2}{(u_1 + u_2)/2} \tag{2.37}$$

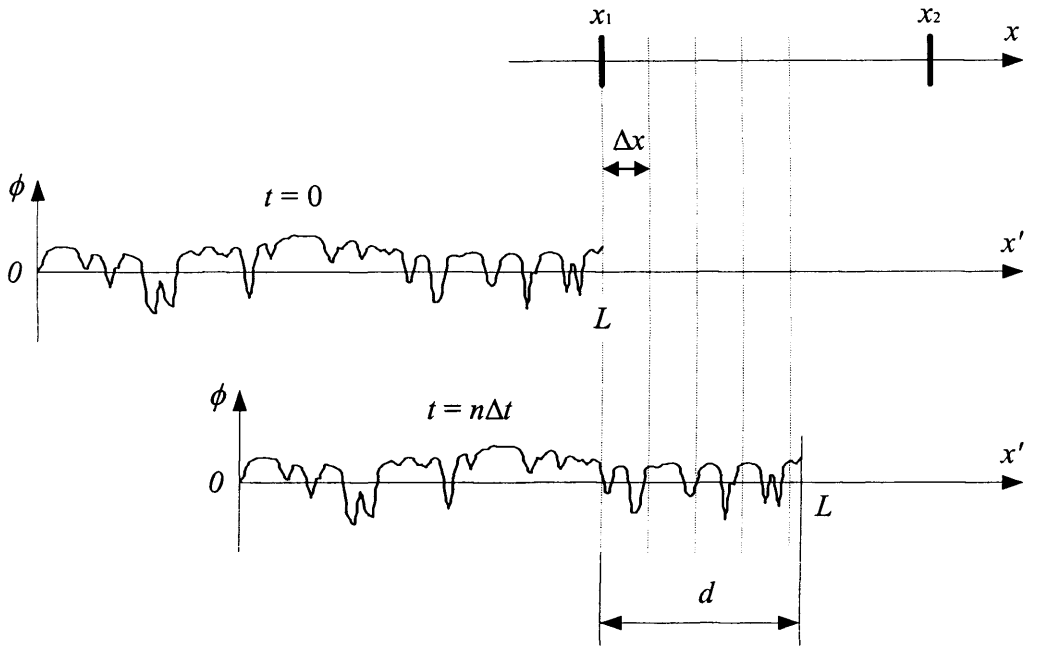


Figure 2.3: Schematic representation of moving rough surface.

As shown in Figure 2.3, the EHL computational domain is defined in the interval $x_1 \leq x \leq x_2$ and the mesh size is Δx. To use Equation (2.32) to obtain the roughness profile values at mesh points, the x value of each mesh point has to be transformed to the local coordinate of the Fourier series, which is denoted as x' to distinguish it from x. The surface is assumed to move from left to right and the point $x' = L$ lies at $x = x_1$ at the time $t = 0$. At the nth timestep, i.e. $t = n\Delta t$, the rough surface moves a distance of d. For the faster surface, d is

$$d = u_1 n\Delta t = \frac{n\Delta x}{2} \tag{2.38}$$

and for the slower surface, d is

$$d = u_2 n\Delta t = \frac{u_2}{u_1} \cdot \frac{n\Delta x}{2} \tag{2.39}$$

From Figure 2.3, the corresponding x' values at mesh points are

$$(L-d),\quad (L-d)+\Delta x,\quad (L-d)+2\Delta x,\quad \cdots,\quad (L-d)+i\Delta x,\quad \cdots$$

Thus, the values of the rough surface profile at mesh points can be obtained by calculating $\phi(x')$ using Equation (2.32).

2.4.2 Numerical Results

In this section, the three rough profiles illustrated in Figure 2.2 are represented by the scheme of Fourier series described in the previous section. The scheme is then incorporated into the EHL line contact analysis procedure, i.e. the coupled solution method with differential deflection technique. The operating conditions and lubricant parameters adopted for EHL analysis are specified in Table 2.1.

Table 2.1: Parameters used for transient EHL line contact analysis.

w'	527 kN/m	τ_0	10 MPa
a	0.335 mm	α	11.1 GPa^{-1}
p_0	1.0 GPa	λ	1.68 GPa^{-1}
E_1, E_2	207 GPa	γ	2.27 GPa^{-1}
ν_1, ν_2	0.3	κ	63.2×10^{-6} Pa·s
R_1, R_2	38.1 mm	χ	5.1 GPa^{-1}
η_0	0.0048 Pa·s		

Fourier series is used to provide surface heights at intermediate points between the measured points as required. Figure 2.4 shows three rough profiles with both measured points and interpolated curves. For clarity, only a small section of each profile in the interval between 1000μm and 1500μm is shown in the figure. The spacings of measured points for profiles (A), (B) and (C) are δx = 3.53, 1.27 and 1.25 μm, respectively. The interpolation

mesh spacing adopted is $\Delta x = a/200 = 1.67\mu m$. It is clear that both profiles (B) and (C) can be perfectly represented by Fourier series. For profile (A) Fourier series does indeed correctly interpolate the measured values, but close examination shows that oscillation between measured points occurs at some positions. This effect is known as aliasing. Aliasing occurs if measured points are not taken sufficiently close together (Brigham 1988). The result is that frequency components outside the range of $[-f_c, f_c]$ are aliased into that range, where f_c is the Nyquist critical frequency and given by (Brigham 1988)

$$f_c = \frac{1}{2\delta x} \tag{2.40}$$

δx is the spacing of measured points. If $\delta x > 1/2f_c$, aliasing will result. Therefore, the only effective way to eliminate aliasing is to adopt a smaller measuring spacing.

To reduce oscillation, one may eliminate high frequency components in the Fourier series by replacing Equation (2.32) with

$$\phi(x) = \frac{a_0}{2} + \sum_{k=1}^{l} \left\{ a_k \cos\left(\frac{2k\pi x}{L}\right) + b_k \sin\left(\frac{2k\pi x}{L}\right) \right\} \tag{2.41}$$

where $l < m$ specifies the selected cut-off point. Figure 2.5 shows the results obtained for profile (A) with 10%, 25% and 50% cut-off, i.e. $l = 0.9m$, $0.75m$ and $0.5m$, respectively. Low frequency components retained in Equation (2.41) capture the essential feature of the original measured data while simultaneously reducing the high frequency terms. Some features of peaks and valleys (if they are not noise) may be lost with the cut-off. For this particular case, the cubic spline approach might be a better choice.

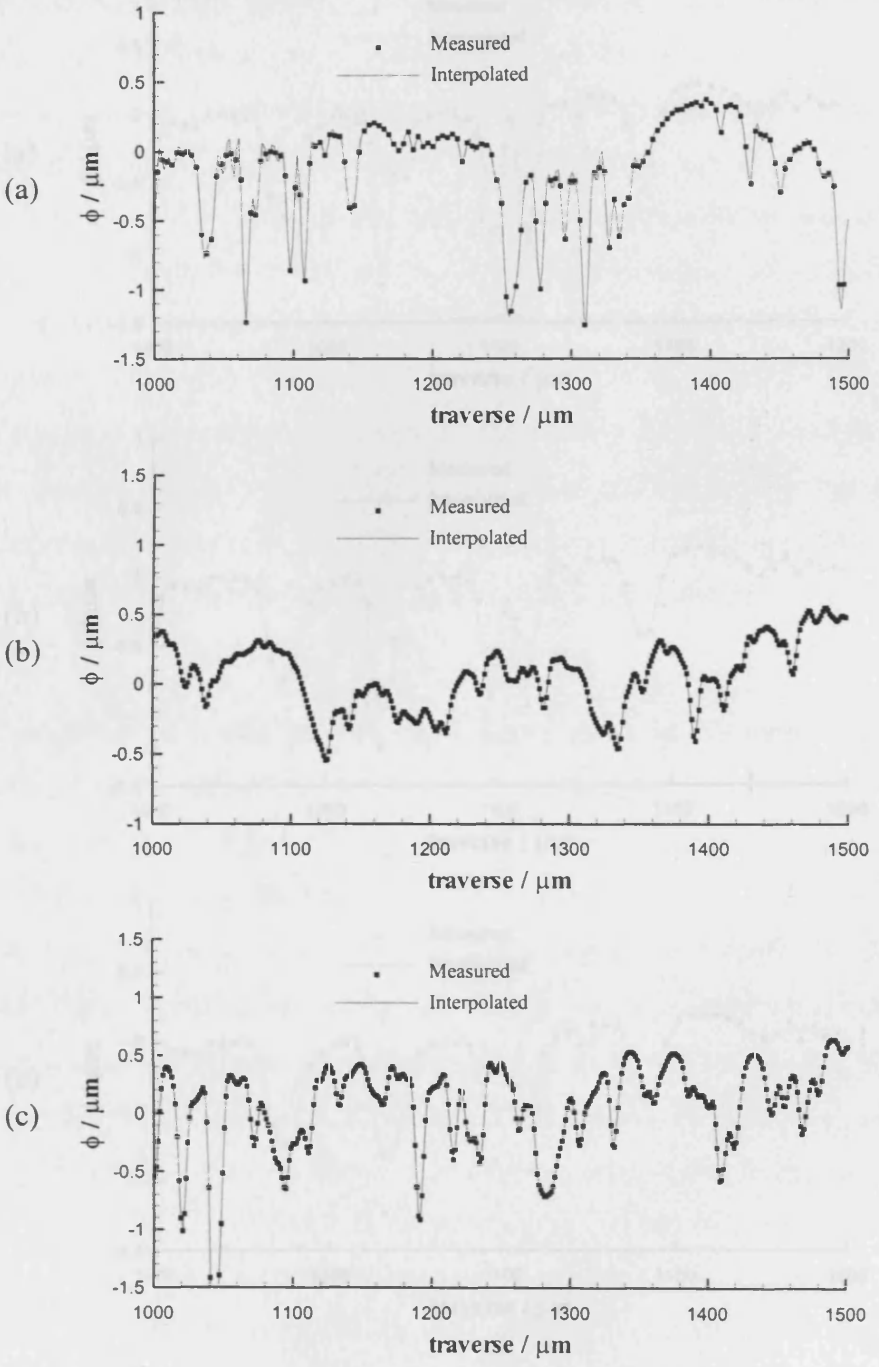


Figure 2.4: Fourier representation for rough profiles with $\Delta x = a/200$: (a) Profile (A), (b) Profile (B) and (c) Profile (C).

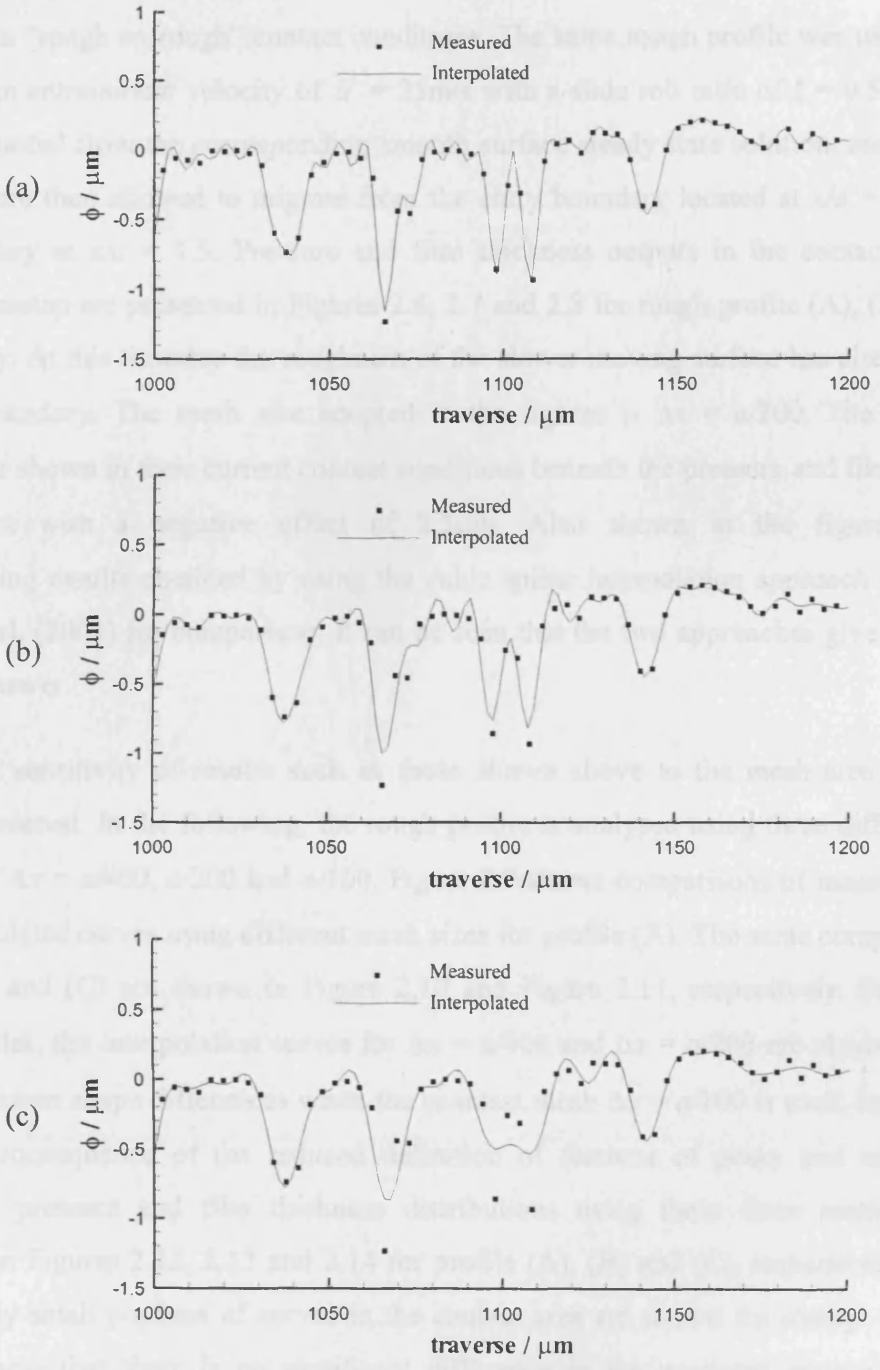


Figure 2.5: Fourier representation for rough profile (A) with high frequency cut-off: (a) 10%, (b) 25% and (c) 50%.

With representation by Fourier series, each of the three profiles was taken in turn and simulated in "rough on rough" contact conditions. The same rough profile was used for both surfaces. An entrainment velocity of $\bar{u}$ = 25m/s with a slide roll ratio of ξ = 0.5 is applied. Analyses started from the corresponding smooth surface steady state solution, and the rough surfaces were then allowed to migrate from the entry boundary located at x/a = -2.5 to the exit boundary at x/a = 1.5. Pressure and film thickness outputs in the contact area at a specific timestep are presented in Figures 2.6, 2.7 and 2.8 for rough profile (A), (B) and (C), respectively. At this timestep the roughness of the slower moving surface has already passed the exit boundary. The mesh size adopted in the figures is $\Delta x = a/200$. The two rough surfaces are shown in their current contact conditions beneath the pressure and film thickness distributions with a negative offset of 2.5μm. Also shown in the figures are the corresponding results obtained by using the cubic spline interpolation approach adopted by Elcoate et al. (2001) for comparison. It can be seen that the two approaches give essentially the same answer.

The sensitivity of results such as those shown above to the mesh size adopted is clearly of interest. In the following, the rough profile is analysed using three different mesh spacings of $\Delta x = a/400$, $a/200$ and $a/100$. Figure 2.9 shows comparisons of measured points with interpolated curves using different mesh sizes for profile (A). The same comparisons for profile (B) and (C) are shown in Figure 2.10 and Figure 2.11, respectively. For all three rough profiles, the interpolation curves for $\Delta x = a/400$ and $\Delta x = a/200$ are almost identical. There are minor shape differences when the coarsest mesh $\Delta x = a/100$ is used, but this is an inevitable consequence of the reduced definition of features of peaks and valleys. The consequent pressure and film thickness distributions using these three mesh sizes are compared in Figures 2.12, 2.13 and 2.14 for profile (A), (B) and (C), respectively. In these figures, only small portions of curves in the contact area are shown for clarity. The results obtained show that there is no significant difference in the predicted pressure and film thickness profiles as the mesh is varied. The largest discrepancies are seen with profile (A) and the main cause of this is the relative coarseness of the profile information where the height measurement spacing was 3.5μm, whereas the finest computing mesh used in this comparison has a mesh spacing of about 0.8μm.

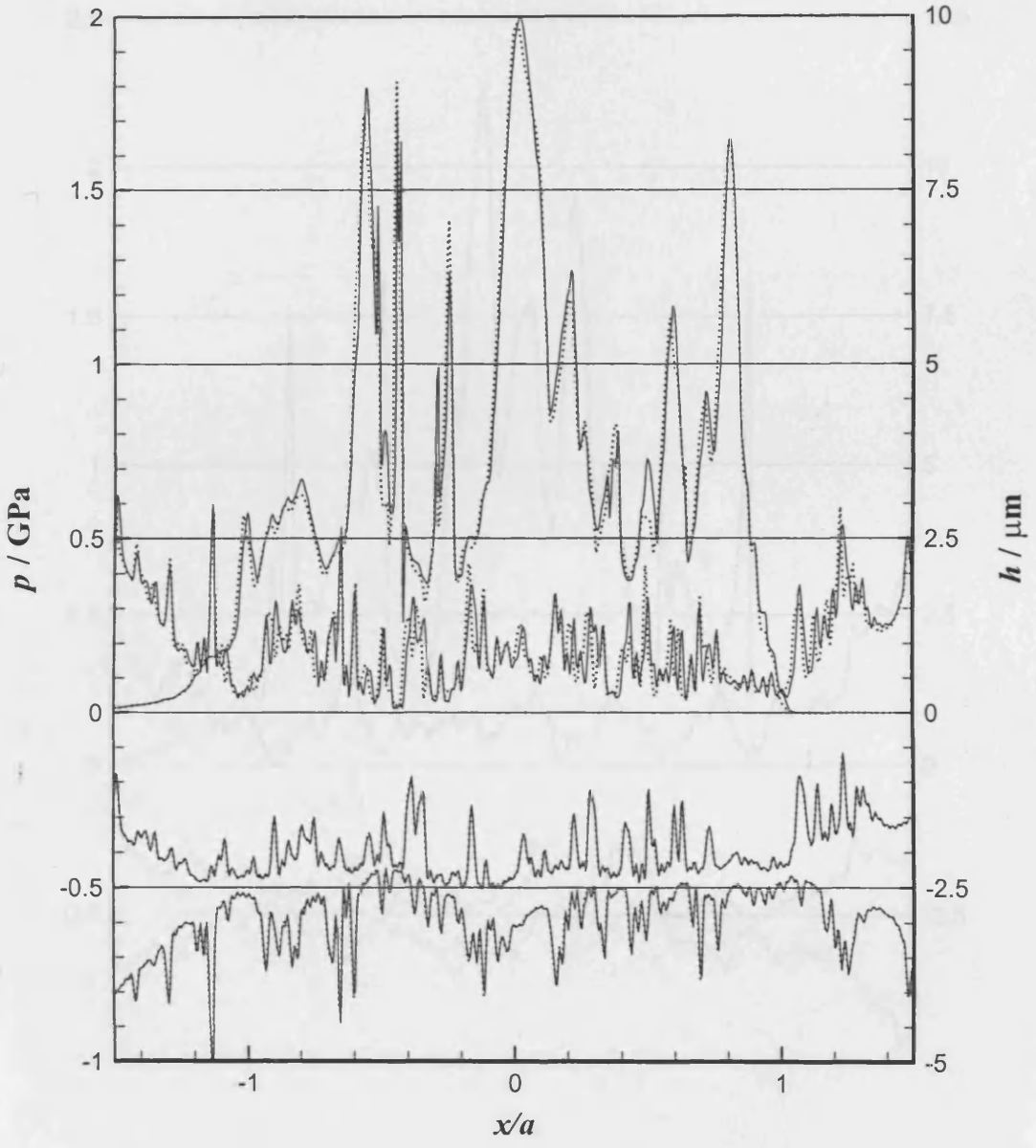


Figure 2.6: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (A). The two rough surfaces are shown below in their deflected positions offset by 2.5μm for clarity. Solid lines are results using Fourier analysis and dotted lines are results obtained by cubic spline approach.

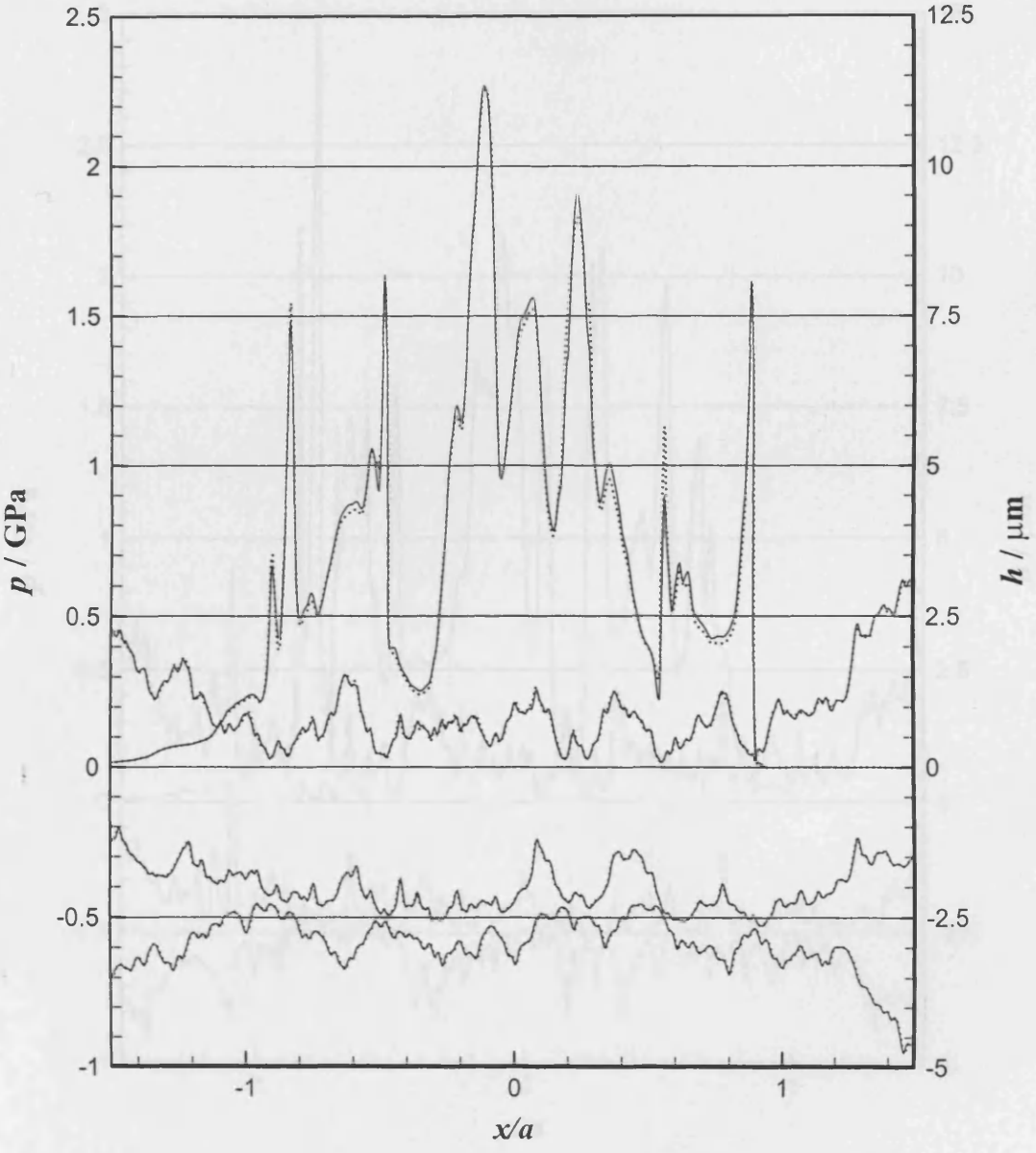


Figure 2.7: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (B). The two rough surfaces are shown below in their deflected positions offset by 2.5μm for clarity. Solid lines are results using Fourier analysis and dotted lines are results obtained by cubic spline approach.

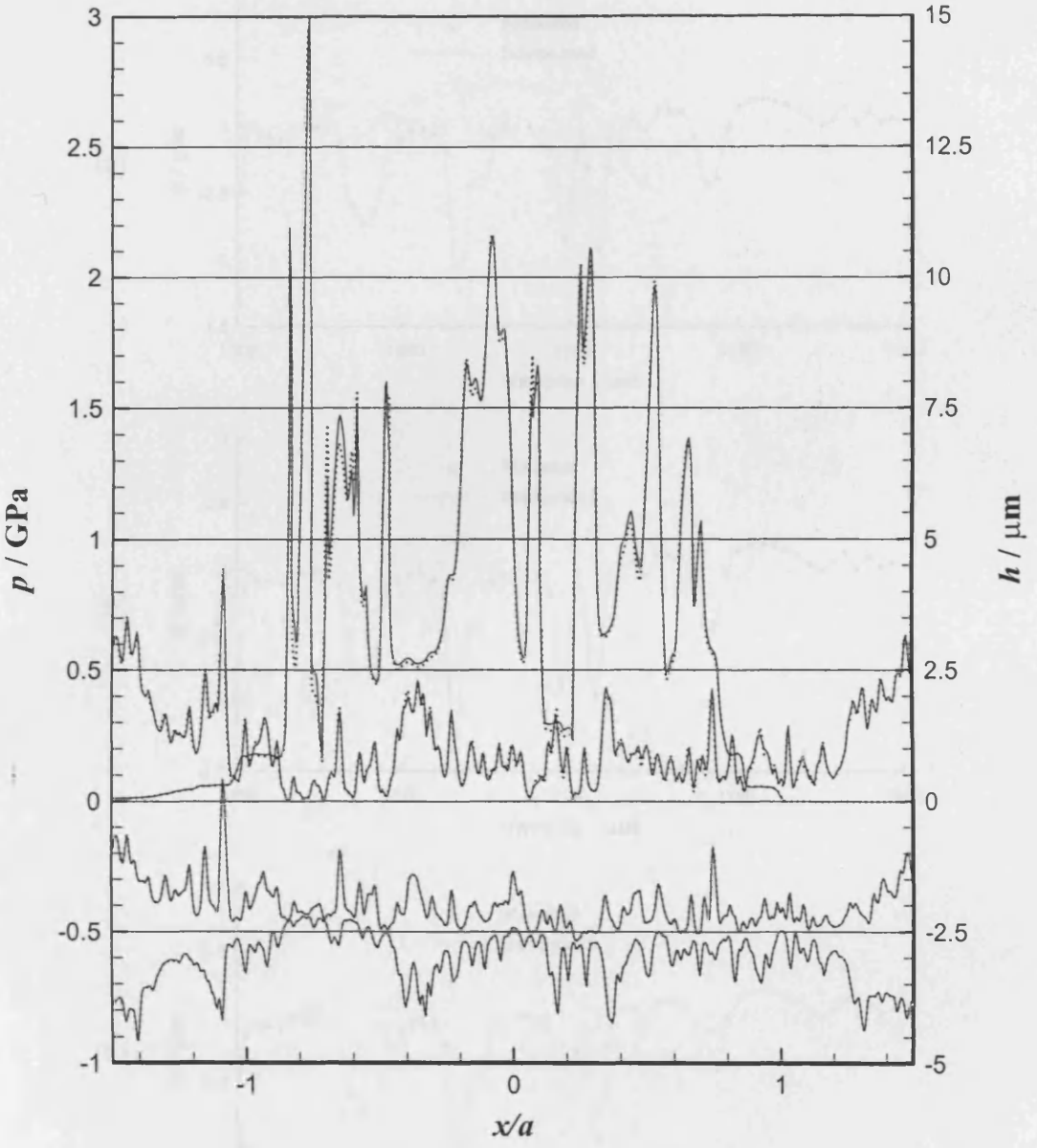


Figure 2.8: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (C). The two rough surfaces are shown below in their deflected positions offset by 2.5μm for clarity. Solid lines are results using Fourier analysis and dotted lines are results obtained by cubic spline approach.

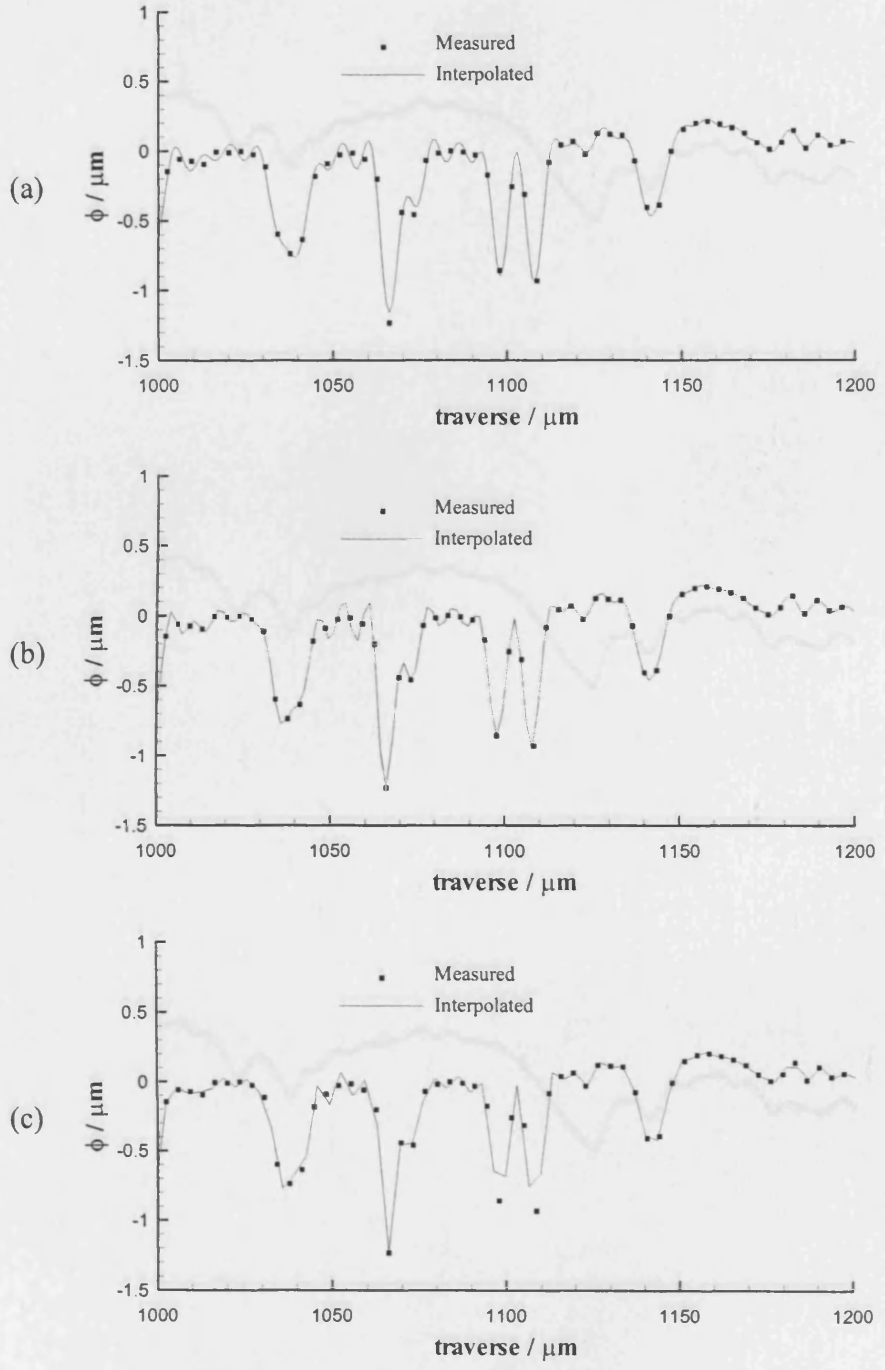


Figure 2.9: Fourier representation for profile (A) using different mesh size: (a) $\Delta x = a/400$, (b) $\Delta x = a/200$ and (c) $\Delta x = a/100$.

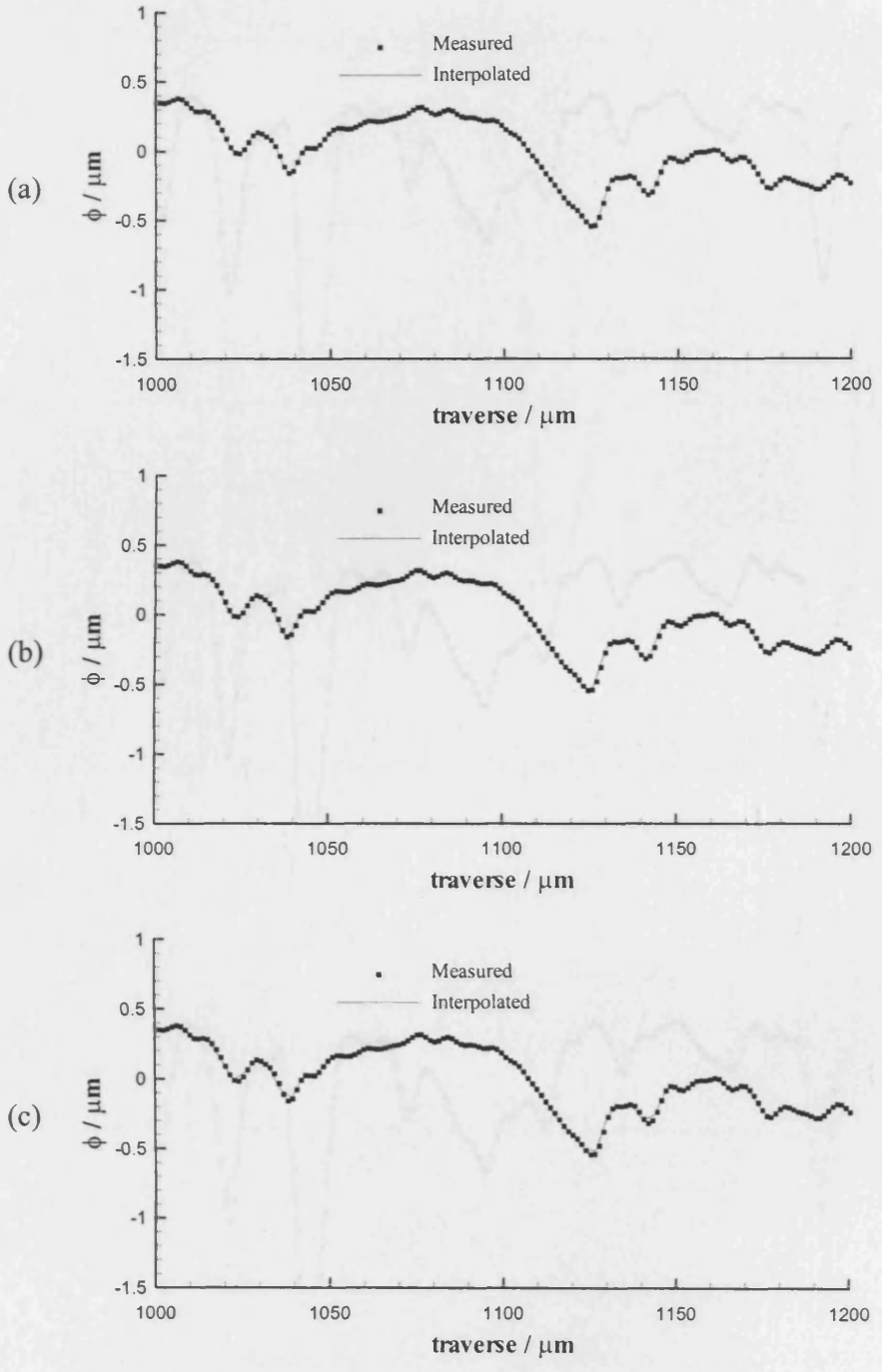


Figure 2.10: Fourier representation for profile (B) using different mesh size: (a) $\Delta x = a/400$, (b) $\Delta x = a/200$ and (c) $\Delta x = a/100$.

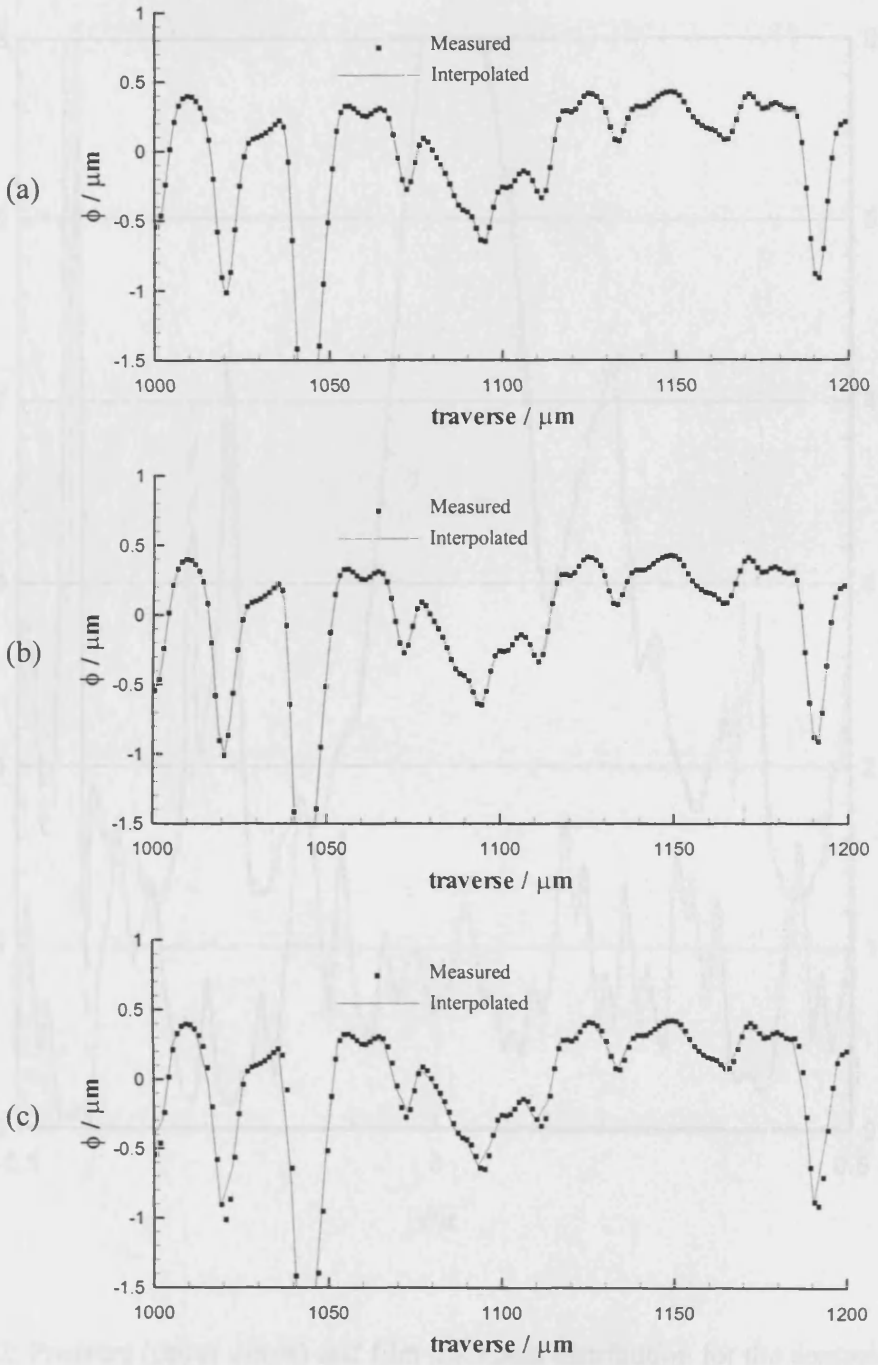


Figure 2.11: Fourier representation for profile (C) using different mesh size: (a) $\Delta x = a/400$, (b) $\Delta x = a/200$ and (c) $\Delta x = a/100$.

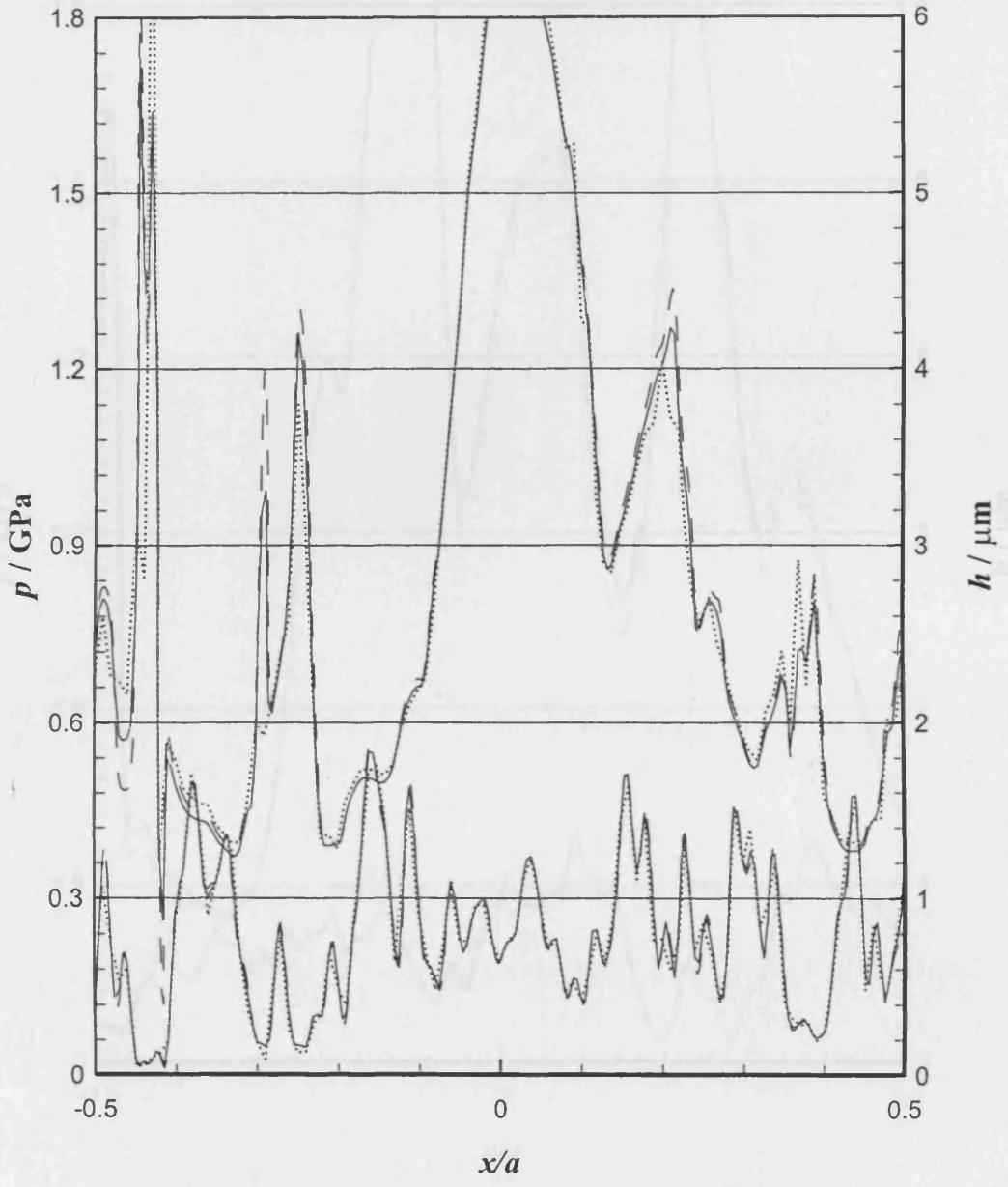


Figure 2.12: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (A) using different mesh size. Dashed lines are for $\Delta x = a/400$, solid lines are for $\Delta x = a/200$ and dotted lines are for $\Delta x = a/100$.

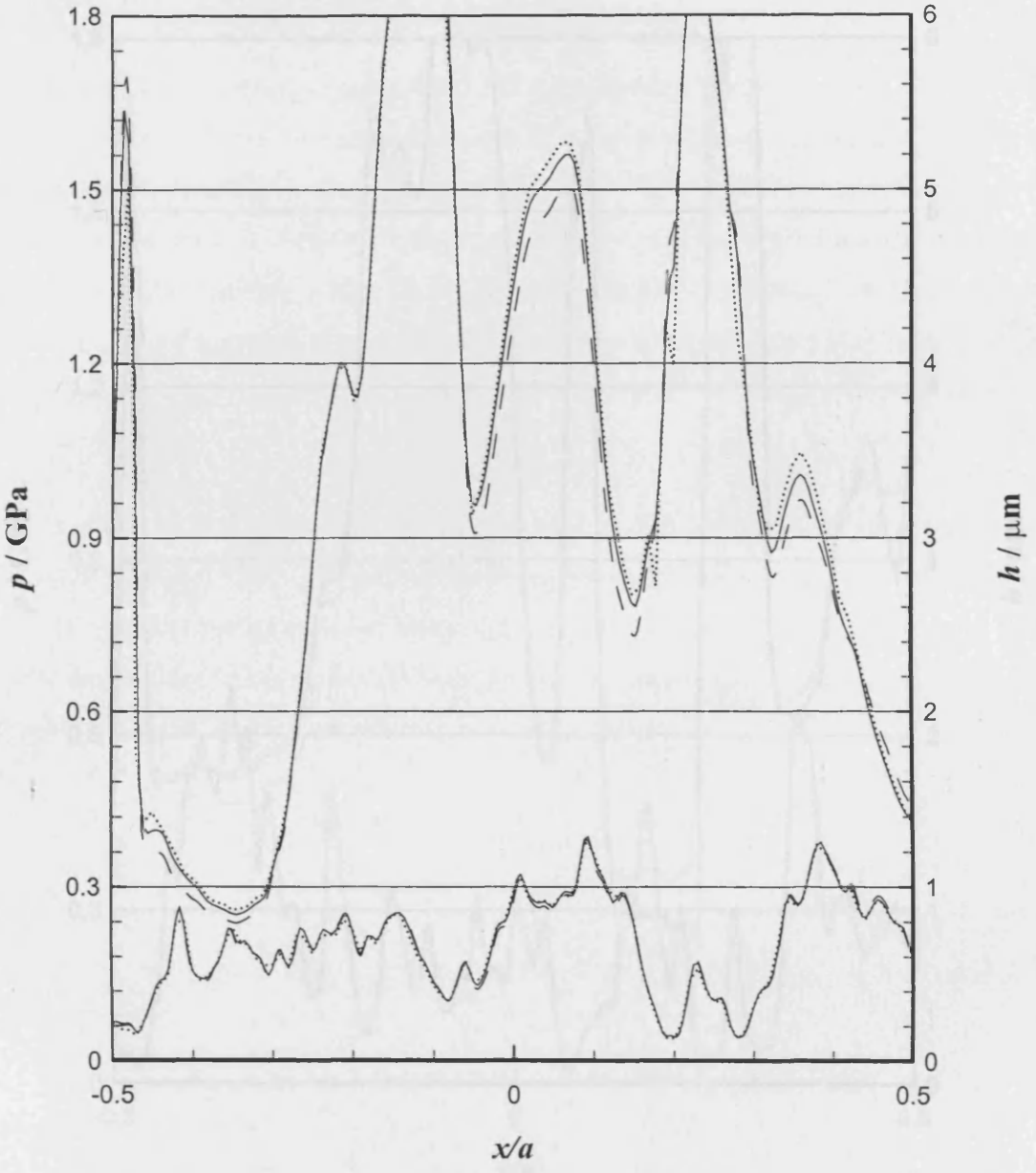


Figure 2.13: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (B) using different mesh size. Dashed lines are for $\Delta x = a/400$, solid lines are for $\Delta x = a/200$ and dotted lines are for $\Delta x = a/100$.

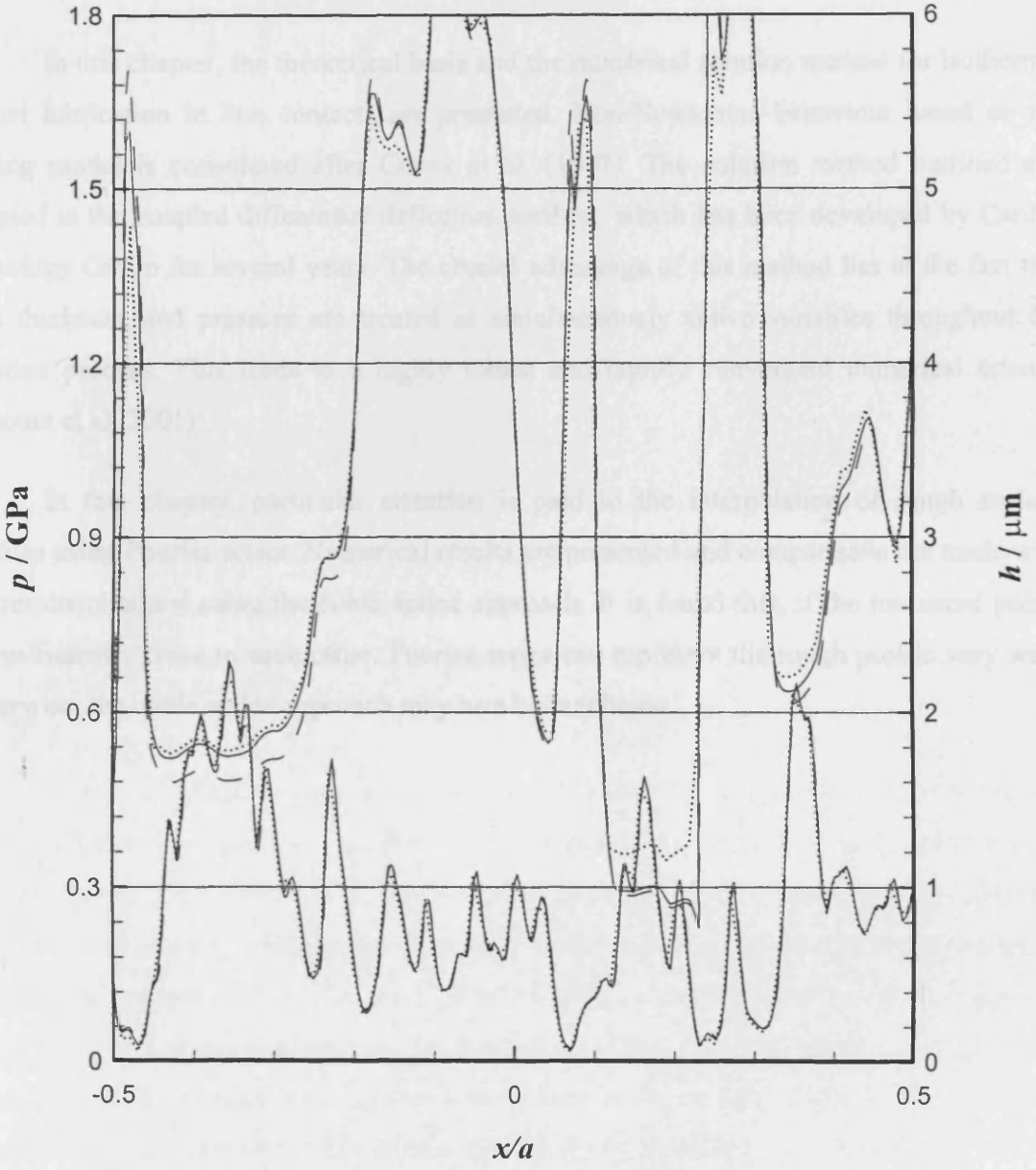


Figure 2.14: Pressure (upper curve) and film thickness distribution for the contact between two rough surfaces each with profile (C) using different mesh size. Dashed lines are for $\Delta x = a/400$, solid lines are for $\Delta x = a/200$ and dotted lines are for $\Delta x = a/100$.

2.5 Summary

In this chapter, the theoretical basis and the numerical solution method for isothermal mixed lubrication in line contacts are presented. Non-Newtonian behaviour based on the Eyring model is considered after Conry et al. (1987). The solution method outlined and adopted is the coupled differential deflection method, which has been developed by Cardiff Tribology Group for several years. The crucial advantage of this method lies in the fact that film thickness and pressure are treated as simultaneously active variables throughout the solution process. This leads to a highly robust and rapidly convergent numerical scheme (Elcoate et al. 2001).

In this chapter, particular attention is paid to the interpolation of rough surface profiles using Fourier series. Numerical results are presented and comparisons are made with the results obtained using the cubic spline approach. It is found that, if the measured points are sufficiently close to each other, Fourier series can represent the rough profile very well. Otherwise, the cubic spline approach may be a better choice.

Chapter 3

Thermal Analysis

3.1 Introduction

In Chapter 2, numerical formulations and solution methods for isothermal EHL with rough surfaces were presented. However, at high rolling speed and high slide roll ratio, thermal effects may become important, and the minimum film thickness obtained under isothermal conditions tends to be overestimated. This chapter will discuss thermal effects in EHL.

Many investigators have attempted to model EHL by considering both thermal and non-Newtonian fluid effects. The fluid rheological models introduced in Chapter 1 were incorporated in many thermal EHL analyses. The Ree-Eyring rheological model is probably the most widely used model in the literature; examples are found in the earlier works by Salehizadeh and Saka (1991), Sui and Sadeghi (1991), Wang et al. (1991, 1992), Yang and Wen (1992) and recent publications by Yang et al. (2004), Kazama (2004), etc. The Bair-Winer model is another well-known rheological model widely used in thermal EHL solutions, such as the work by Khonsari and Hua (1994), Ma (1997), etc. Sharif et al. (2001a) used a different form of non-Newtonian model, which includes both Eyring behaviour and a limiting shear stress behaviour, in their thermal non-Newtonian EHL analysis of worm gears. Other non-Newtonian rheological models incorporated in thermal EHL analyses include the circular fluid model (Lee & Hamrock 1990, Hsiao & Hamrock 1992, 1994), the power law fluid model (Prasad & Chhabra 1991, Prasad et al. 1993, Rodkiewicz & Yang 1995), etc. Yang & Wen (1990) derived a generalized Reynolds equation to incorporate different rheological models. Ma (1998) examined the influence of non-Newtonian behaviour of

lubricants on the temperature, film thickness and traction in an EHL contact using some well known rheological models. Kumar et al. (2002) investigated the behaviour of thermal EHL under dynamic loading and assumed the lubricant to be a mixture of Newtonian and Ree-Eyring fluids. Sharif et al. (2001b, 2004) compared results obtained using different non-Newtonian models with experiments and showed that the most encouraging agreement between experiment and theory under the range of conditions considered was obtained by adopting an Eyring type model.

Three methods, as described by Chang (1992), have been used to incorporate thermal effects into an EHL model. The simplest method uses the mean temperature across the lubricant film to evaluate the effective viscosity across the lubricant film. This method has been widely used due to its numerical efficiency and stability, such as the work by Salehizadeh and Saka (1991), Chang (1992), Hsiao and Hamrock (1992), Lee and Hsu (1993) and many of others. More accurate methods accommodate variation of viscosity across the lubricant film and, therefore, increase computational costs. Mihailidis et al. (1999) calculated the viscosity at every discrete point in the film as a function of the local temperature and pressure. Yang et al. (2001) solved the steady-state line contact thermal EHL problem using a multilevel method by taking account of the variations of viscosity and density across the film. Sharif et al. (2001b) considered both a mean temperature model and a local temperature model. In order to achieve the compromise between accuracy and efficiency, approximations have been made to account for variations of viscosity across the film. Wang et al. (1991) derived a thermal Reynolds equation, assuming that lubricant viscosity varies parabolically across the film, for example.

The solution of the energy equation is challenging and requires large computational time. Furthermore, numerical stability problems can arise considering the dependence of lubricant viscosity on pressure and temperature. To reduce the complexity of the thermal problem, Salehizadeh and Saka (1991) assumed a parabolic lubricant temperature profile across the film. Such an approximation was also employed by Lee et al. (1995), Kazama et al. (2001) and Kazama (2004). Pandey and Ghosh (1996, 1998) described the temperature distribution across the fluid film using a Legendre polynomial.

Most of the above work dealt with the steady-state thermal EHL problem with smooth surfaces. That is, the energy equation was solved for the film region bounded by the two smooth surfaces. The surface temperatures were taken as the boundary conditions, which can be obtained by a well-known integral equation derived by Carslaw and Jaeger (1959) for a moving heat source. This integral equation has been widely used in thermal EHL models to determine the steady-state surface temperature. Recently, some researchers have also used this integral equation to determine the transient flash temperatures for rough surface dry contacts (Gao et al. 2000, Liu et al. 2001b) and mixed lubricated contacts (Qiu & Cheng 1998, Liu et al. 2002, Wang et al. 2004b). To perform the numerical integration for temperature efficiently, the fast Fourier transform (FFT) technique has also been widely used. However, the use of this integral equation for transient thermal problems is impractical computationally if large timesteps are involved since the whole history of the heat source needs to be considered at each timestep. To overcome this, Chang and Farnum (1992) devised the thermal micro-EHL model by extending the thermal domain into the two bounding solids. In this way, the surface temperature is calculated in the thermal domain without resorting to the integral equation. Chang (1992) used this thermal model to study the traction behaviour of EHL line contacts between two rough surfaces. Later, Zhai and Chang (2000) extended this thermal model for mixed lubricated contacts and Chang and Zhao (2000) and Zhai and Chang (2001) applied the model to study asperity temperatures in mixed lubrication. Wang et al. (2004c) also presented a model which calculates temperatures by solving the transient energy equations in both the solids and the fluid without considering surface roughness.

In this chapter, a relatively complete thermal EHL model is presented based on the previous work by Sharif et al. (2001a) and Davies (2005). In this model, the method of domain extension into the bounding solids, as adopted by Chang and Farnum (1992) and Zhai and Chang (2000), is applied. This thermal model is then integrated into the isothermal micro or mixed EHL model presented in Chapter 2 to study the thermal effects due to shear heating for rough surface EHL line contact problems. Factors affecting temperature, pressure and film thickness in EHL will be investigated by numerical simulations.

3.2 Mathematical Modeling

The computing model for thermal EHL line contact problem is shown in Figure 3.1. The model consists of three domains, Ω, Ω_1, and Ω_2, which represent lubricant film, solid 1 and solid 2, respectively.

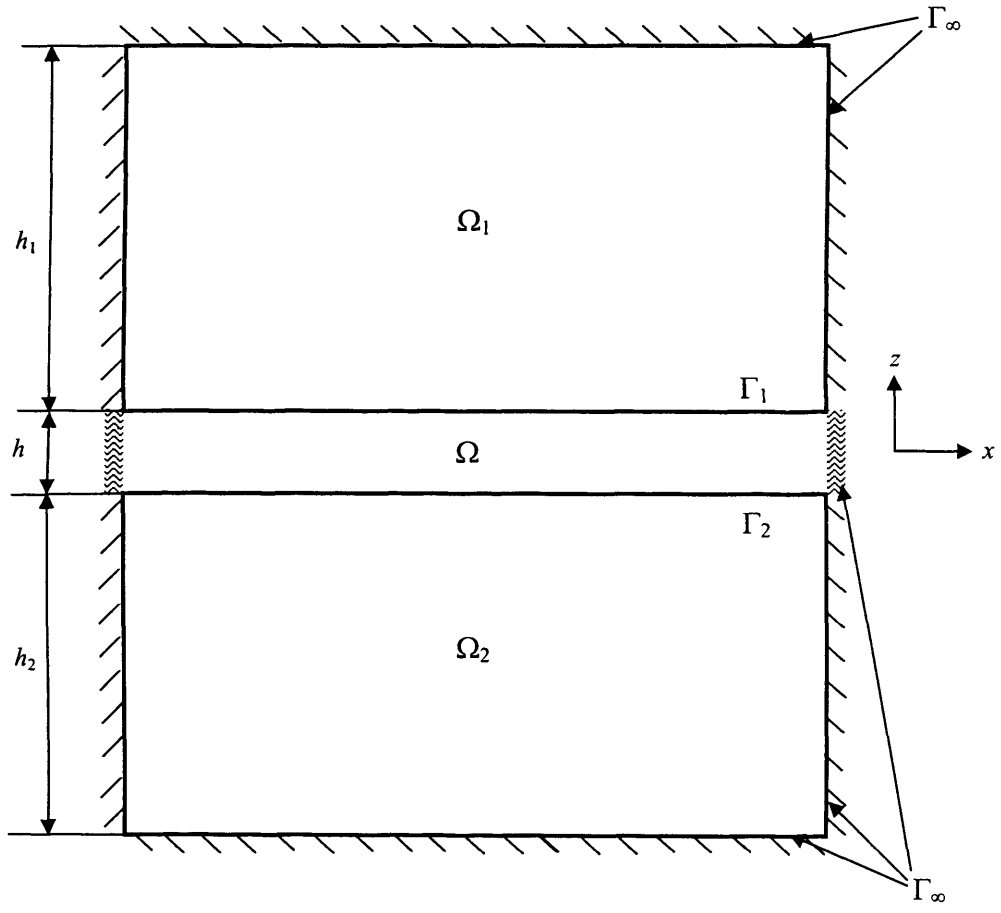


Figure 3.1: The computational domains for thermal analysis.

3.2.1 Energy Equations

The derivation of the energy equation is not given here and readers may refer to the book by Suryanarayana (1994) for details. The energy equations governing the temperature variations in these domains can be written as partial differential equations (PDE):

- **Lubricant film (Ω):**

$$\rho c\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}\right)=\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+\varepsilon T\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)+\tau\dot{\gamma} \tag{3.1}$$

where: ρ is density of lubricant
c is specific heat of lubricant
k is thermal conductivity of lubricant
ε is thermal expansion coefficient of lubricant
u is velocity of lubricant
τ is shear stress
$\dot{\gamma}$ is shear strain rate
t is time
T is temperature
p is fluid pressure

- **Solid 1 (Ω_1):**

$$\rho_1 c_1\left(\frac{\partial T}{\partial t}+u_1\frac{\partial T}{\partial x}\right)=\frac{\partial}{\partial x}\left(k_1\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial z}\left(k_1\frac{\partial T}{\partial z}\right) \tag{3.2}$$

where: ρ_1 is density of Solid 1
c_1 is specific heat of Solid 1
k_1 is thermal conductivity of Solid 1
u_1 is velocity of Solid 1

- **Solid 2 (Ω_2):**

$$\rho_2 c_2\left(\frac{\partial T}{\partial t}+u_2\frac{\partial T}{\partial x}\right)=\frac{\partial}{\partial x}\left(k_2\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial z}\left(k_2\frac{\partial T}{\partial z}\right) \tag{3.3}$$

where: ρ_2 is density of Solid 2

c_2 is specific heat of Solid 2

k_2 is thermal conductivity of Solid 2

u_2 is velocity of Solid 2

In these equations, the axis set is fixed relative to the contact position. The effects of changes in specific heat, thermal conductivity and thermal expansivity of lubricant are neglected and kept constant, and so are the properties of the solids. The density and viscosity of lubricant are expressed as functions of temperature and pressure.

3.2.2 Boundary Conditions

Besides the governing equations, the boundary conditions must be specified to solve the thermal EHL problem. The boundaries are defined as illustrated in Figure 3.1. Γ_1 is the interface between the film and Solid 1, and Γ_2 is the interface between the film and Solid 2. Γ_∞ is the outside boundary which is sufficiently far away from the contact area so that its temperature variation can be neglected. This boundary condition can be expressed as

$$T(x,z) = T_\infty, \quad (x,z) \in \Gamma_\infty \tag{3.4}$$

where T_∞ can take the value of the ambient temperature.

To determine the boundary Γ_∞ in the z-direction, the effective depth of heat penetration needs to be estimated so that the boundary positions h_1 and h_2 can be set below these depths. The temperature in the solid, as an approximation, is governed by the one-dimensional heat conduction equation:

$$\rho c \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial z^2} \tag{3.5}$$

Assuming the solid has zero initial temperature and a constant heat flux q_0 is applied at the surface $z = 0$, the following temperature distribution (temperature rise) is obtained for the solid (Carslaw & Jaeger 1959):

$$T(z,t) = \frac{2q_0}{k}\left\{\left(\frac{\alpha t}{\pi}\right)^{\frac{1}{2}} e^{-z^2/4\alpha t} - \frac{z}{2}\operatorname{erfc}\frac{z}{2\sqrt{\alpha t}}\right\} \tag{3.6}$$

where the diffusivity $\alpha = \dfrac{k}{\rho c}$.

From Equation (3.6), the temperature rise at $z = 0$ is given by

$$T_0 = \frac{2q_0}{k}\left(\frac{\alpha t}{\pi}\right)^{\frac{1}{2}} \tag{3.7}$$

We define H to be the depth where the temperature rise is $R\%$ of surface temperature rise, i.e.

$$T(H,t) = R\% \cdot T_0 \tag{3.8}$$

This gives the penetration depth H as a function of time t for a prescribed percentage $R\%$. Substituting Equations (3.6) and (3.7) into Equation (3.8) yields

$$\left(\frac{\alpha t}{\pi}\right)^{\frac{1}{2}}\left(e^{-H^2/4\alpha t} - R\%\right) - \frac{H}{2}\operatorname{erfc}\frac{H}{2\sqrt{\alpha t}} = 0 \tag{3.9}$$

This gives an H-t relationship for a specified $R\%$.

Figure 3.2 shows curves of H-t for different values of $R\%$ = 0.1%, 1%, 5%, 10%. The material parameters used for these curves are those for a typical steel namely, ρ = 7850 kg/m^3, k = 48 W/m·K and c = 450 J/kg·K. If the Hertzian contact semi-width a = 335μm and surface speed u = 25m/s, then the heat penetration depth during traverse of the contact area (i.e. $t = 2a/u$) is $H \approx 60$ μm for $R\%$ = 1%.

On the two film-solid interfaces, heat flux continuity needs to be satisfied and thus the following boundary conditions are established:

$$-k\frac{\partial T}{\partial z} = -k_1\frac{\partial T}{\partial z}, \quad z \in \Gamma_1 \tag{3.10a}$$

$$-k\frac{\partial T}{\partial z} = -k_2\frac{\partial T}{\partial z}, \quad z \in \Gamma_2 \tag{3.10b}$$

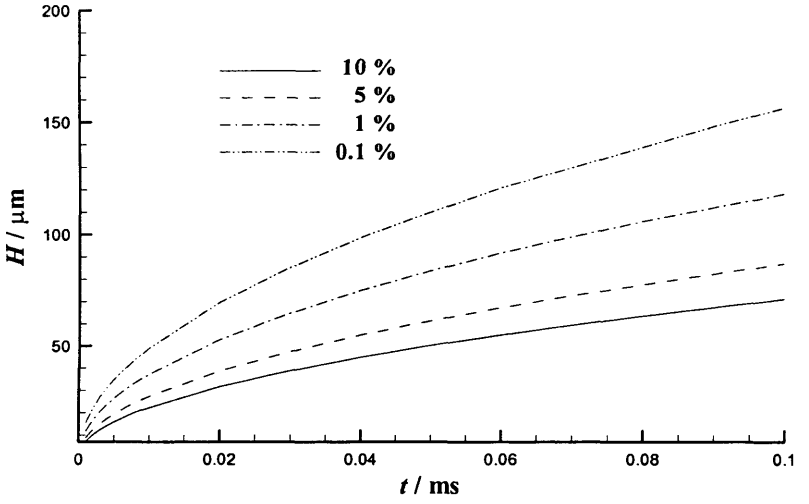


Figure 3.2: An example of heat penetration curves for different percentage values.

3.3 Numerical Formulations

3.3.1 Control Volume Formulations

The energy equations are discretised using the finite difference method based on the control volume (finite volume) approach. The problem domain is to be divided up into control volumes with associated grid points at the centres of the volumes. An example of the control volume for a point (x_i, z_j) is shown in Figure 3.3.

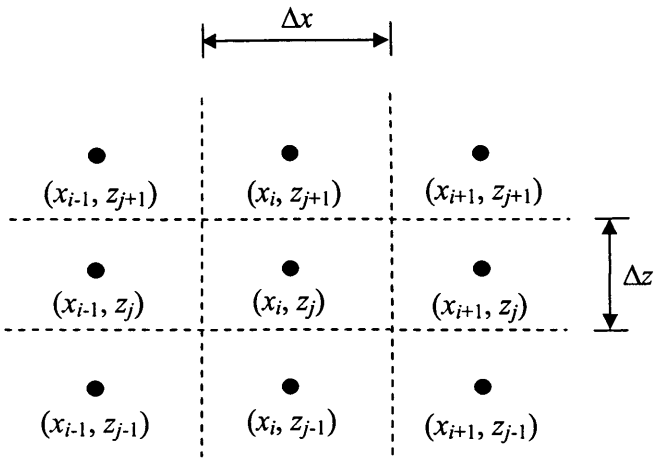


Figure 3.3: The control volumes in the domain.

According to the location of the point (x_i, z_j), there are six cases to be considered in the formulations as follows.

(1) $(x_i, z_j) \in \Omega$

Consider a control volume for a point (x_i, z_j) in the film region, as shown in Figure 3.3. Integrating the PDE for the film, i.e. Equation (3.1), over this control volume gives

$$\rho c\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}\right)\Delta x\Delta z = -\Delta z k\frac{\partial T}{\partial x}\bigg|_{i-\frac{1}{2},j} + \Delta z k\frac{\partial T}{\partial x}\bigg|_{i+\frac{1}{2},j} - \Delta x k\frac{\partial T}{\partial z}\bigg|_{i,j-\frac{1}{2}} + \Delta x k\frac{\partial T}{\partial z}\bigg|_{i,j+\frac{1}{2}} + \varepsilon T\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)\Delta x\Delta z + \tau\dot{\gamma}\Delta x\Delta z \tag{3.11}$$

The ½ in the subscripts refers to evaluation at the boundaries of the control volume, which is half-way between mesh points.

The time-dependent temperature term is formulated using the Crank-Nicolson scheme (Tannehill et al. 1997), i.e.

$$\frac{\partial T}{\partial t} = \frac{1}{2}\left[\left.\frac{\partial T}{\partial t}\right|^{n+1} + \left.\frac{\partial T}{\partial t}\right|^{n}\right] \tag{3.12}$$

where n is the timestep. The spatial derivatives are approximated by finite difference scheme. For the convection term (i.e. $\rho cu\partial T/\partial x$), if $u > 0$ the backward difference is used; otherwise (reverse flow), the forward difference is used. For other spatial terms (conduction in both x and z directions), the central difference is used. Thus Equation (3.11) can be written as

$$\begin{aligned} 2\rho c\frac{T_{i,j}^{n+1}-T_{i,j}^{n}}{\Delta t} = &\left\{\begin{array}{l} \dfrac{k}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}-2T_{i,j}^{n+1}\right)+\dfrac{k}{\Delta z^2}\left(T_{i,j-1}^{n+1}+T_{i,j+1}^{n+1}-2T_{i,j}^{n+1}\right) \\ +\varepsilon\left(\dfrac{\partial p}{\partial t}+u\dfrac{\partial p}{\partial x}\right)\left|^{n+1}\right. T_{i,j}^{n+1}+\tau\dot{\gamma}\left|^{n+1}\right. - \rho cu\dfrac{T_{i,j}^{n+1}-T_{i-1,j}^{n+1}}{\Delta x} \end{array}\right\} \\ &+\left\{\begin{array}{l} \dfrac{k}{\Delta x^2}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)+\dfrac{k}{\Delta z^2}\left(T_{i,j-1}^{n}+T_{i,j+1}^{n}-2T_{i,j}^{n}\right) \\ +\varepsilon\left(\dfrac{\partial p}{\partial t}+u\dfrac{\partial p}{\partial x}\right)\left|^{n}\right. T_{i,j}^{n}+\tau\dot{\gamma}\left|^{n}\right. - \rho cu\dfrac{T_{i,j}^{n}-T_{i-1,j}^{n}}{\Delta x} \end{array}\right\} \end{aligned} \tag{3.13}$$

Equation (3.13) can be rearranged in terms of unknowns as

$$L_\Omega T_{i,j-1}^{n+1} + D_\Omega T_{i,j}^{n+1} + U_\Omega T_{i,j+1}^{n+1} = R_\Omega \tag{3.14}$$

where

$$D_\Omega = \frac{2\rho c}{\Delta t} + \frac{\rho cu}{\Delta x} + \frac{2k}{\Delta x^2} + \frac{2k}{\Delta z^2} - \varepsilon\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)\Bigg|^{n+1} \tag{3.15a}$$

$$L_\Omega = -\frac{k}{\Delta z^2} \tag{3.15b}$$

$$U_\Omega = -\frac{k}{\Delta z^2} \tag{3.15c}$$

$$R_\Omega = \frac{k}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\tau\dot{\gamma}\left|^{n+1}\right. + \frac{\rho cu}{\Delta x}T_{i-1,j}^{n+1} + R_\Omega^{n} \tag{3.15d}$$

$$R_{\Omega}^{n}=\left\{\begin{array}{l}\dfrac{k}{\Delta x^{2}}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)+\dfrac{k}{\Delta z^{2}}\left(T_{i,j-1}^{n}+T_{i,j+1}^{n}-2T_{i,j}^{n}\right)\\ +\varepsilon\left(\dfrac{\partial p}{\partial t}+u\dfrac{\partial p}{\partial x}\right)\Big|^{n}T_{i,j}^{n}+\tau\dot{\gamma}\Big|^{n}-\rho cu\dfrac{T_{i,j}^{n}-T_{i-1,j}^{n}}{\Delta x}+\dfrac{2\rho c}{\Delta t}T_{i,j}^{n}\end{array}\right\} \tag{3.15e}$$

(2) $(x_i, z_j) \in \Omega_1$

If point (x_i, z_j) is located in Solid 1, the differential equation is Equation (3.2) which does not include the compressive heating and shear heating terms and where the material velocity is constant. A similar equation to Equation (3.14) can be obtained in the same way for Solid 1:

$$L_{\Omega_1}T_{i,j-1}^{n+1}+D_{\Omega_1}T_{i,j}^{n+1}+U_{\Omega_1}T_{i,j+1}^{n+1}=R_{\Omega_1} \tag{3.16}$$

where

$$D_{\Omega_1}=\frac{2\rho_1 c_1}{\Delta t}+\frac{\rho_1 c_1 u_1}{\Delta x}+\frac{2k_1}{\Delta x^{2}}+\frac{2k_1}{\Delta z_1^{2}} \tag{3.17a}$$

$$L_{\Omega_1}=-\frac{k_1}{\Delta z_1^{2}} \tag{3.17b}$$

$$U_{\Omega_1}=-\frac{k_1}{\Delta z_1^{2}} \tag{3.17c}$$

$$R_{\Omega_1}=\frac{k_1}{\Delta x^{2}}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\frac{\rho_1 c_1 u_1}{\Delta x}T_{i-1,j}^{n+1}+R_{\Omega_1}^{n} \tag{3.17d}$$

$$R_{\Omega_1}^{n}=\left\{\begin{array}{l}\dfrac{k_1}{\Delta x^{2}}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)+\dfrac{k_1}{\Delta z_1^{2}}\left(T_{i,j-1}^{n}+T_{i,j+1}^{n}-2T_{i,j}^{n}\right)\\ -\rho_1 c_1 u_1\dfrac{T_{i,j}^{n}-T_{i-1,j}^{n}}{\Delta x}+\dfrac{2\rho_1 c_1}{\Delta t}T_{i,j}^{n}\end{array}\right\} \tag{3.17e}$$

(3) $(x_i, z_j) \in \Omega_2$

Similarly, if point (x_i, z_j) is located in Solid 2, the following equation can be obtained:

$$L_{\Omega_2}T_{i,j-1}^{n+1}+D_{\Omega_2}T_{i,j}^{n+1}+U_{\Omega_2}T_{i,j+1}^{n+1}=R_{\Omega_2} \tag{3.18}$$

where

$$D_{\Omega_2} = \frac{2\rho_2 c_2}{\Delta t} + \frac{\rho_2 c_2 u_2}{\Delta x} + \frac{2k_2}{\Delta x^2} + \frac{2k_2}{\Delta z_2^{\,2}} \tag{3.19a}$$

$$L_{\Omega_2} = -\frac{k_2}{\Delta z_2^{\,2}} \tag{3.19b}$$

$$U_{\Omega_2} = -\frac{k_2}{\Delta z_2^{\,2}} \tag{3.19c}$$

$$R_{\Omega_2} = \frac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \frac{\rho_2 c_2 u_2}{\Delta x} T_{i-1,j}^{n+1} + R_{\Omega_2}^{n} \tag{3.19d}$$

$$R_{\Omega_2}^{n} = \begin{Bmatrix} \dfrac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n} + T_{i+1,j}^{n} - 2T_{i,j}^{n}\right) + \dfrac{k_2}{\Delta z_2^{\,2}}\left(T_{i,j-1}^{n} + T_{i,j+1}^{n} - 2T_{i,j}^{n}\right) \\ -\rho_2 c_2 u_2 \dfrac{T_{i,j}^{n} - T_{i-1,j}^{n}}{\Delta x} + \dfrac{2\rho_2 c_2}{\Delta t} T_{i,j}^{n} \end{Bmatrix} \tag{3.19e}$$

(4) $(x_i, z_j) \in \Gamma_1$

For a point (x_i, z_j) on the boundary, the control volume consists of half fluid and half solid. Figure 3.4 shows the control volume on the boundary Γ_1. Integrating the PDEs over the control volume gives

$$\begin{aligned} &\rho c\left(\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x}\right)\Delta x\frac{\Delta z}{2} + \rho_1 c_1\left(\frac{\partial T}{\partial t} + u_1\frac{\partial T}{\partial x}\right)\Delta x\frac{\Delta z_1}{2} \\ &= \left\{-\frac{\Delta z}{2}k\frac{\partial T}{\partial x}\bigg|_{i-\frac{1}{2},j} + \frac{\Delta z}{2}k\frac{\partial T}{\partial x}\bigg|_{i+\frac{1}{2},j} - \Delta x k\frac{\partial T}{\partial z}\bigg|_{i,j-\frac{1}{2}} + \varepsilon T\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x}\right)\Delta x\frac{\Delta z}{2} + \tau\dot{\gamma}\Delta x\frac{\Delta z}{2}\right\} \\ &+ \left\{-\frac{\Delta z_1}{2}k_1\frac{\partial T}{\partial x}\bigg|_{i-\frac{1}{2},j} + \frac{\Delta z_1}{2}k_1\frac{\partial T}{\partial x}\bigg|_{i+\frac{1}{2},j} + \Delta x k_1\frac{\partial T}{\partial z}\bigg|_{i,j+\frac{1}{2}}\right\} \end{aligned} \tag{3.20}$$

Applying the Crank-Nicolson scheme, Equation (3.20) can be written as the following finite difference equation:

$$
\begin{aligned}
&2\left(\rho c\frac{\Delta z}{2}+\rho_1 c_1\frac{\Delta z_1}{2}\right)\frac{T_{i,j}^{n+1}-T_{i,j}^{n}}{\Delta t}\\
&=\frac{\Delta z}{2}\left\{\begin{aligned}&\frac{k}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}-2T_{i,j}^{n+1}\right)-\frac{2k}{\Delta z^2}\left(T_{i,j}^{n+1}-T_{i,j-1}^{n+1}\right)\\&+\varepsilon\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)\Bigg|^{n+1}T_{i,j}^{n+1}+\tau\dot{\gamma}\Big|^{n+1}-\frac{\rho c u}{\Delta x}\left(T_{i,j}^{n+1}-T_{i-1,j}^{n+1}\right)\end{aligned}\right\}\\
&+\frac{\Delta z_1}{2}\left\{\frac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}-2T_{i,j}^{n+1}\right)+\frac{2k_1}{\Delta z_1^{\,2}}\left(T_{i,j+1}^{n+1}-T_{i,j}^{n+1}\right)-\frac{\rho_1 c_1 u_1}{\Delta x}\left(T_{i,j}^{n+1}-T_{i-1,j}^{n+1}\right)\right\}\\
&+\frac{\Delta z}{2}R_{F_1}^{n}+\frac{\Delta z_1}{2}R_{S_1}^{n}
\end{aligned}
\tag{3.21}
$$

where

$$
R_{F_1}^{n}=\left\{\begin{aligned}&\frac{k}{\Delta x^2}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)+\frac{2k}{\Delta z^2}\left(T_{i,j-1}^{n}-T_{i,j}^{n}\right)\\&+\varepsilon\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)\Bigg|^{n}T_{i,j}^{n}+\tau\dot{\gamma}\Big|^{n}-\frac{\rho c u}{\Delta x}\left(T_{i,j}^{n}-T_{i-1,j}^{n}\right)+\frac{2\rho c}{\Delta t}T_{i,j}^{n}\end{aligned}\right\}
\tag{3.22a}
$$

$$
R_{S_1}^{n}=\left\{\begin{aligned}&\frac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)+\frac{2k_1}{\Delta z_1^{\,2}}\left(T_{i,j+1}^{n}-T_{i,j}^{n}\right)\\&-\frac{\rho_1 c_1 u_1}{\Delta x}\left(T_{i,j}^{n}-T_{i-1,j}^{n}\right)+\frac{2\rho_1 c_1}{\Delta t}T_{i,j}^{n}\end{aligned}\right\}
\tag{3.22b}
$$

In the same way as before, Equation (3.21) can be rearranged as

$$
L_{\Gamma_1}T_{i,j-1}^{n+1}+D_{\Gamma_1}T_{i,j}^{n+1}+U_{\Gamma_1}T_{i,j+1}^{n+1}=R_{\Gamma_1}
\tag{3.23}
$$

where

$$
D_{\Gamma_1}=\left\{\begin{aligned}&\frac{\Delta z}{2}\left[\frac{2\rho c}{\Delta t}+\frac{\rho c u}{\Delta x}+\frac{2k}{\Delta x^2}+\frac{2k}{\Delta z^2}-\varepsilon\left(\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial x}\right)\Bigg|^{n+1}\right]\\&+\frac{\Delta z_1}{2}\left[\frac{2\rho_1 c_1}{\Delta t}+\frac{\rho_1 c_1 u_1}{\Delta x}+\frac{2k_1}{\Delta x^2}+\frac{2k_1}{\Delta z_1^{\,2}}\right]\end{aligned}\right\}
\tag{3.24a}
$$

$$
L_{\Gamma_1}=-\frac{k}{\Delta z}
\tag{3.24b}
$$

$$U_{\Gamma_1} = -\frac{k_1}{\Delta z_1} \tag{3.24c}$$

$$R_{\Gamma_1} = \left\{ \begin{array}{l} \frac{\Delta z}{2}\left[\frac{k}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \tau\dot{\gamma}\Big|^{n+1} + \frac{\rho c u}{\Delta x} T_{i-1,j}^{n+1} + R_{F_1}^{n}\right] \\ + \frac{\Delta z_1}{2}\left[\frac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \frac{\rho_1 c_1 u_1}{\Delta x} T_{i-1,j}^{n+1} + R_{S_1}^{n}\right] \end{array} \right\} \tag{3.24d}$$

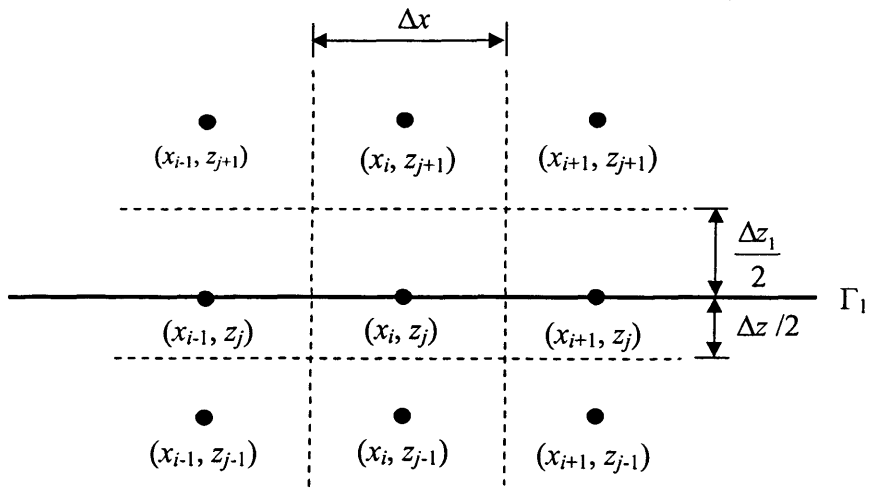


Figure 3.4: The control volumes on the boundary.

(5) $(x_i, z_j) \in \Gamma_2$

Similarly, the equation for the point (x_i, z_j) on the boundary Γ_2 is

$$L_{\Gamma_2} T_{i,j-1}^{n+1} + D_{\Gamma_2} T_{i,j}^{n+1} + U_{\Gamma_2} T_{i,j+1}^{n+1} = R_{\Gamma_2} \tag{3.25}$$

where

$$D_{\Gamma_2} = \left\{ \begin{array}{l} \frac{\Delta z}{2}\left[\frac{2\rho c}{\Delta t} + \frac{\rho c u}{\Delta x} + \frac{2k}{\Delta x^2} + \frac{2k}{\Delta z^2} - \varepsilon\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x}\right)\Big|^{n+1}\right] \\ + \frac{\Delta z_2}{2}\left[\frac{2\rho_2 c_2}{\Delta t} + \frac{\rho_2 c_2 u_2}{\Delta x} + \frac{2k_2}{\Delta x^2} + \frac{2k_2}{\Delta z_2^{\,2}}\right] \end{array} \right\} \tag{3.26a}$$

$$L_{\Gamma_2} = -\frac{k_2}{\Delta z_2} \tag{3.26b}$$

$$U_{\Gamma_2} = -\frac{k}{\Delta z} \tag{3.26c}$$

$$R_{\Gamma_2} = \left\{ \begin{array}{l} \frac{\Delta z}{2}\left[\frac{k}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \tau\dot{\gamma}\Big|^{n+1} + \frac{\rho c u}{\Delta x} T_{i-1,j}^{n+1} + R_{F_2}^{n}\right] \\ + \frac{\Delta z_2}{2}\left[\frac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \frac{\rho_2 c_2 u_2}{\Delta x} T_{i-1,j}^{n+1} + R_{S_2}^{n}\right] \end{array} \right\} \tag{3.26d}$$

$$R_{F_2}^{n} = \left\{ \begin{array}{l} \frac{k}{\Delta x^2}\left(T_{i-1,j}^{n} + T_{i+1,j}^{n} - 2T_{i,j}^{n}\right) + \frac{2k}{\Delta z^2}\left(T_{i,j+1}^{n} - T_{i,j}^{n}\right) \\ + \varepsilon\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x}\right)\Big|^{n} T_{i,j}^{n} + \tau\dot{\gamma}\Big|^{n} - \frac{\rho c u}{\Delta x}\left(T_{i,j}^{n} - T_{i-1,j}^{n}\right) + \frac{2\rho c}{\Delta t} T_{i,j}^{n} \end{array} \right\} \tag{3.26e}$$

$$R_{S_2}^{n} = \left\{ \begin{array}{l} \frac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n} + T_{i+1,j}^{n} - 2T_{i,j}^{n}\right) + \frac{2k_2}{\Delta z_2^{\,2}}\left(T_{i,j-1}^{n} - T_{i,j}^{n}\right) \\ - \frac{\rho_2 c_2 u_2}{\Delta x}\left(T_{i,j}^{n} - T_{i-1,j}^{n}\right) + \frac{2\rho_2 c_2}{\Delta t} T_{i,j}^{n} \end{array} \right\} \tag{3.26f}$$

(6) $\Gamma_1 = \Gamma_2$

For the special case where direct contact occurs between the two solids, i.e. $\Gamma_1 = \Gamma_2$, the fluid film disappears and simple dry frictional heating at the surface between two solids needs to be taken into account. The control volume at the interface is illustrated in Figure 3.5. It should be noted that direct contact may not occur at the left and/or right side neighbours. If this is the case, point $(i\text{-}1, j)$ is two surface points and/or so is point $(i\text{+}1, j)$. Conservation of energy in the control volume then gives

$$\rho_1 c_1\left(\frac{\partial T}{\partial t}+u_1\frac{\partial T}{\partial x}\right)\Delta x\frac{\Delta z_1}{2}+\rho_2 c_2\left(\frac{\partial T}{\partial t}+u_2\frac{\partial T}{\partial x}\right)\Delta x\frac{\Delta z_2}{2}$$
$$=\left\{-\frac{\Delta z_1}{2}k_1\left.\frac{\partial T}{\partial x}\right|_{i-\frac{1}{2},j}+\frac{\Delta z_1}{2}k_1\left.\frac{\partial T}{\partial x}\right|_{i+\frac{1}{2},j}+\Delta x k_1\left.\frac{\partial T}{\partial z}\right|_{i,j+\frac{1}{2}}\right\} \tag{3.27}$$
$$+\left\{-\frac{\Delta z_2}{2}k_2\left.\frac{\partial T}{\partial x}\right|_{i-\frac{1}{2},j}+\frac{\Delta z_2}{2}k_2\left.\frac{\partial T}{\partial x}\right|_{i+\frac{1}{2},j}-\Delta x k_2\left.\frac{\partial T}{\partial z}\right|_{i,j-\frac{1}{2}}\right\}+\mu p|u_1-u_2|\Delta x$$

where μ is the coefficient of friction between the solids. The above equation can also be discretised using the Crank-Nicolson approach and rearranged in the form

$$L_{\Gamma_1=\Gamma_2}T_{i,j-1}^{n+1}+D_{\Gamma_1=\Gamma_2}T_{i,j}^{n+1}+U_{\Gamma_1=\Gamma_2}T_{i,j+1}^{n+1}=R_{\Gamma_1=\Gamma_2} \tag{3.28}$$

where

$$D_{\Gamma_1=\Gamma_2}=\left\{\begin{array}{l}\frac{\Delta z_1}{2}\left[\frac{2\rho_1 c_1}{\Delta t}+\frac{\rho_1 c_1 u_1}{\Delta x}+\frac{2k_1}{\Delta x^2}+\frac{2k_1}{\Delta {z_1}^2}\right]+\\ \frac{\Delta z_2}{2}\left[\frac{2\rho_2 c_2}{\Delta t}+\frac{\rho_2 c_2 u_2}{\Delta x}+\frac{2k_2}{\Delta x^2}+\frac{2k_2}{\Delta {z_2}^2}\right]\end{array}\right\} \tag{3.29a}$$

$$L_{\Gamma_1=\Gamma_2}=-\frac{k_2}{\Delta z_2} \tag{3.29b}$$

$$U_{\Gamma_1=\Gamma_2}=-\frac{k_1}{\Delta z_1} \tag{3.29c}$$

$$R_{\Gamma_1=\Gamma_2}=\left\{\begin{array}{l}\frac{\Delta z_1}{2}\left[\frac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\frac{\rho_1 c_1 u_1}{\Delta x}T_{i-1,j}^{n+1}+R_{S_1}^{n}\right]+\\ \frac{\Delta z_2}{2}\left[\frac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\frac{\rho_2 c_2 u_2}{\Delta x}T_{i-1,j}^{n+1}+R_{S_2}^{n}\right]+\mu p|u_1-u_2|\end{array}\right\} \tag{3.29d}$$

In Equation (3.29d), $R_{S_1}^n$ and $R_{S_2}^n$ are defined by Equations (3.22b) and (3.26f).

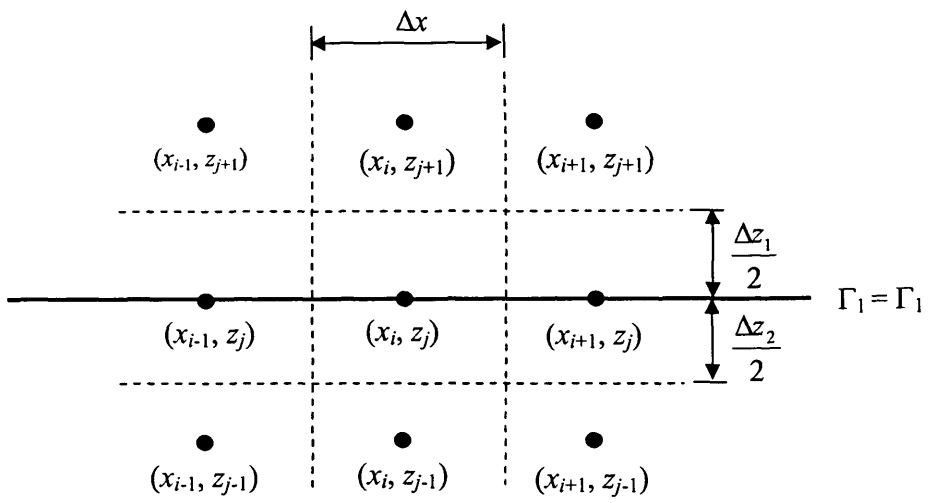


Figure 3.5: The control volumes on the boundary where direct contact occurs.

Assembly of all the equations for points in the *i*th column, i.e. Equations (3.14), (3.16), (3.18), (3.23) and (3.25) (or Equation (3.28) for a special case), leads to a system of equations in matrix form:

$$\begin{bmatrix} \ddots & \ddots & & & & \\ \ddots & \ddots & \ddots & & 0 & \\ & \ddots & \ddots & \ddots & & \\ & & L & D & U & \\ & & & \ddots & \ddots & \ddots \\ & 0 & & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \vdots \\ T_{i,j-1} \\ T_{i,j} \\ T_{i,j+1} \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ R \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} \tag{3.30}$$

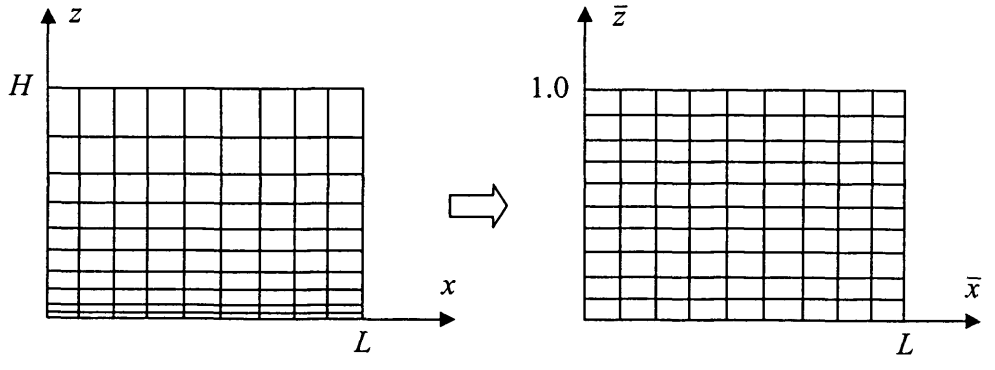
The coefficient matrix in Equation (3.30) is a tridiagonal matrix. For this tridiagonal matrix system, a technique called the Thomas algorithm (Tannehill et al. 1997), which modifies the Gaussian elimination procedure to take advantage of the zeros in the matrix of coefficients, can be used to solve the system very rapidly. In this way, the temperature fields in the three domains are determined simultaneously column by column by applying the boundary condition of Equation (3.4). The film-solid boundary conditions (Equations (3.10)) are not applied explicitly because they are already incorporated in the formulations.

The R terms on the right hand side of the equation contain temperatures at the new time step (n+1) for the upstream and downstream columns (i-1 and i+1). These have a minor influence on the energy balance as the predominant heat transfer mechanism is conduction perpendicular to the film.

To solve the thermal problem for a given EHL pressure / film thickness solution, the column by column scheme would need to be applied iteratively with the upstream temperatures taken from the current cycle and the downstream ones from the previous cycle. This is not the task being undertaken, however, as the temperature distribution within the lubricant is part of the EHL problem so that the equations need to be solved simultaneously as discussed in section 3.3.3.

3.3.2 Domain Transformation for Solids

All the above formulations are based on the uniformly-spaced grids in each domain. For the solids, usually a larger number of grid points are necessary due to the larger size of domain. To reduce the computation cost without loss of accuracy, a simple stretching transformation for solids can be used to cluster grid points in regions of large temperature gradients, i.e. near the surfaces Γ_1 and Γ_2. This strategy is illustrated in Figure 3.6.



(a) Physical domain (x, z) (b) Computational domain ($\bar{x}$, $\bar{z}$)

Figure 3.6: Domain transformation for the solids.

An example of such a transformation can be defined as (Tannehill et al. 1997)

$$\begin{cases} \bar{x} = x \\ \bar{z} = 1 - \dfrac{\ln\{(\beta + 1 - z/H)/(\beta - 1 + z/H)\}}{\ln\{(\beta + 1)/(\beta - 1)\}} \end{cases} \tag{3.31}$$

where the stretching parameter $1 < \beta < \infty$. The inverse of the transformation can be found from Equation (3.31) as

$$\begin{cases} x = \bar{x} \\ z = H \dfrac{(\beta + 1) - (\beta - 1)\{(\beta + 1)/(\beta - 1)\}^{1-\bar{z}}}{\{(\beta + 1)/(\beta - 1)\}^{1-\bar{z}} + 1} \end{cases} \tag{3.32}$$

This transformation clusters more points near $z = 0$ as β approaches 1, as shown in Figure 3.7 where different β values are used to perform the above transformation with $H = 50$. The number of mesh points is 11 in this demonstration.

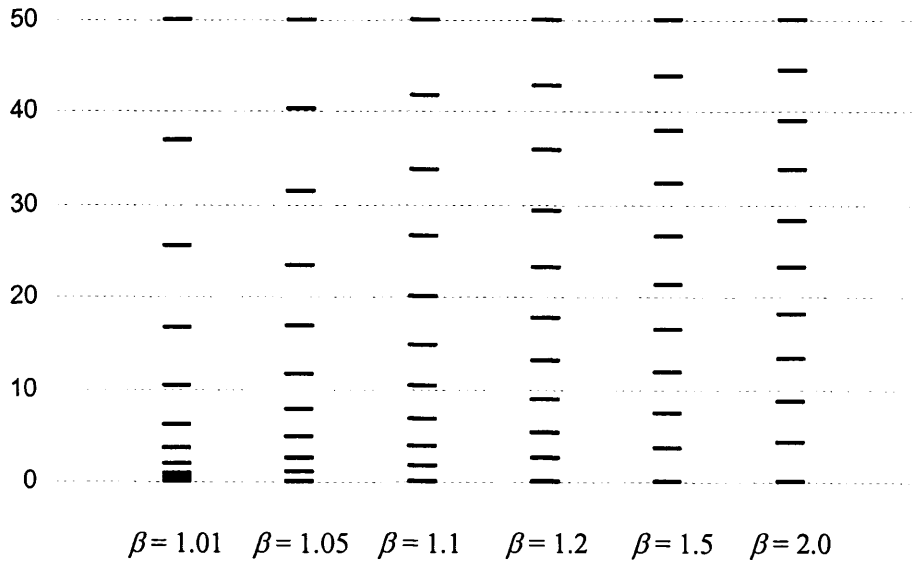


Figure 3.7: Mesh grids after domain transformation.

In order to apply this transformation to the governing equations of the solids, i.e. Equations (3.2) and (3.3), the partial derivatives with regard to z need to be transformed in the following way:

$$\frac{\partial}{\partial z}\left(\frac{\partial T}{\partial z}\right)=\frac{\partial}{\partial z}\left(\frac{\partial T}{\partial \bar{z}}\cdot\frac{\partial \bar{z}}{\partial z}\right)=\frac{\partial^2 \bar{z}}{\partial z^2}\cdot\frac{\partial T}{\partial \bar{z}}+\frac{\partial \bar{z}}{\partial z}\cdot\frac{\partial}{\partial z}\left(\frac{\partial T}{\partial \bar{z}}\right)=\frac{\partial^2 \bar{z}}{\partial z^2}\cdot\frac{\partial T}{\partial \bar{z}}+\left(\frac{\partial \bar{z}}{\partial z}\right)^2\frac{\partial^2 T}{\partial \bar{z}^2} \tag{3.33}$$

where

$$\frac{\partial \bar{z}}{\partial z}=\frac{2\beta}{H\left\{\beta^2-(1-z/H)^2\right\}\ln\{(\beta+1)/(\beta-1)\}} \tag{3.34a}$$

$$\frac{\partial^2 \bar{z}}{\partial z^2}=\frac{\partial}{\partial z}\left(\frac{\partial \bar{z}}{\partial z}\right)=\frac{-4\beta(1-z/H)}{H^2\left\{\beta^2-(1-z/H)^2\right\}^2\ln\{(\beta+1)/(\beta-1)\}} \tag{3.34b}$$

After transformation, the governing equation can be differenced again on the uniformly spaced grid in the computational domain. With this transformation, modifications are needed for the formulations where mesh points are located in the solids or at the interface.

If a point (i, j) is in the solids, an additional term, $\frac{\partial T}{\partial \bar{z}}$, needs to be formulated. For Solid 1, $\frac{\partial T}{\partial \bar{z}}=\frac{T_{i,j+1}-T_{i,j}}{\Delta \bar{z}_1}$; for Solid 2, $\frac{\partial T}{\partial \bar{z}}=\frac{T_{i,j-1}-T_{i,j}}{\Delta \bar{z}_2}$. Thus, the coefficients for Solid 1, i.e. Equations (3.17), become

$$D_{\Omega_1}=\frac{2\rho_1 c_1}{\Delta t}+\frac{\rho_1 c_1 u_1}{\Delta x}+\frac{2k_1}{\Delta x^2}+\frac{2k_1}{\Delta \bar{z}_1^{\,2}}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right)^2+\frac{k_1}{\Delta \bar{z}_1}\left(\frac{\partial^2 \bar{z}_1}{\partial z_1^{\,2}}\right) \tag{3.35a}$$

$$L_{\Omega_1}=-\frac{k_1}{\Delta \bar{z}_1^{\,2}}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right)^2 \tag{3.35b}$$

$$U_{\Omega_1}=-\frac{k_1}{\Delta \bar{z}_1^{\,2}}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right)^2-\frac{k_1}{\Delta \bar{z}_1}\left(\frac{\partial^2 \bar{z}_1}{\partial z_1^{\,2}}\right) \tag{3.35c}$$

$$R_{\Omega_1} = \frac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \frac{\rho_1 c_1 u_1}{\Delta x} T_{i-1,j}^{n+1} + R_{\Omega_1}^{n} \tag{3.35d}$$

$$R_{\Omega_1}^{n} = \left\{ \begin{array}{l} \dfrac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n} + T_{i+1,j}^{n} - 2T_{i,j}^{n}\right) - \rho_1 c_1 u_1 \dfrac{T_{i,j}^{n} - T_{i-1,j}^{n}}{\Delta x} + \dfrac{2\rho_1 c_1}{\Delta t} T_{i,j}^{n} \\ + \dfrac{k_1}{\Delta \bar{z}_1^{\,2}}\left(\dfrac{\partial \bar{z}_1}{\partial z_1}\right)^2 \left(T_{i,j-1}^{n} + T_{i,j+1}^{n} - 2T_{i,j}^{n}\right) + \dfrac{k_1}{\Delta \bar{z}_1}\left(\dfrac{\partial^2 \bar{z}_1}{\partial z_1^{\,2}}\right)\left(T_{i,j+1}^{n} - T_{i,j}^{n}\right) \end{array} \right\} \tag{3.35e}$$

The coefficients for Solid 2, i.e. Equations (3.19), become

$$D_{\Omega_2} = \frac{2\rho_2 c_2}{\Delta t} + \frac{\rho_2 c_2 u_2}{\Delta x} + \frac{2k_2}{\Delta x^2} + \frac{2k_2}{\Delta \bar{z}_2^{\,2}}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right)^2 + \frac{k_2}{\Delta \bar{z}_2}\left(\frac{\partial^2 \bar{z}_2}{\partial z_2^{\,2}}\right) \tag{3.36a}$$

$$L_{\Omega_2} = -\frac{k_2}{\Delta \bar{z}_2^{\,2}}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right)^2 - \frac{k_2}{\Delta \bar{z}_2}\left(\frac{\partial^2 \bar{z}_2}{\partial z_2^{\,2}}\right) \tag{3.36b}$$

$$U_{\Omega_2} = -\frac{k_2}{\Delta \bar{z}_2^{\,2}}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right)^2 \tag{3.36c}$$

$$R_{\Omega_2} = \frac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n+1} + T_{i+1,j}^{n}\right) + \frac{\rho_2 c_2 u_2}{\Delta x} T_{i-1,j}^{n+1} + R_{\Omega_2}^{n} \tag{3.36d}$$

$$R_{\Omega_2}^{n} = \left\{ \begin{array}{l} \dfrac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n} + T_{i+1,j}^{n} - 2T_{i,j}^{n}\right) - \rho_2 c_2 u_2 \dfrac{T_{i,j}^{n} - T_{i-1,j}^{n}}{\Delta x} + \dfrac{2\rho_2 c_2}{\Delta t} T_{i,j}^{n} \\ + \dfrac{k_2}{\Delta \bar{z}_2^{\,2}}\left(\dfrac{\partial \bar{z}_2}{\partial z_2}\right)^2 \left(T_{i,j-1}^{n} + T_{i,j+1}^{n} - 2T_{i,j}^{n}\right) + \dfrac{k_2}{\Delta \bar{z}_2}\left(\dfrac{\partial^2 \bar{z}_2}{\partial z_2^{\,2}}\right)\left(T_{i,j-1}^{n} - T_{i,j}^{n}\right) \end{array} \right\} \tag{3.36e}$$

For the control volume on the surface, contributions from Solid 1 to the coefficients are

$$D_{S_1} = \frac{\Delta z_1}{2}\left\{\frac{2\rho_1 c_1}{\Delta t} + \frac{\rho_1 c_1 u_1}{\Delta x} + \frac{2k_1}{\Delta x^2}\right\} + \frac{k_1}{\Delta \bar{z}_1}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right) \tag{3.37a}$$

$$L_{S_1} = 0 \tag{3.37b}$$

$$U_{S_1} = -\frac{k_1}{\Delta \bar{z}_1}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right) \tag{3.37c}$$

$$R_{S_1} = \frac{\Delta z_1}{2}\left\{\begin{array}{l}\dfrac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\dfrac{\rho_1 c_1 u_1}{\Delta x}T_{i-1,j}^{n+1} \\ +\dfrac{k_1}{\Delta x^2}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)-\dfrac{\rho_1 c_1 u_1}{\Delta x}\left(T_{i,j}^{n}-T_{i-1,j}^{n}\right)+\dfrac{2\rho_1 c_1}{\Delta t}T_{i,j}^{n}\end{array}\right\} \\ +\frac{k_1}{\Delta \bar{z}_1}\left(\frac{\partial \bar{z}_1}{\partial z_1}\right)\left(T_{i,j+1}^{n}-T_{i,j}^{n}\right) \tag{3.37d}$$

and contributions from Solid 2 to the coefficients are

$$D_{S_2} = \frac{\Delta z_2}{2}\left\{\frac{2\rho_2 c_2}{\Delta t}+\frac{\rho_2 c_2 u_2}{\Delta x}+\frac{2k_2}{\Delta x^2}\right\}+\frac{k_2}{\Delta \bar{z}_2}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right) \tag{3.38a}$$

$$L_{S_2} = -\frac{k_2}{\Delta \bar{z}_2}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right) \tag{3.38b}$$

$$U_{S_2} = 0 \tag{3.38c}$$

$$R_{S_2} = \frac{\Delta z_2}{2}\left\{\begin{array}{l}\dfrac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n+1}+T_{i+1,j}^{n}\right)+\dfrac{\rho_2 c_2 u_2}{\Delta x}T_{i-1,j}^{n+1} \\ +\dfrac{k_2}{\Delta x^2}\left(T_{i-1,j}^{n}+T_{i+1,j}^{n}-2T_{i,j}^{n}\right)-\dfrac{\rho_2 c_2 u_2}{\Delta x}\left(T_{i,j}^{n}-T_{i-1,j}^{n}\right)+\dfrac{2\rho_2 c_2}{\Delta t}T_{i,j}^{n}\end{array}\right\} \\ +\frac{k_2}{\Delta \bar{z}_2}\left(\frac{\partial \bar{z}_2}{\partial z_2}\right)\left(T_{i,j-1}^{n}-T_{i,j}^{n}\right) \tag{3.38d}$$

In Equations (3.37) and (3.38), $\frac{\Delta z_1}{2}$ and $\frac{\Delta z_2}{2}$ are the physical values corresponding to $\frac{\Delta \bar{z}_1}{2}$ and $\frac{\Delta \bar{z}_2}{2}$ respectively, and they can be obtained from Equation (3.32).

The above formulations for the solids with domain transformation can be changed back to formulations with uniform mesh by simply setting $\bar{z} = z$, $\frac{\partial \bar{z}}{\partial z} = 1$, and $\frac{\partial^2 \bar{z}}{\partial z^2} = 0$.

3.3.3 Solution Procedure

The thermal model is developed based on the numerical formulations described in the previous sections. This model is then incorporated into the EHL line contact solution program and the whole thermal EHL solution procedure is illustrated in Figure 3.8. The dashed box is the thermal module. The iterative scheme used is the coupled solution method described in Chapter 2. The thermal analysis is carried out once for every trial pressure in the overall iterative scheme. Energy equations for both film and solids are solved to determine the temperature distribution within the film corresponding to the current pressure and film thickness distribution. The downstream temperatures at the new timestep that are incorporated in the right hand side of the energy equation are taken as those from the previous cycle. Then the viscosity and density of the lubricant are updated using these temperatures and pressures to determine new pressure and film thickness distributions. This iterative process continues until a solution is obtained which simultaneously satisfies all EHL equations.

For temperature convergence, two levels of criteria are employed. First, changes of maximum values of the mean fluid temperature and two surface temperatures are checked. If these changes are less than 0.05°C, then mid-plane temperature and surface temperature at all mesh points are checked using the following criterion:

$$\max_i\left\{\frac{\left|T_{mi}^{new}-T_{mi}^{old}\right|}{T_{mi}^{old}}\right\}+\max_i\left\{\frac{\left|T_{1i}^{new}-T_{1i}^{old}\right|}{T_{1i}^{old}}\right\}+\max_i\left\{\frac{\left|T_{2i}^{new}-T_{2i}^{old}\right|}{T_{2i}^{old}}\right\}\le 10^{-4} \tag{3.39}$$

where T_m is the mid-plane temperature in the fluid and T_1 and T_2 are the two surface temperatures respectively.

The cross-film mean (or effective) temperature is obtained from (Sharif et al. 1999)

$$T_e=\frac{1}{h}\int_0^h T(z)dz \tag{3.40}$$

which is used to determine the effective viscosity and density via Equations (1.14) and (1.16), respectively.

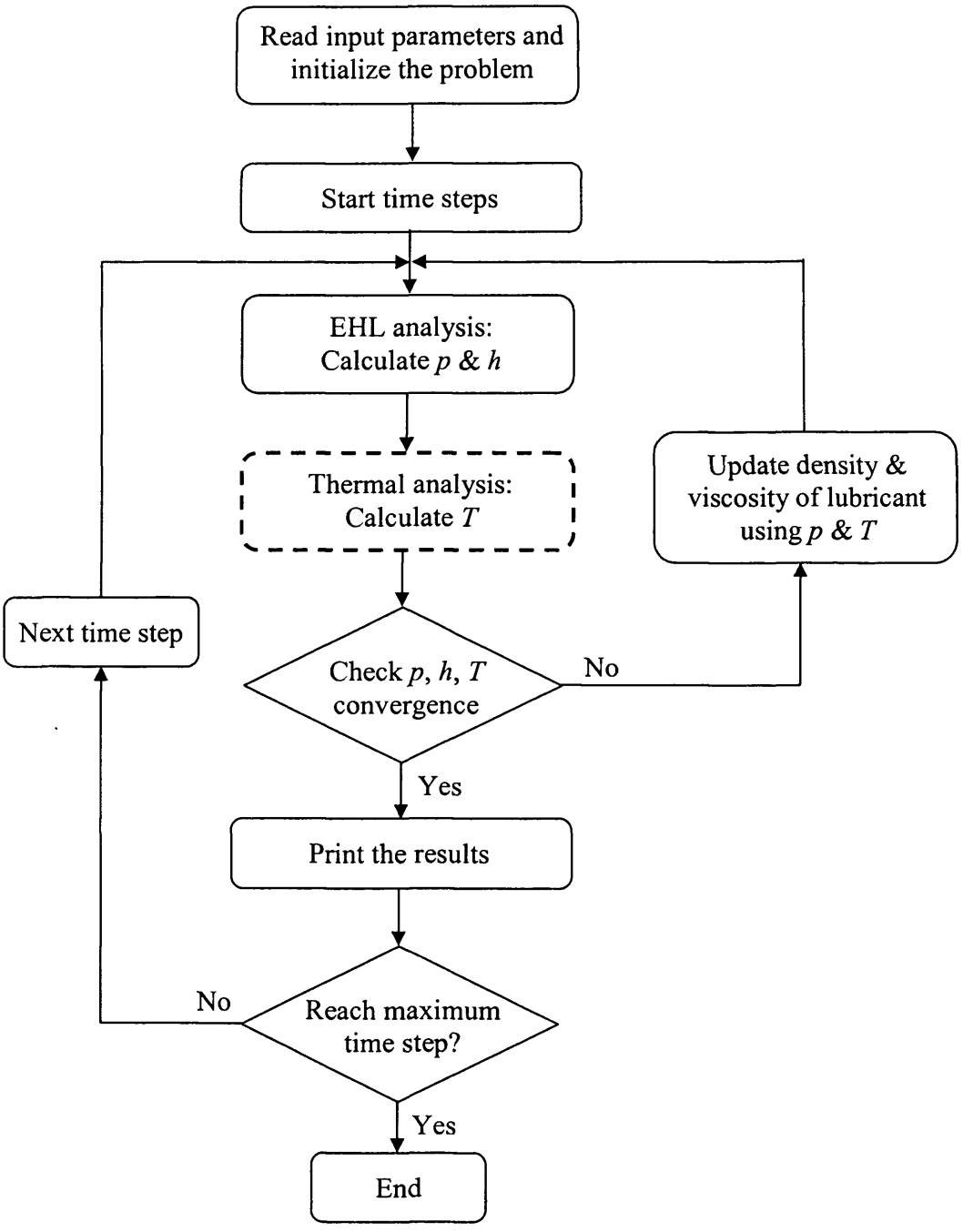


Figure 3.8: Thermal analysis procedure for EHL line contact.

3.4 Results and Discussions

The thermal analysis for the rough surface EHL line contact problem is performed based on the numerical formulations described in the previous section. The rough profiles analysed are those shown in Figure 2.2 of Chapter 2. An entrainment velocity of $\bar{u}$ = 25m/s with a slide roll ratio of ξ = 0.5 is applied unless otherwise stated. The same load is applied as used in Chapter 2 so that the maximum Hertzian pressure in smooth dry contact is about p_0 = 1GPa with the semi-contact width a = 0.335mm. The inlet and bulk temperature is 100°C, which corresponds to conditions used in scuffing tests by Patching et al. (1995). The material properties for both lubricant and solids are listed in Table 3.1. The mesh spacing used is $\Delta x = a/200$. The timestep used, as in Chapter 2, is $\Delta t = u_1/2\Delta x$, where u_1 is the velocity of the faster moving surface. For the solution of the energy equation, 21 uniformly distributed mesh points across the oil film (in the z direction) are used. The number of nodes in the solids will be discussed later.

Table 3.1: Material properties used for thermal EHL analysis.

η_0	0.0048 Pa·s	λ	1.68 GPa^{-1}
α	11.1 GPa^{-1}	ε	6.4×10^{-4} K^{-1}
τ_0	10 MPa	k	0.131 W/mK
κ	63.2×10^{-6} Pa·s	c	2000 J/kgK
χ	5.1 GPa^{-1}	ρ_1, ρ_2	7850 kg/m^3
ρ_0	846 kg/m^3	k_1, k_2	48 W/mK
γ	2.27 GPa^{-1}	c_1, c_2	450 J/kgK

3.4.1 Determination of Boundary in Solids

From Equation (3.9), the thermal boundary depths of the two solids can be estimated. As stated in section 3.2.2, this depth is typically around 60μm. Based on this estimation, a simple test is required to check if the depth selected is sufficiently deep so that temperature variations are negligibly small with increasing boundary depth. For this, thermal analyses are performed for contact between two smooth surfaces (steady-state) with six boundary depths of 30, 40, 50, 60, 70, and 80μm. The same boundary depth is selected for both surfaces. To eliminate the influence of mesh size from the comparison, a uniform mesh with a spacing of 1μm is selected across the solids (in the z direction) for all the cases. Figure 3.9 shows the temperature rise at the two surfaces for these boundary depths. ΔT_1 (upper part) is the temperature rise for the upper surface and ΔT_2 (lower part) is for the lower surface. It can be seen that from 50μm, the temperature rise is kept unchanged with the increasing thermal boundary depth. Therefore, the depth of 50μm is deep enough to maintain the accuracy of thermal analysis.

3.4.2 Influence of Mesh in Solids

In the film region, a uniform mesh of a relatively fine grid spacing (21 nodes across the film) is used. In the solid region, a non-uniform mesh with fine grids near the surface and coarse grids towards the thermal boundary in the solid can be used to capture the possible high temperature gradient at the interfaces. This idea has already been employed by Zhai and Chang (2000) and Wang et al. (2004c) in their thermal EHL analysis. In this section, the domain transform described in section 3.3.2 is used to investigate the influence of mesh in solids on efficiency and accuracy of thermal analysis results.

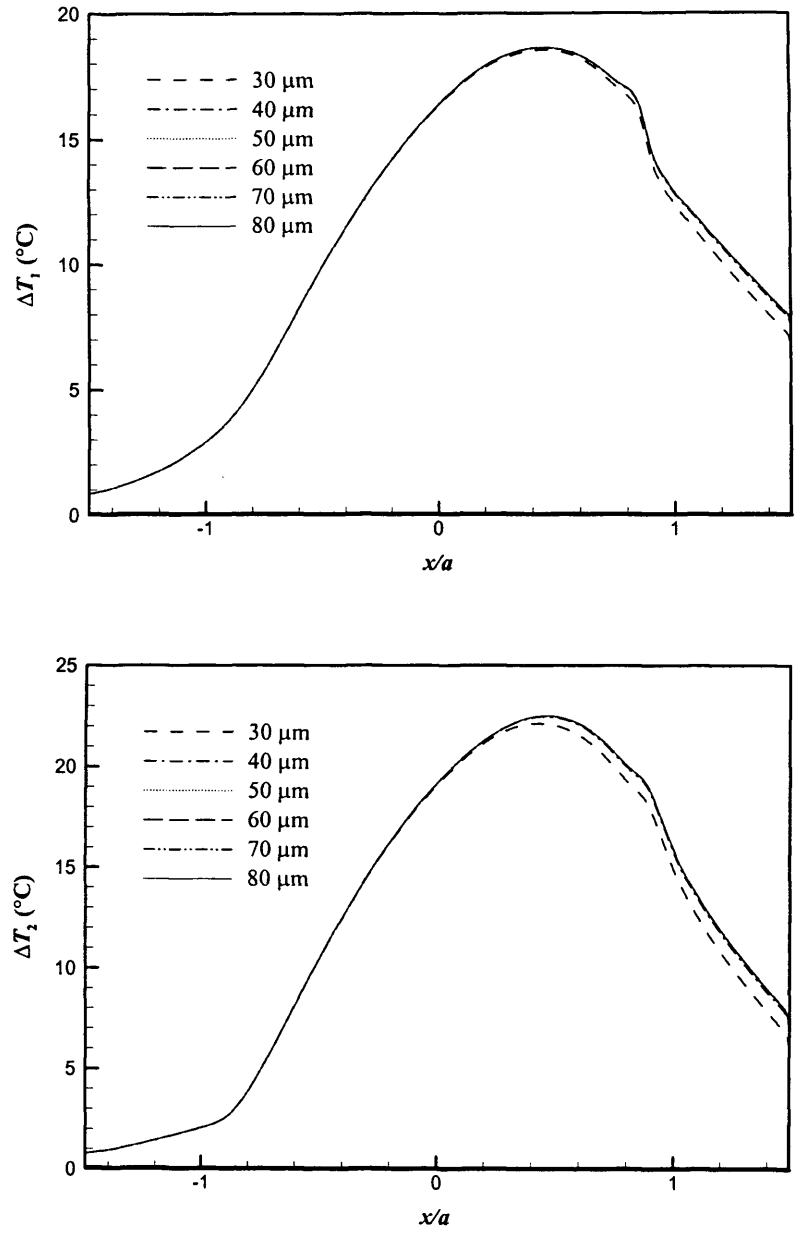


Figure 3.9: Comparison of temperature rise at the two surfaces for smooth-smooth EHL contact using different boundary depths. ΔT_1 (upper part) is the temperature rise for the upper surface and ΔT_2 (lower part) is for the lower surface.

Table 3.2 lists the CPU time to solve the same thermal EHL problem using different numbers of nodes across the solids, NZ, and different β values for domain transformation. The results are obtained from running profile (A) against a smooth surface for 4000 timesteps. The corresponding solution time for an isothermal analysis is 5'23.4". The boundary position for the two solids are $h_1 = h_2 = 50\mu m$. It can be seen that CPU time for this comparison is mainly dependent on NZ, not β.

Figures 3.10 – 3.12 compare the results obtained using different mesh spacings. The number of nodes across the solids applied in these figures are NZ = 21, 11, and 6, respectively. In these figures, curves denoted as 'reference' represent the results obtained using a uniform mesh with NZ = 51. It is clear that accuracy increases with increasing NZ. For the fine mesh (e.g. NZ = 21), the transformation is not necessary and may reduce accuracy. For the coarse mesh (e.g. NZ = 6), the transformation can increase accuracy. Based on these comparisons it was decided that a reasonable balance between efficiency and accuracy was obtained with a uniform mesh with 11 grid points.

Table 3.2: CPU time for a series of values of NZ and β.

	$NZ = 51$	$NZ = 21$	$NZ = 11$	$NZ = 6$
$\beta = 1.1$	17'31.7"	13'0.4"	11'32.1"	10'50.9"
$\beta = 1.2$	17'36.1"	12'57.6"	11'31.8"	10'48.2"
$\beta = 1.5$	17'48.6"	13'4.0"	11'25.9"	10'38.3"
$\beta = 2.0$	17'47.5"	12'54.4"	11'27.6"	10'48.3"
uniform	17'28.5"	13'1.7"	11'28.3"	10'44.0"

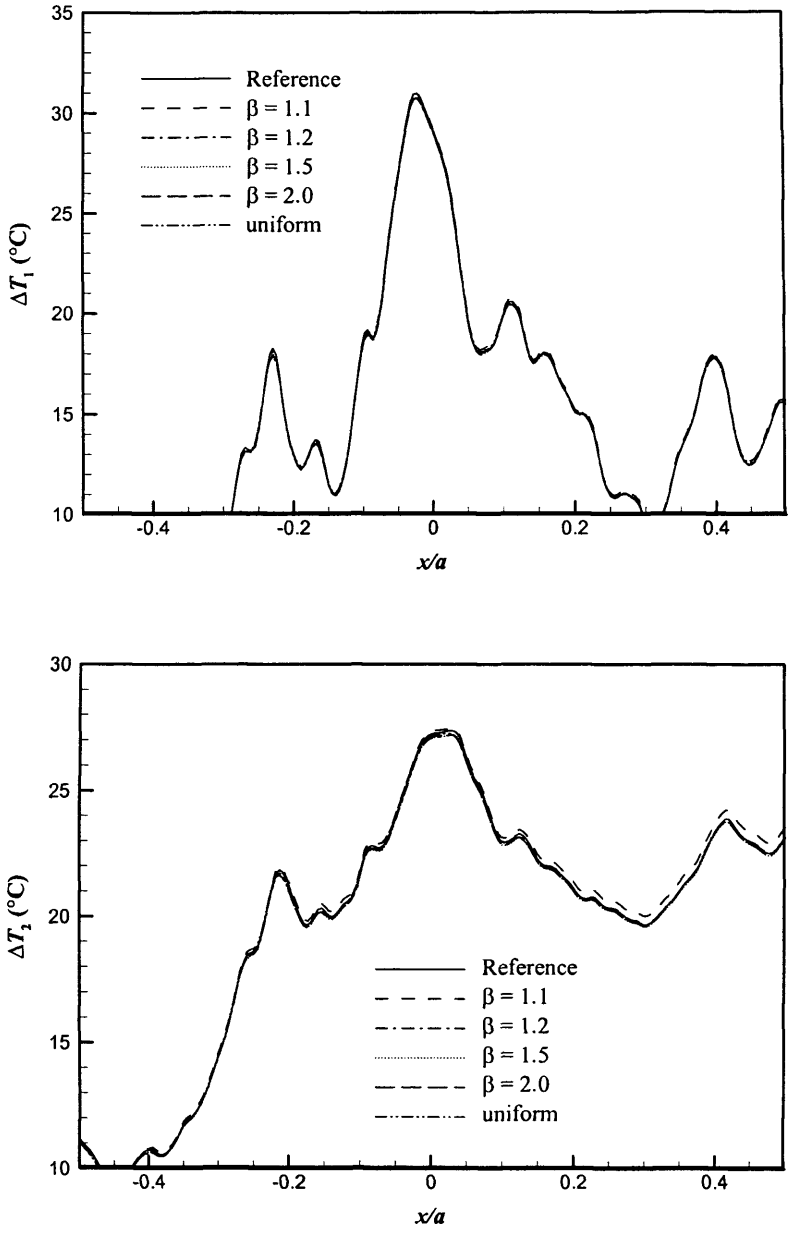


Figure 3.10: Comparison of temperature rise at the two surfaces for the case of $NZ = 21$ using different domain transformations. ΔT_1 (upper part) is the temperature rise for the upper surface and ΔT_2 (lower part) is for the lower surface.

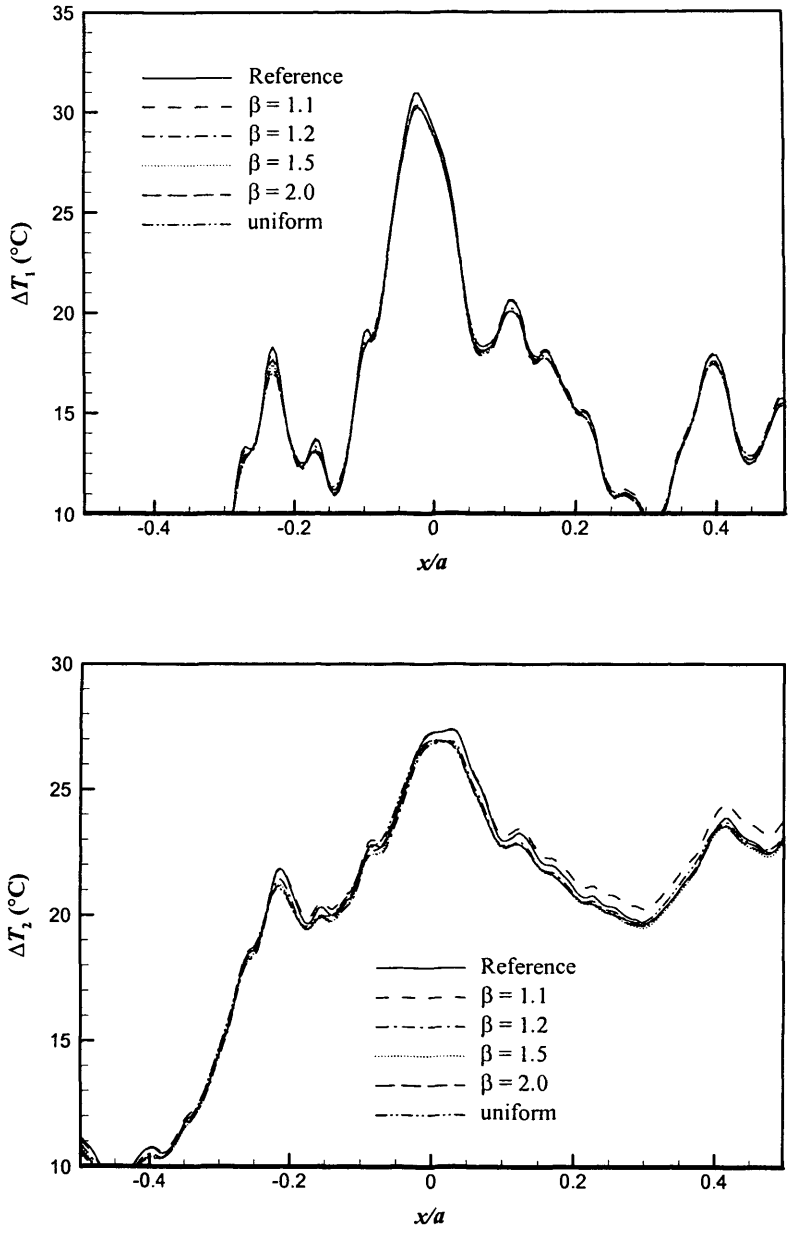


Figure 3.11: Comparison of temperature rise at the two surfaces for the case of $NZ = 11$ using different domain transformations. ΔT_1 (upper part) is the temperature rise for the upper surface and ΔT_2 (lower part) is for the lower surface.

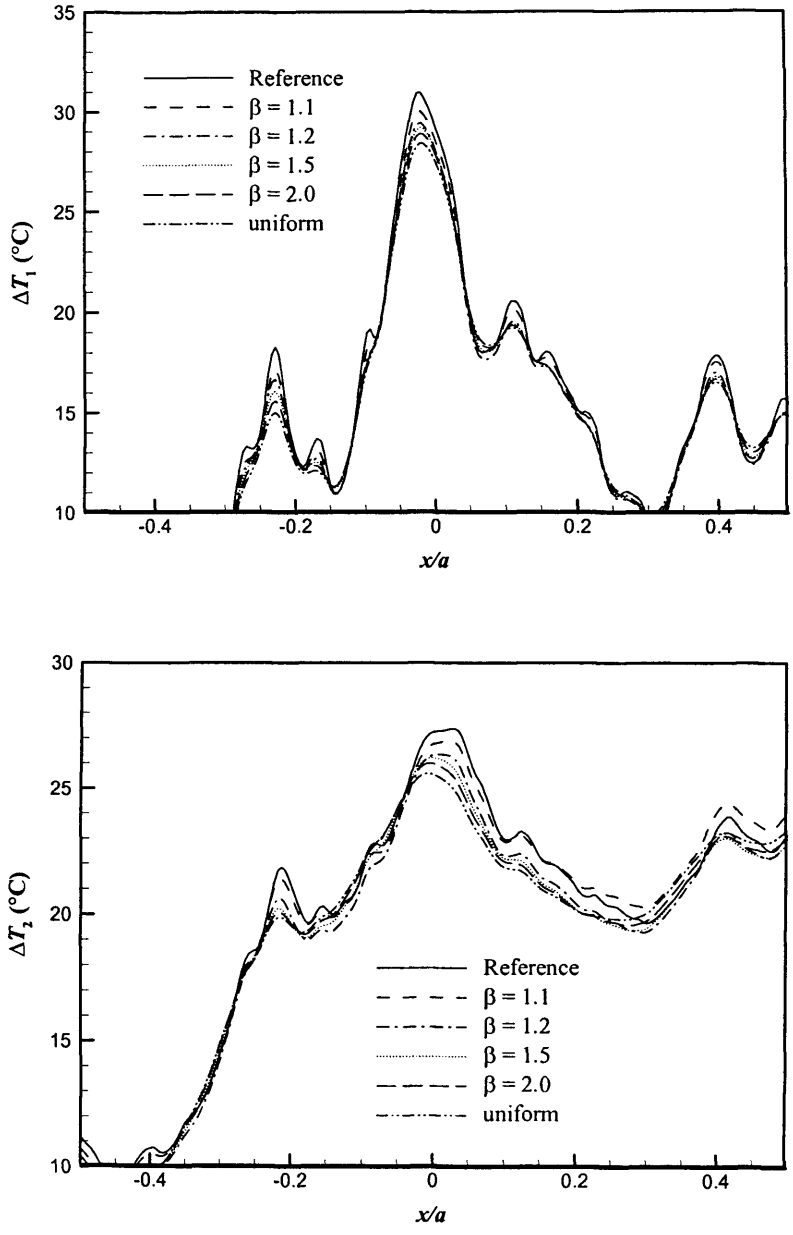


Figure 3.12: Comparison of temperature rise at the two surfaces for the case of $NZ = 6$ using different domain transformations. ΔT_1 (upper part) is the temperature rise for the upper surface and ΔT_2 (lower part) is for the lower surface.

3.4.3 Simulation of Thermal Behaviour

In this section, numerical results are presented for thermal EHL analysis. Based on the previous investigations, $h_1 = h_2 = 50\mu m$ are chosen for the thermal boundary in the solids and 11 uniformly distributed points across the solids are used for analysis.

Figure 3.13 shows the results obtained from running profile (A) against a smooth surface at timestep 4000, when the faster surface (rough surface) has moved a distance of $10a$. The upper part of the figure shows the pressure and film thickness distributions in the contact area. The lower part of the figure shows the mean lubricant temperature rise and temperature rise for two surfaces. Also shown in the figure are the corresponding results for the case of smooth-smooth contact for comparison. It can be seen that substantial perturbation to the smooth surface pressure, film thickness and temperature rise has been developed due to the moving rough surfaces. It is clearly seen that the rise and fall of mean lubricant temperature is broadly similar to the pressure pattern in the lubricant. This phenomenon is also observed by using different rough profiles, such as the profile (B) against a smooth surface (shown in Figure 3.14) and the profile (C) against a smooth surface (shown in Figure 3.15). This can be explained by the fact that, in the region of contact the dominant heat generating mechanism is shear heating which is related to the pressure distribution through its effect on the lubricant viscosity. Figure 3.16 shows the pressure distribution and corresponding compressive heating term, *HC*, and shear heating term, *HS*, for the smooth contact and the rough contact. *HC* and *HS* are defined by

$$HC = \int_0^h \varepsilon T\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x}\right)dz \tag{3.41a}$$

$$HS = \int_0^h \tau\dot{\gamma}\, dz \tag{3.41b}$$

It is clear that *HS* has the same distribution pattern as p and is generally much higher than *HC*.

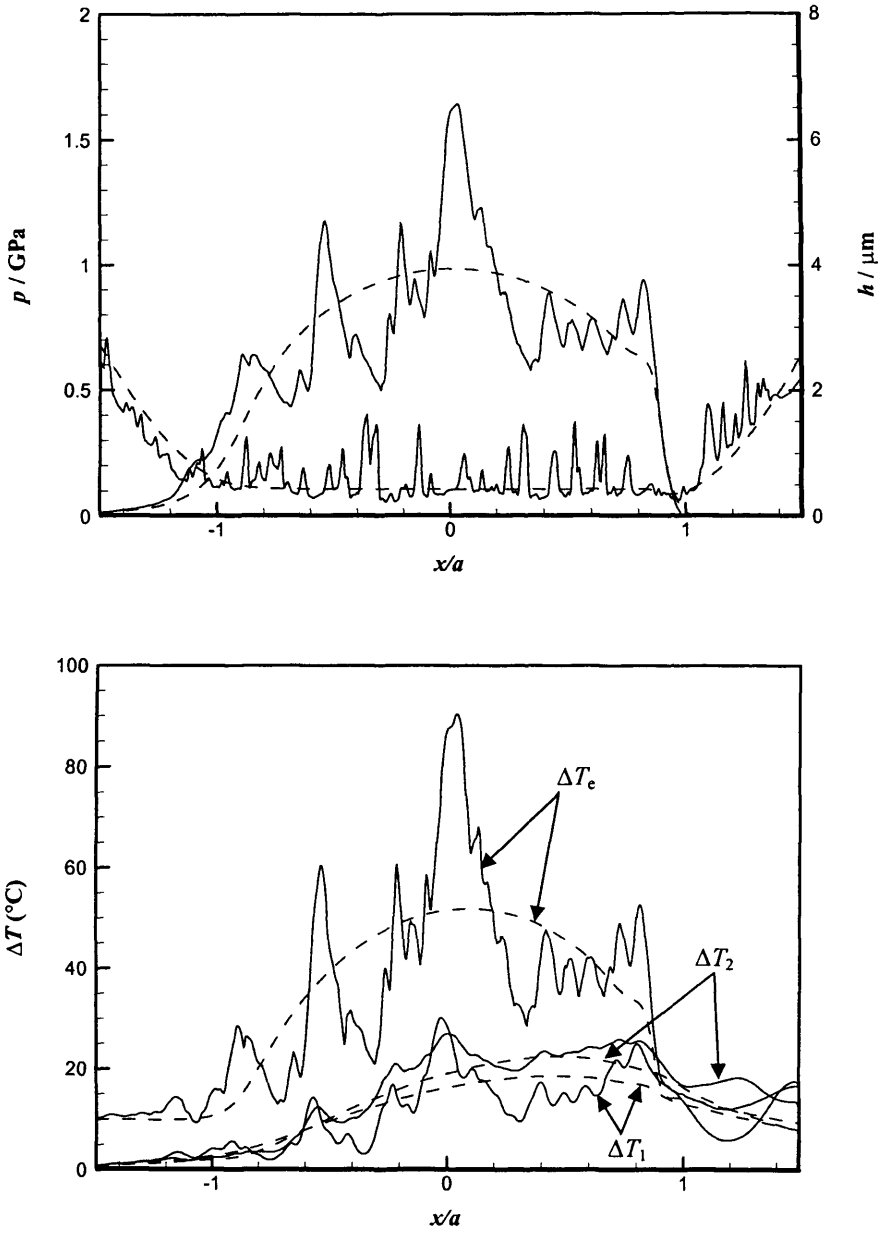


Figure 3.13: Pressure, film thickness and temperature rise distributions at one timestep for the contact between profile (A) and a smooth surface. ΔT_e, ΔT_1 and ΔT_2 are the temperature rise for oil mean temperature, upper surface and lower surface, respectively.

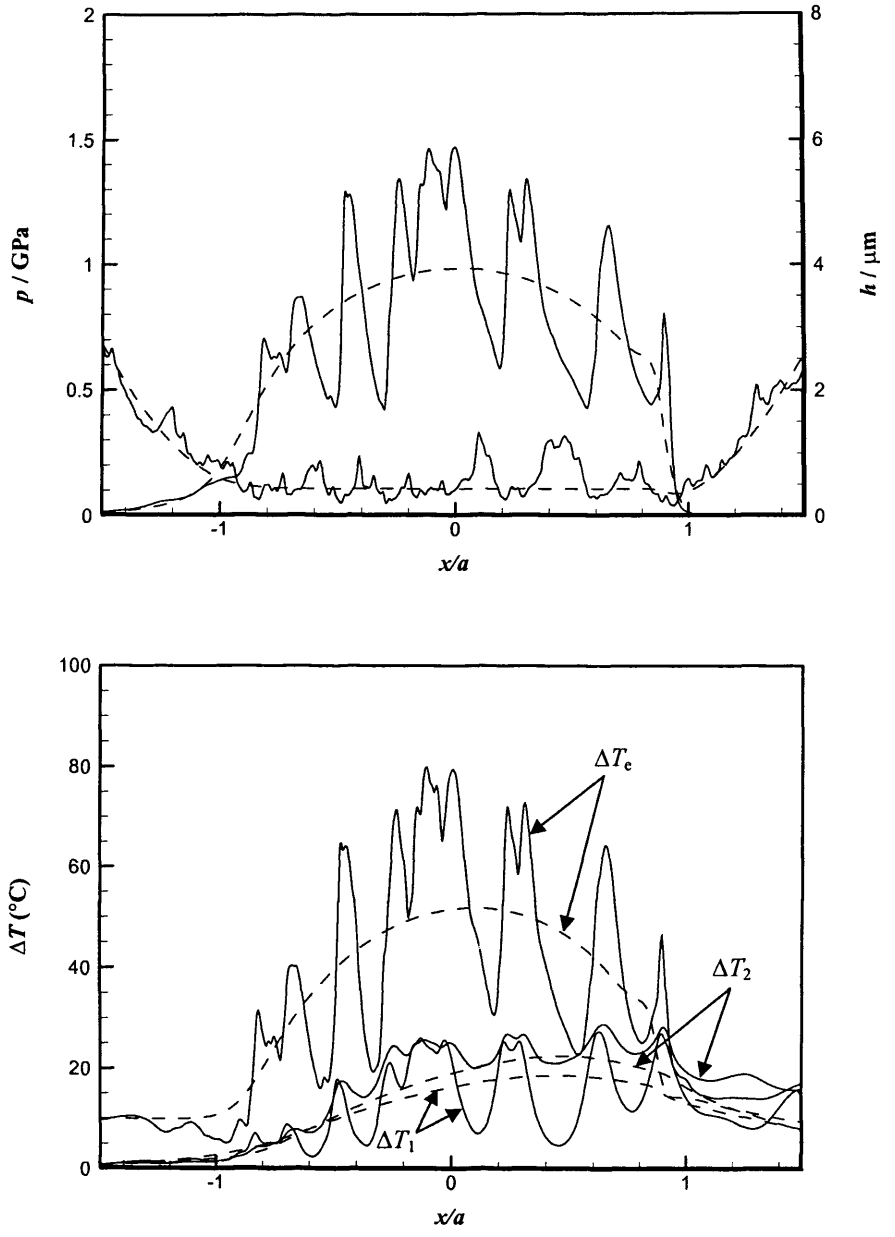


Figure 3.14: Pressure, film thickness and temperature rise distributions at one timestep for the contact between profile (B) and a smooth surface. ΔT_e, ΔT_1 and ΔT_2 are the temperature rise for oil mean temperature, upper surface and lower surface, respectively.

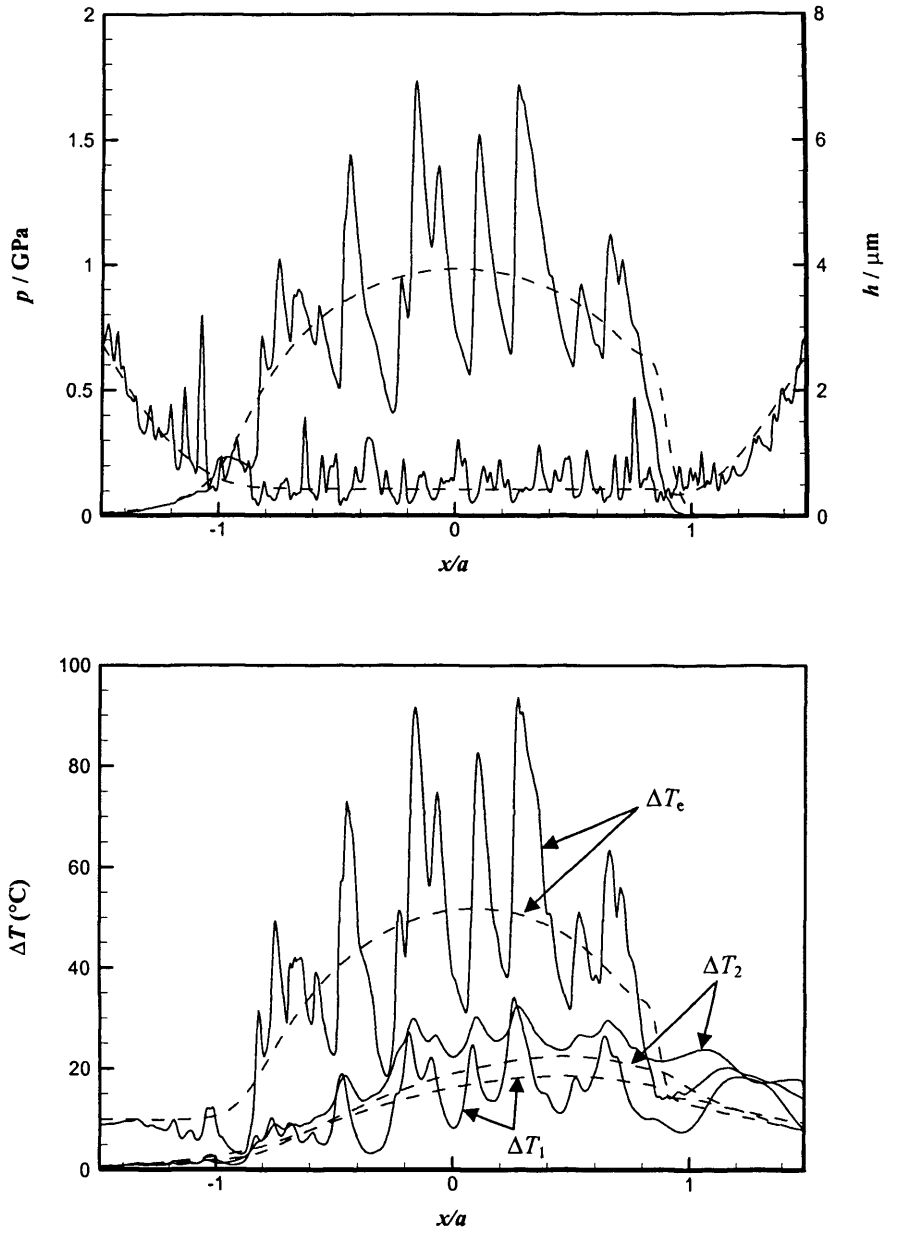


Figure 3.15: Pressure, film thickness and temperature rise distributions at one timestep for the contact between profile (C) and a smooth surface. ΔT_e, ΔT_1 and ΔT_2 are the temperature rise for oil mean temperature, upper surface and lower surface, respectively.

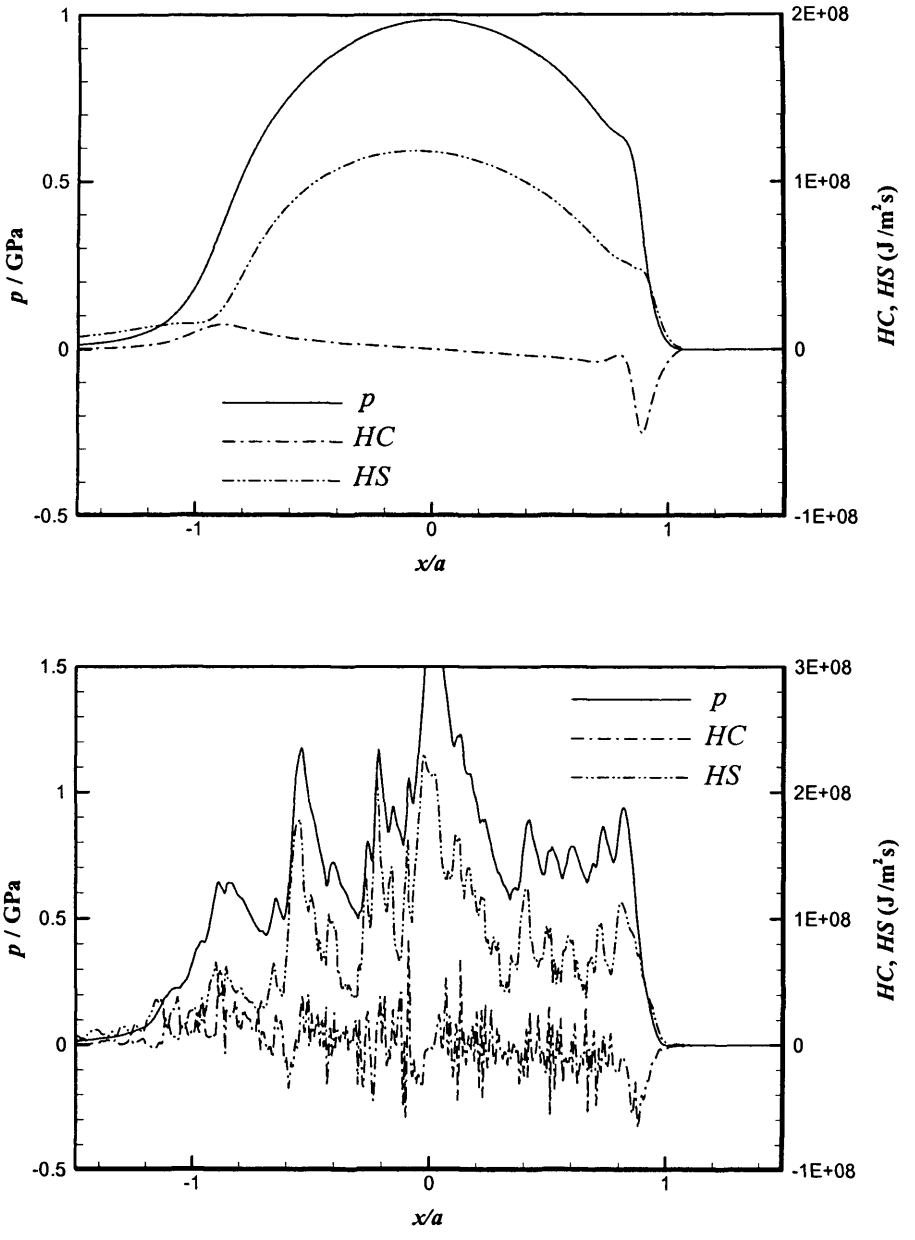


Figure 3.16: Comparison between compressive heating term, *HC*, and shear heating term, *HS*, for the smooth contact (upper figure) and the rough contact (lower figure).

Sliding in the EHL contact affects the shear stresses and thus the rheological behaviour of the lubricant and heat generation in the film. The influence of slide roll ratio ξ on the temperature rise at timestep 4000 is examined in Figure 3.17. Results obtained for three slide roll ratios of ξ = 0.25, 0.5 and 1.0 are compared with the same entrainment velocity of $\bar{u}$ = 25m/s. It is clear that as the ratio ξ increases, the temperature rises for both the lubricant and solids increase substantially.

Figure 3.18 shows the influence of the entrainment velocity on the temperature rise. Three velocity values of $\bar{u}$ = 5, 10 and 25m/s are used with the same slide roll ratio ξ = 0.5. The results shown in the figure are the temperature rises at timestep 4000. It can be seen that as $\bar{u}$ increases, the temperature rise increases due to viscous dissipation.

In the above comparisons, the temperature for boundary conditions (the oil inlet temperature and the solid ambient temperature) is T_∞ = 100°C. Important physical properties of the lubricant such as viscosity, pressure coefficient of viscosity and density are temperature dependent. Therefore it is necessary to study the influence of T_∞ on temperature rise. Table 3.3, which is taken from Patching et al. (1995), summarises viscosity (η_0) and pressure coefficient of viscosity (α) at three values of temperature T_∞ = 40, 100 and 140°C. The density is assumed not to change with temperature in this study. Figure 3.19 shows results obtained at timestep 4000. The slide roll ratio is ξ = 0.5 with the velocity $\bar{u}$ = 25m/s. It can be seen that with decreased T_∞ the temperature rise in the lubricant increases due to viscous dissipation with higher viscosity (η_0) and pressure coefficient of viscosity (α). However, the effect of T_∞ on the temperature rise of the two surfaces is very limited.

Table 3.3: Viscosity (η_0) and pressure coefficient of viscosity (α) as a function of temperature.

Temperature (°C)	η_0 (Pa·s)	α (m^2/N)
40	0.0250	1.60×10^{-8}
100	0.0048	1.11×10^{-8}
140	0.0024	0.96×10^{-8}

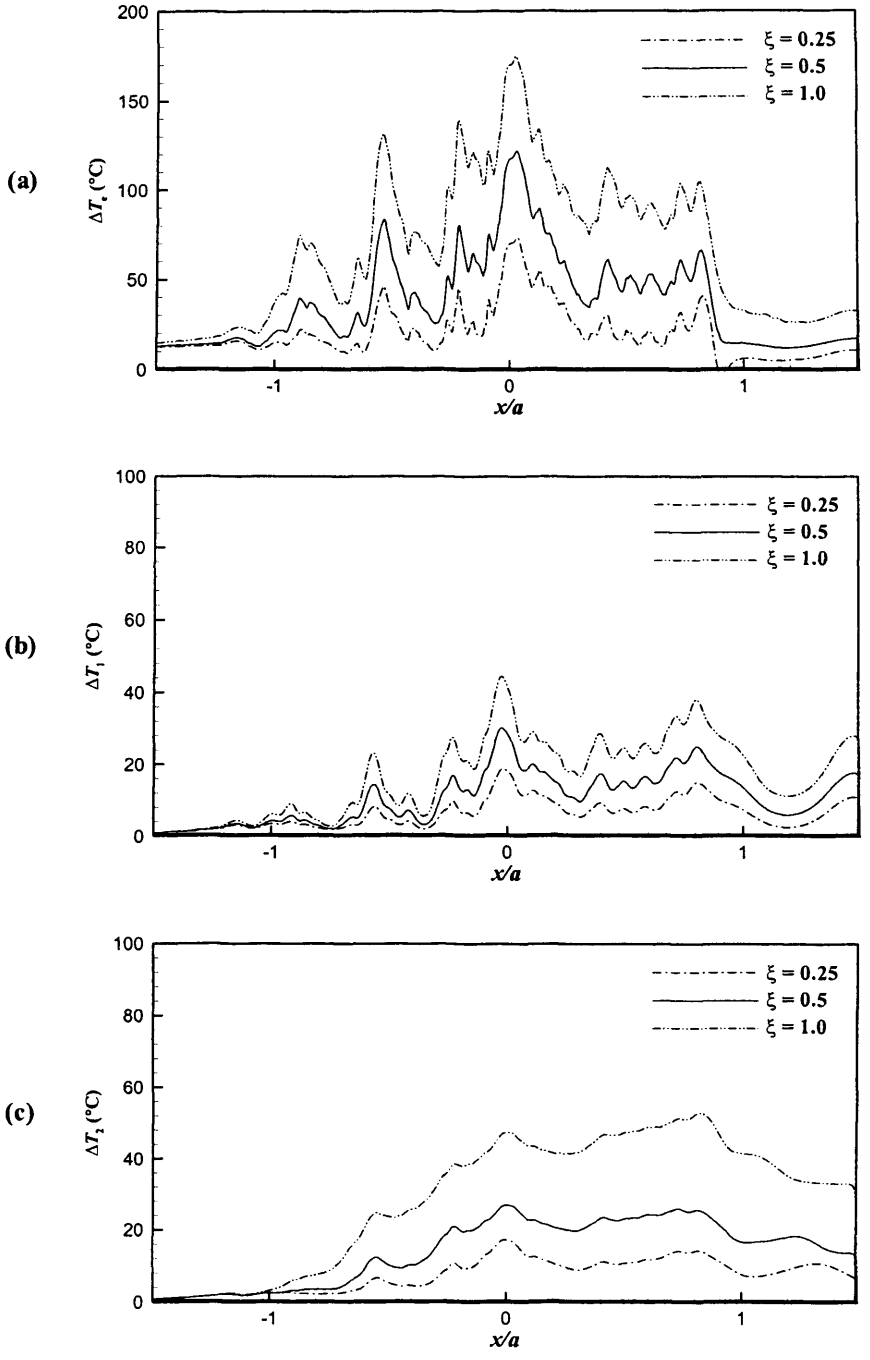


Figure 3.17: Temperature rise distributions at one timestep for the contact between profile (A) and a smooth surface with different slid roll ratios ξ = 0.25, 0.5, and 1.0: (a) oil mean temperature, (b) upper surface, and (c) lower surface.

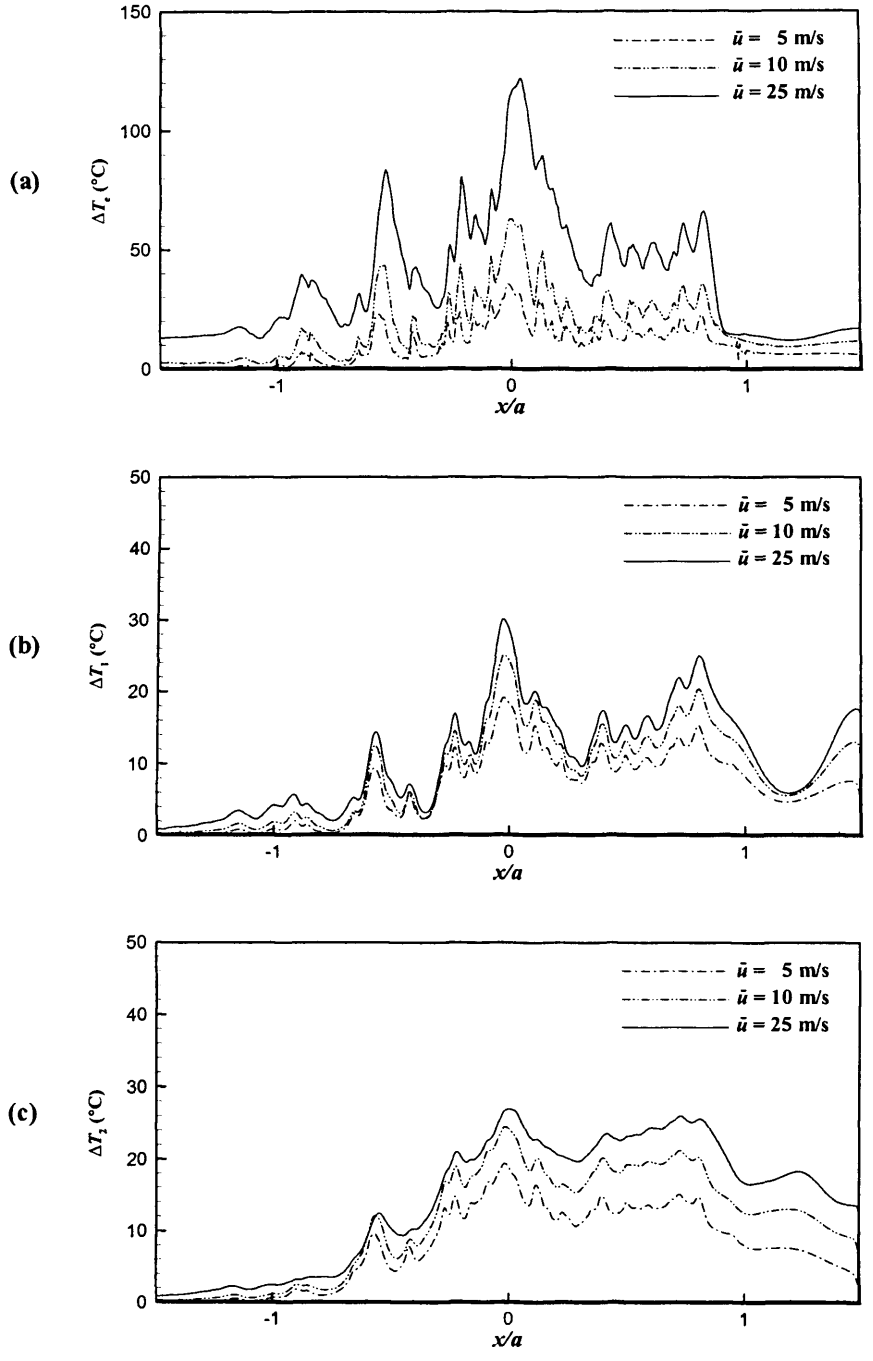


Figure 3.18: Temperature rise distributions at one timestep for the contact between profile (A) and a smooth surface with different entrainment speeds $\overline{u}$ = 5, 10, and 25m/s: (a) oil mean temperature, (b) upper surface, and (c) lower surface.

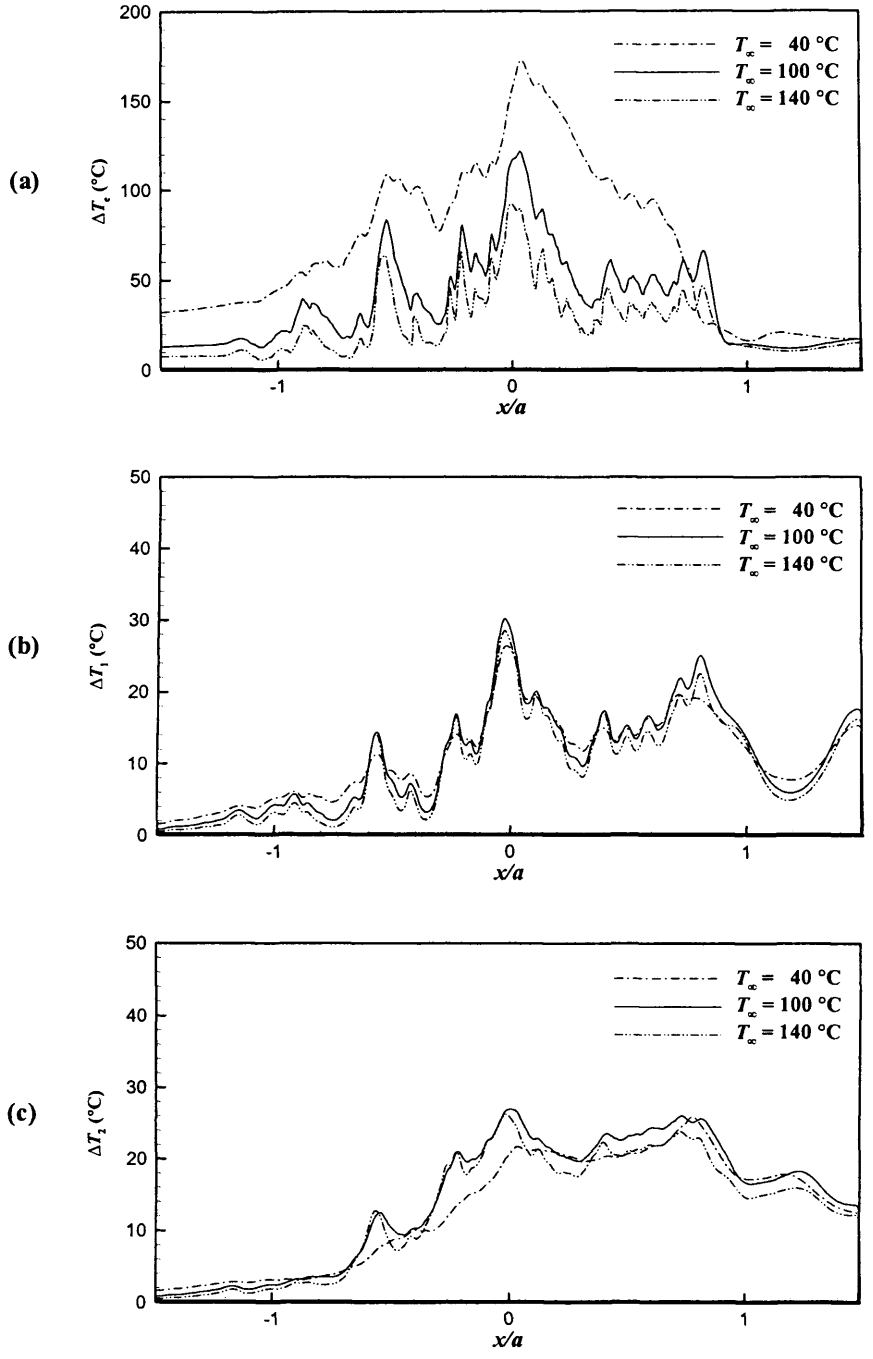


Figure 3.19: Temperature rise distributions at one timestep for the contact between profile (A) and a smooth surface with different inlet temperature values T_∞ = 40, 100, and 140°C: (a) oil mean temperature, (b) upper surface, and (c) lower surface.

The predicted pressure and film thickness distribution under thermal EHL conditions examined above is compared with corresponding isothermal results in Figures 3.20 to 3.23. In these figures, solid lines are thermal results and dotted lines are corresponding isothermal results. Figure 3.20 shows the results at one timestep for the three rough profiles against a smooth surface. The slide roll ratio is ξ = 0.5, the velocity is $\bar{u}$ = 25m/s and the inlet temperature is T_∞ = 100°C. It can be seen that considering thermal effects does not significantly alter the pressure and film thickness distribution for these cases. Figure 3.21 shows the results obtained for the contact between profile (A) and a smooth surface with three slide roll ratio values, i.e. ξ = 0.25, 0.5 and 1.0. The same velocity and inlet temperature are maintained. It can be seen that with increased slide roll ratio, thermal effects become stronger and the differences between thermal and isothermal results increase. Figure 3.22 shows the results obtained for the contact between profile (A) and a smooth surface with three velocity values of $\bar{u}$ = 5, 10 and 25m/s and ξ = 0.5 and T_∞ = 100°C. Once again, only a small difference in the pressure and film thickness distributions between thermal and isothermal results for these cases. Figure 3.23 shows results obtained using different inlet temperature values of T_∞ = 40, 100 and 140°C. For the case of T_∞ = 40°C, the viscosity is much higher than that in the other two cases (see Table 3.3). This increases the predicted film thickness significantly and also slightly reduces the pressure. The thermal effects cannot be ignored for this case. For the other two cases, inclusion of thermal effects does not significantly alter the pressure and film thickness distribution.

The pressure, film thickness and temperature distributions obtained from contacts between two rough surfaces each with the same profile are shown in Figures 3.24 to 3.26 for the three profiles. The slide roll ratio is ξ = 0.5, the velocity is $\bar{u}$ = 25m/s and the inlet temperature is T_∞ = 100°C. In all three cases, the mean oil temperature can be seen to follow the variation of pressure. Corresponding pressure and film thickness distributions for isothermal conditions are also shown in the upper part of each figure for comparison. It is seen that the pressure and film distribution predictions are essentially unchanged by the inclusion of the thermal analysis in these cases.

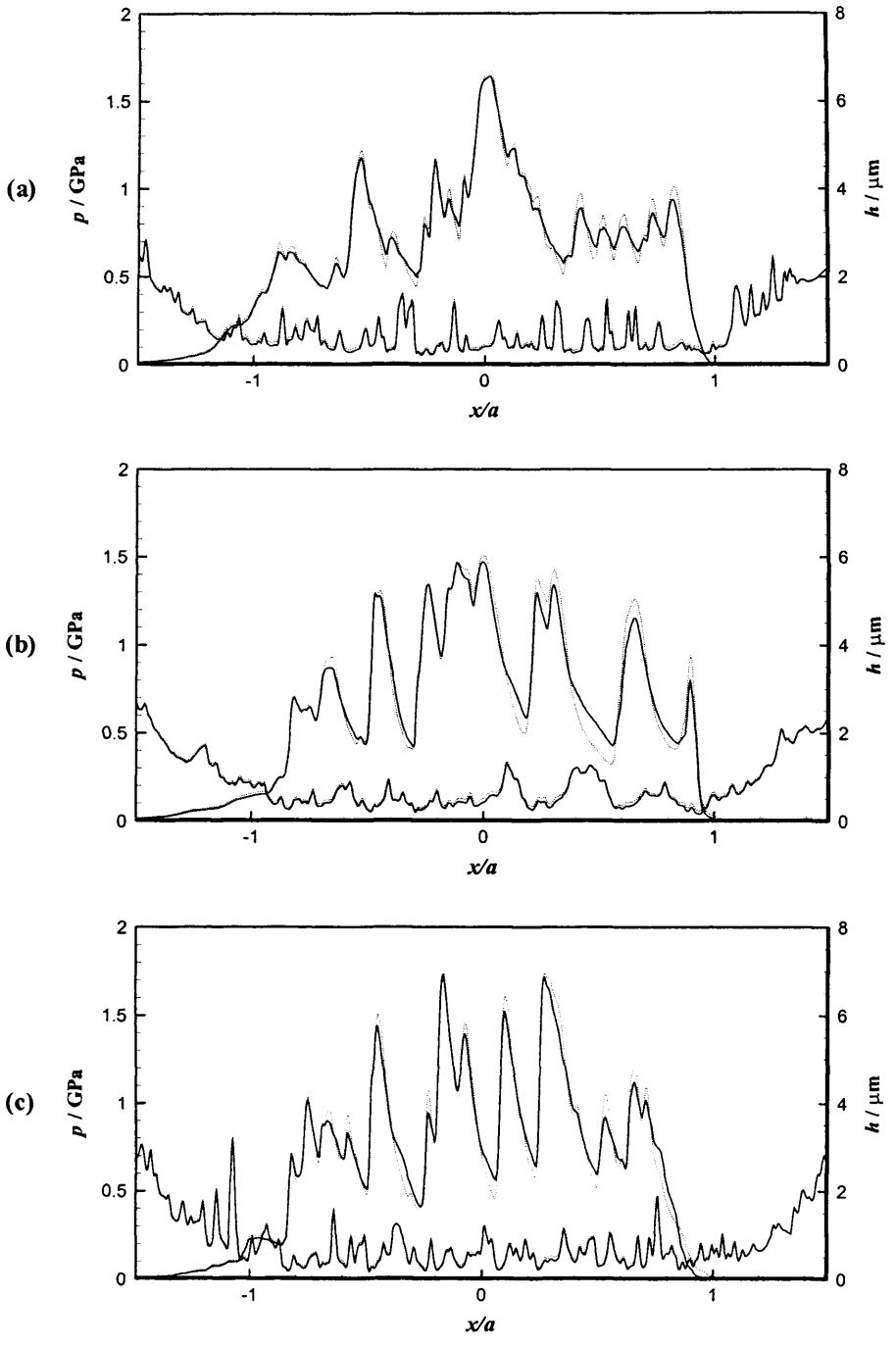


Figure 3.20: Pressure and film thickness distributions at one timestep for the contact between: (a) profile (A) and a smooth surface, (b) profile (B) and a smooth surface, and (c) profile (C) and a smooth surface. Solid lines are the thermal results and dotted lines are the isothermal results.

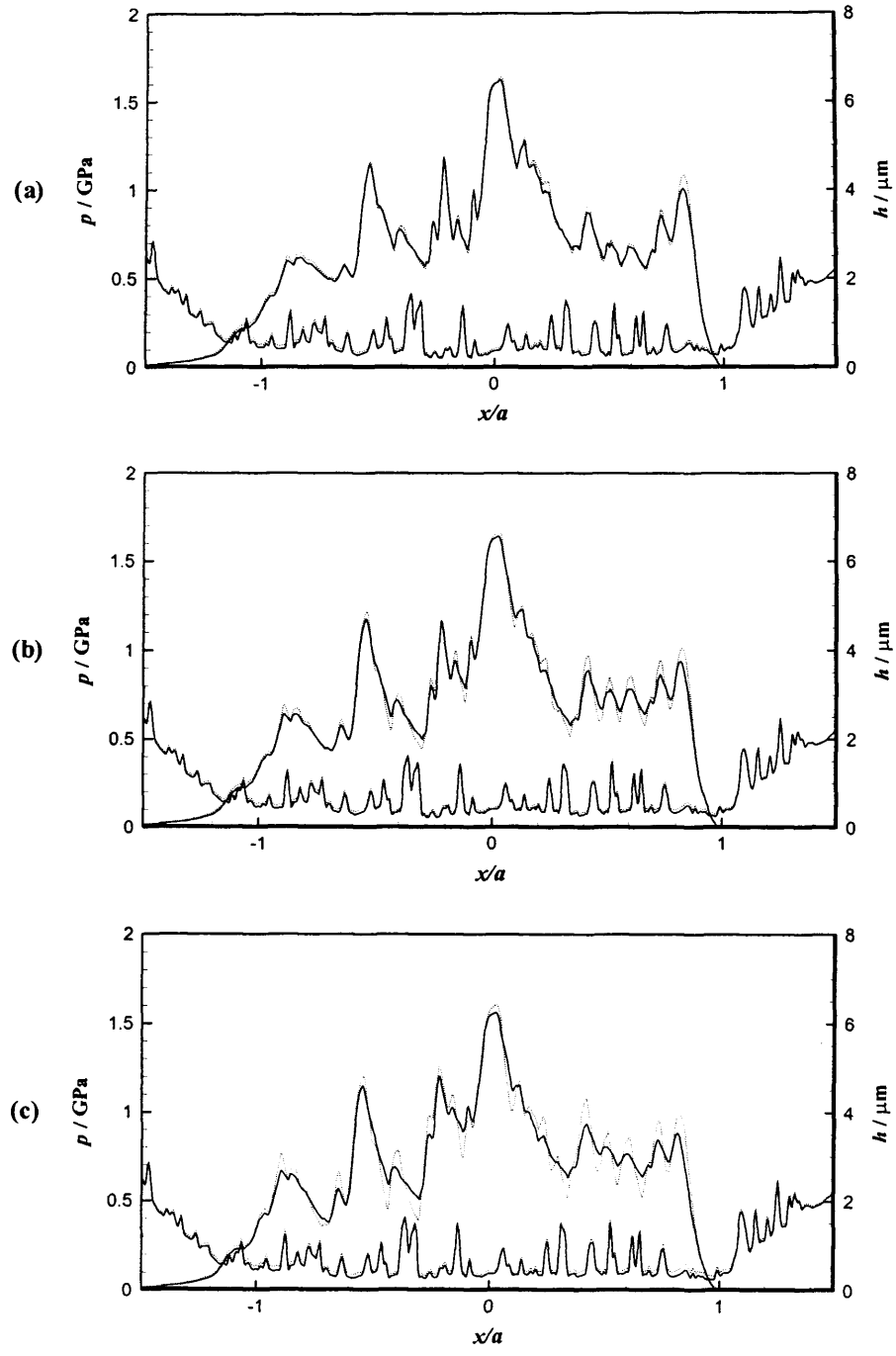


Figure 3.21: Pressure and film thickness distributions at one timestep for the contact between profile (A) and a smooth surface with different slide roll ratio: (a) ξ = 0.25, (b) ξ = 0.5, and (c) ξ = 1.0. Solid lines are the thermal results and dotted lines are the isothermal results.

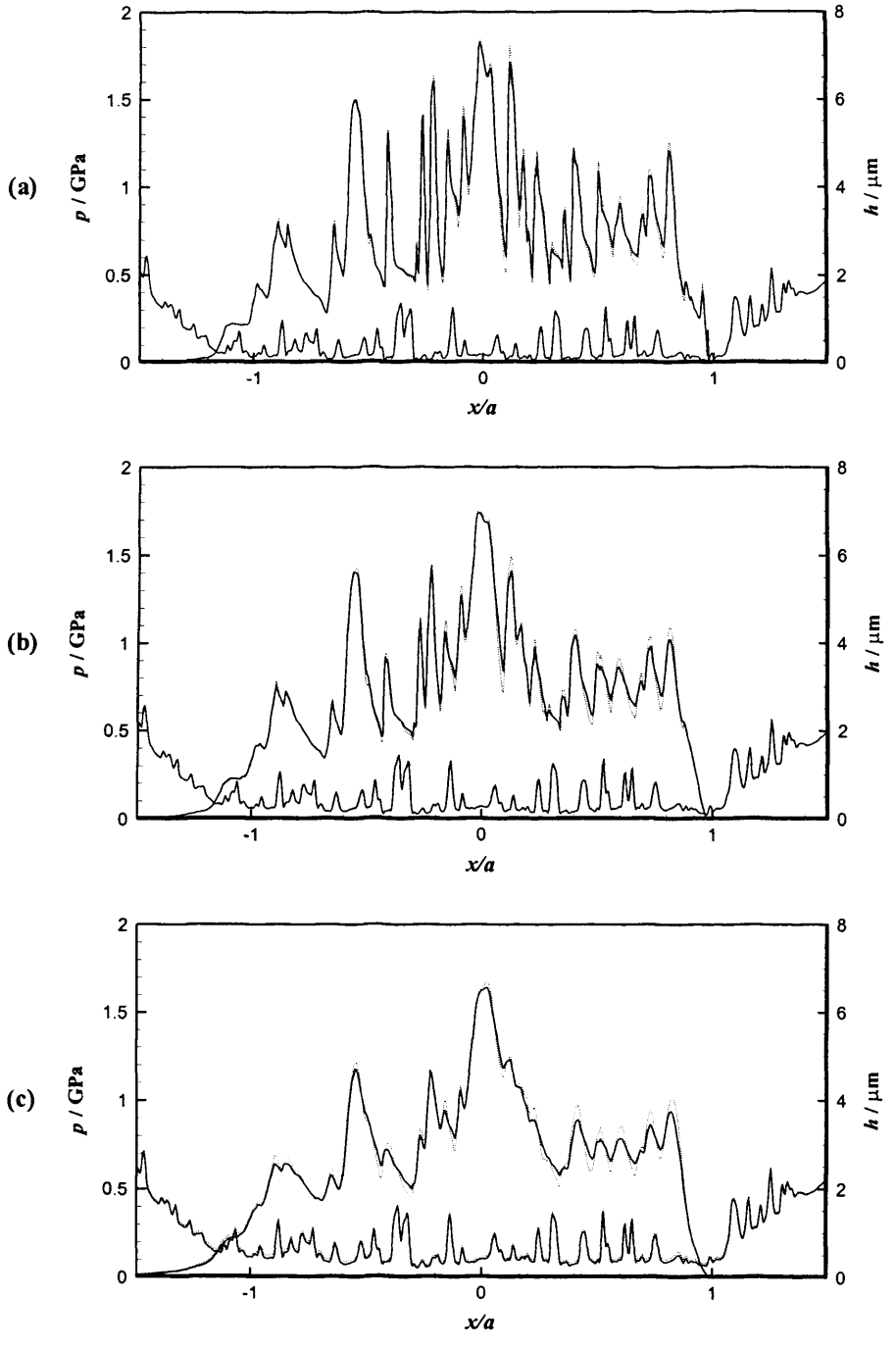


Figure 3.22: Pressure and film thickness distributions at one timestep for the contact between profile (A) and a smooth surface with different entrainment speed: (a) $\bar{u}$ = 5m/s, (b) $\bar{u}$ = 10m/s, and (c) $\bar{u}$ = 25m/s. Solid lines are the thermal results and dotted lines are the isothermal results.

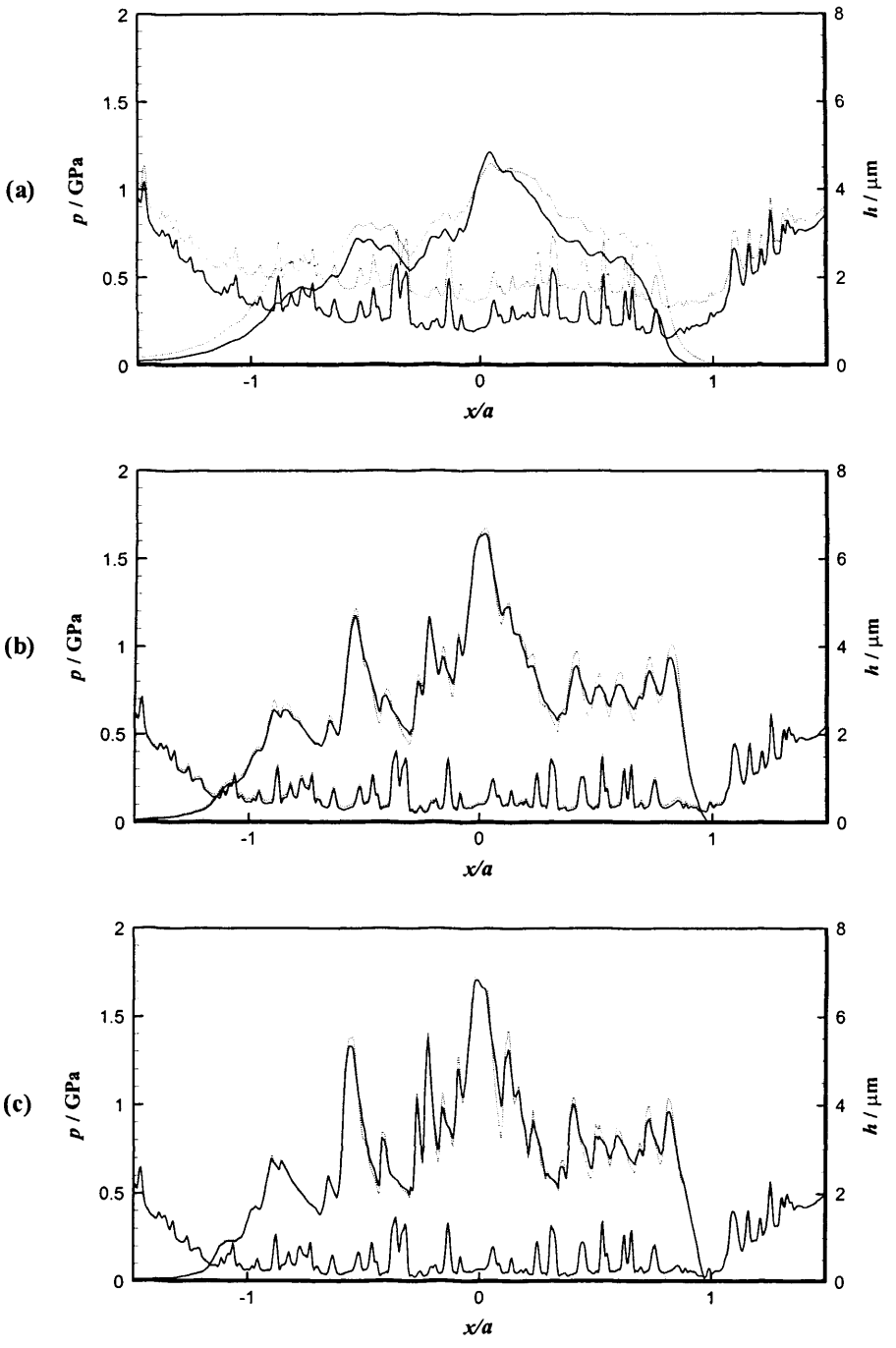


Figure 3.23: Pressure and film thickness distributions at one timestep for the contact between profile (A) and a smooth surface with different inlet temperature: (a) T_∞ = 40°C, (b) T_∞ = 100°C, and (c) T_∞ = 140°C. Solid lines are the thermal results and dotted lines are the isothermal results.

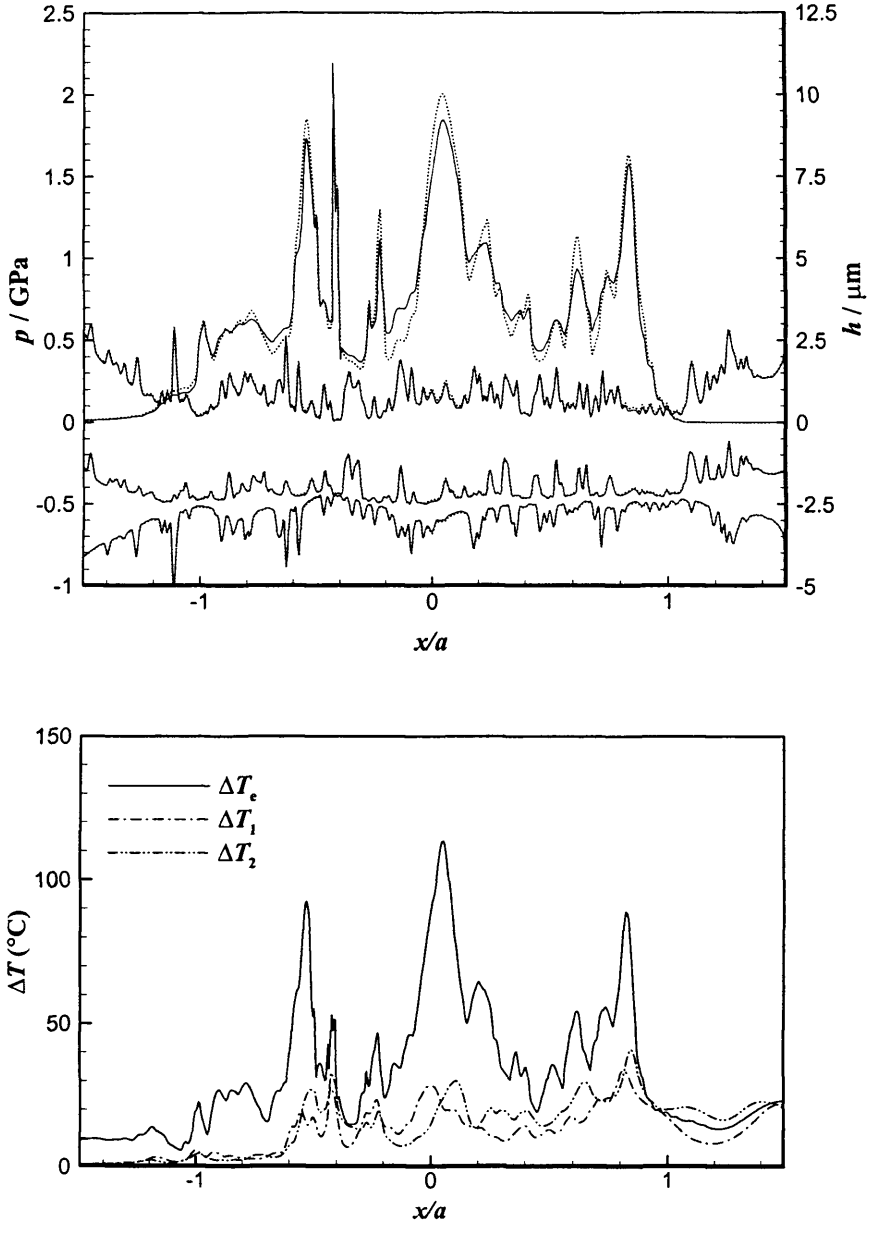


Figure 3.24: Pressure, film thickness and temperature rise distributions at one timestep for the contact between two rough surfaces each with profile (A). Pressure and film thickness distribution for corresponding isothermal case is also shown in the upper figure in dotted lines for comparison.

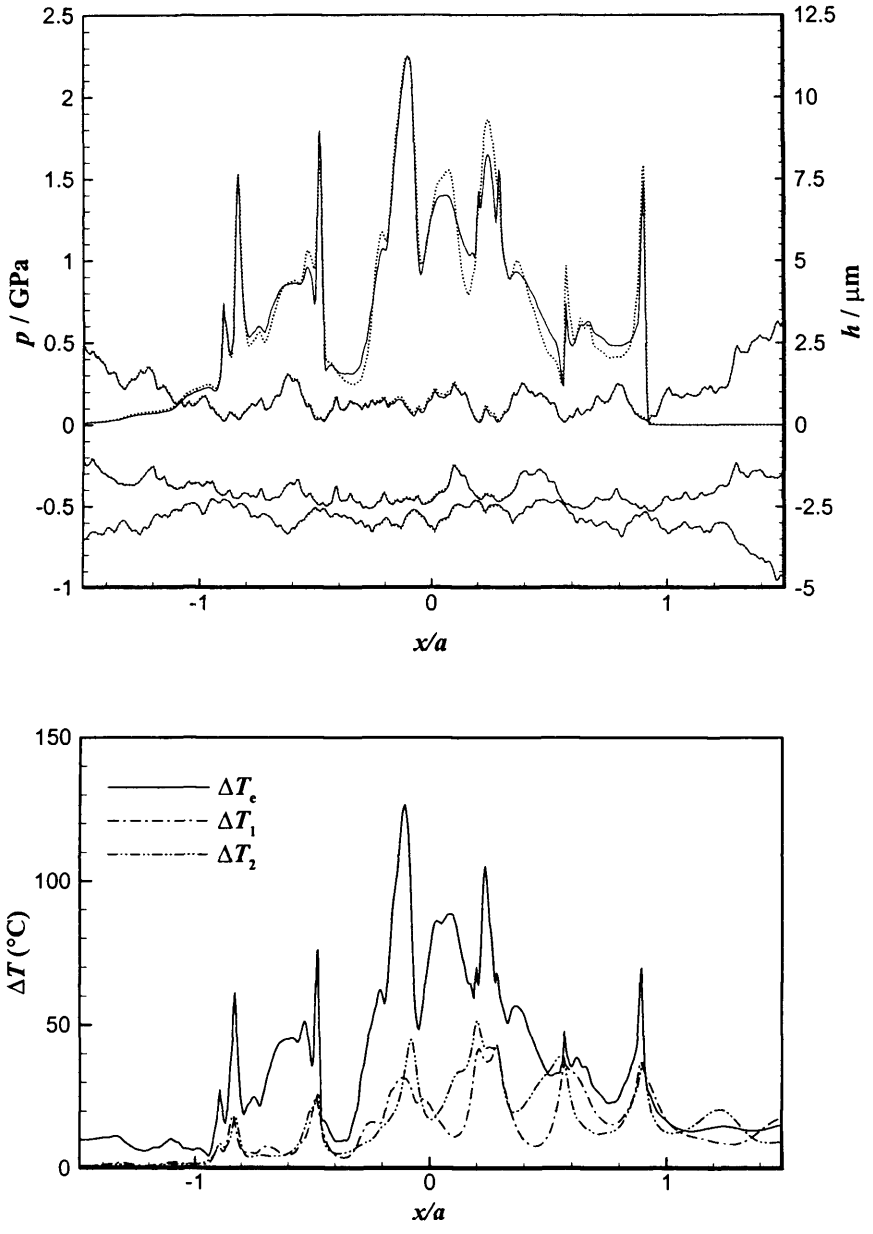


Figure 3.25: Pressure, film thickness and temperature rise distributions at one timestep for the contact between two rough surfaces each with profile (B). Pressure and film thickness distribution for corresponding isothermal case is also shown in the upper figure in dotted lines for comparison.

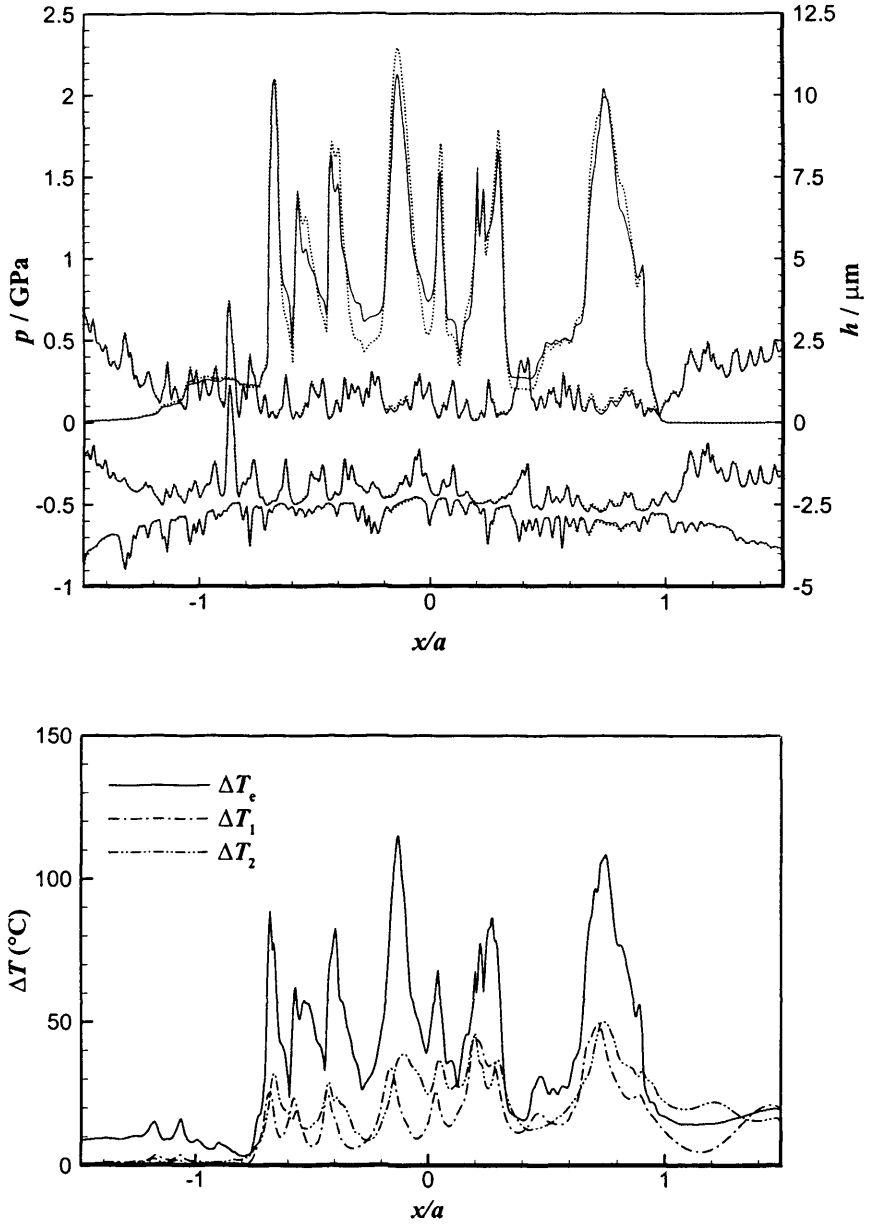


Figure 3.26: Pressure, film thickness and temperature rise distributions at one timestep for the contact between two rough surfaces each with profile (C). Pressure and film thickness distribution for corresponding isothermal case is also shown in the upper figure in dotted lines for comparison.

3.5 Summary

A solution method for thermal effects in transient, rough surface EHL line contacts has been presented. The numerical formulation is based upon the full energy equation for the fluid film and solid surfaces. A control volume approach, which includes a domain transformation in the solid surfaces, is used within an efficient solution scheme.

The formulation includes the effect of transitory "dry" contacts between asperities in which the fluid film breaks down and heating is due to Coulomb friction. Results are presented which demonstrate thermal effects in disc machine tests using roughness profiles acquired from the surfaces used. Under realistic operating conditions or load, rolling and sliding speed, temperature increases of the order 100°C are predicted, but these changes do not produce a significant effect on the pressure and film thickness profiles within the contact.

Chapter 4

Elastic Stress Analysis

4.1 Introduction

When two bodies in contact under load are non-conforming, the resulting contact areas are very small and the contact pressures are very high. Therefore it is essential to know the values of stresses acting in such contacts in order to assist machine design. For the Hertzian contact, these stresses can be determined from the analytical formulae as described in Chapter 1. For the real surfaces which are rough rather than smooth, the contact stresses will significantly differ from the Hertzian stress distribution. Therefore, the analysis of roughness-induced stresses is a fundamental problem in the study of mixed lubrication.

The solution to a rough contact problem is to find the surface displacement that satisfies the condition of compatibility of deformation and the associated contact pressure that satisfies the equilibrium condition. As described by Liu et al. (1999), three aspects of work for rough contact problems are needed: (a) descriptions of rough surface profiles; (b) expressions for the pressure-deformation relationship; and (c) methods and techniques to solve the contact equations. A large number of works on solving rough contact problems have been published in the literature (see review papers by Bhushan (1998), Liu et al. (1999) and Adams and Nosonovsky (2000)). A detailed account of the principles involved in using numerical elastic contact techniques on measured rough surfaces was presented by Sayles (1996). More recently, Allwood (2005) reviewed solution methods for elastic rough contact problems and assessed their performance with emphasis on their solution speed, accuracy and memory requirements.

Once the pressure distribution is determined, the calculation of subsurface stress fields for a line contact problem is straightforward using the formulae given by Johnson (1985). Bailey and Sayles (1991) evaluated the complete subsurface stress field under a pressure distribution obtained using the contact model of Webster and Sayles (1986) for real rough surface contacts. Their studies have shown that roughness-induced stresses are significantly higher than the maximum stress in a comparable smooth contact. The same contact model was also used by Kim and Olver (1998) to determine the pressure distribution in the contact. The pressure distribution was then used in turn to determine the stress distributions and the stress histories of a specimen loaded by a counterbody rolling-sliding over it. Their numerical results show that stresses near the surface layer can be many times higher than in an equivalent smooth contact, and the roughness on the counterbody greatly increases the number of stress cycles during each passage of the contact when sliding is present. Hu et al. (1999) presented a numerical model for the elastic contact of rough surfaces based on fast Fourier transform (FFT) and minimisation of complementary energy. The subsurface stresses caused by the contact were analysed and the results showed that high frequency roughness could lead to very sharp impulses and significant oscillations in the pressure distribution.

Recently, many contact models have been developed to address more complex rough contact problems such as layered contact, three-dimensional contact, elastic-plastic contact, etc. Yu and Bhushan (1996) calculated surface and subsurface stresses for nominally flat on flat rough surface contact. Three-dimensional rough surfaces were generated by the computer with specified standard deviation of surface heights and correlation length. Mao et al. (1997) developed a numerical technique for evaluating the subsurface stresses arising from the two-dimensional sliding contact of multilayered elastic bodies with rough surfaces under both normal and tangential forces. Nogi and Kato (1997, 2002) investigated the influence of a hard surface layer on the limit of elastic contact and concluded that the limit of elastic contact is highly dependent on the layer thickness and the surface roughness. Kadiric et al. (2003) employed a layered contact model developed by Cole and Sayles (1992), which is an extension of the numerical rough surface contact model of Webster and Sayles (1986), to investigate the effects of a coating layer in the substrate on the contact pressure and subsurface stress distributions. Peng and Bhushan (2001, 2002) investigated three-

dimensional contact behaviour of layered elastic/plastic solids with rough surfaces. The surface deformation and pressure distributions were obtained based on a variation principle. The surface and subsurface stresses were determined and the von Mises stress yield criterion used to determine the onset of yield. Mihailidis et al. (2001) presented a model to determine the asperity pressure distribution at the contact of two rough cylinders, assuming elastic-perfectly plastic deformation. The effect of roughness on the subsurface stress field was investigated using numerically generated rough surfaces. Mayeur et al. (1995) presented a numerical elasto-plastic model for rough contact based on a boundary integral formulation. The material was assumed to obey the von Mises yield criterion with linear kinematic hardening. Liu et al. (2001a) developed an elasto-plastic rough contact model based on an elasto-plastic finite element method (FEM) with the incremental form of a simplex-type algorithm. They also used the von Mises yield criterion to determine the onset of plastic deformation. Kogut and Etsion (2003) studied the contact of rough surfaces based on the Greenwood-Williamson plasticity index Ψ and found that up to $\Psi = 0.6$ the contact is purely elastic; $\Psi = 1.4$ marks the transition from entirely elastic to elastic-plastic; and $\Psi = 8$ marks the transition from elastic-plastic to fully plastic behaviour.

In a rough surface EHL contact, an appropriate EHL analysis is used to calculate the pressure, surface shear stress, and film thickness profiles in the contact area. The calculated pressure and surface shear stress are then used to evaluate the surface and subsurface stress distributions in the contacting bodies. The stresses are affected by the lubricating film separating the opposing rough surfaces and the level of rolling and sliding. Elsharkawy and Hamrock (1991) studied subsurface stresses in micro-EHL line contacts using a simple sinusoidal wavy surface. The effect of the amplitude and wavelength of the surface on the pressure, surface shear stress and film thickness profiles, and hence on the subsurface stress pattern in the solids were presented. Tao et al. (2003) presented numerical results for EHL of gear teeth using real surface profiles obtained from micropitting tests carried out on an FZG gear testing machine. The EHL model adopted was based on the line contact model of Elcoate et al. (2001) that solves the elastic and hydrodynamic film thickness equations simultaneously. The subsurface maximum shear stress was found to be much higher than for the corresponding smooth surface case and is close to the surface. Hooke and Li (2002, 2003)

presented an inverse approach that allows the pressures and stresses predicted by EHL analyses to be checked. A Fourier transform method was applied to calculate stresses from pressures in order to reduce computational time. Zhu et al. (2003) performed transient contact stress analyses for different types of rough surfaces in mixed lubrication problems of point contacts.

To reduce the computational time, fast numerical methods have been utilized for solving rough contact equations and subsequent stress evaluation. Polonsky and Keer (1999) developed a fast method for the numerical analysis of rough contacts by combining a multilevel multisummation (MLMS) algorithm with an iteration scheme based on the conjugate gradient method (CGM). Another fast method was developed by Nogi and Kato (1997) by combining the fast Fourier transform (FFT) with a CGM-based iteration scheme. A serious limitation imposed on FFT-based methods is that they are not strictly applicable to nonperiodic contact problems. FFT application to such problems introduces a periodicity error in the numerical solution. The periodicity error can be reduced by extending the contact domain, as suggested by Nogi and Kato (1997), Hooke and Li (2002), etc. Polonsky and Keer (2000) compared MLMS and FFT by applying them to an example contact problem. Colin and Lubrecht (2001) also compared the two methods for elastic deformation calculations. It has been shown that to achieve accuracy comparable to that of MLMS, the grid on which FFT is performed needs to be extended far beyond the contact area, which results in a dramatic increase in the computation time. To overcome periodical extension error associated with FFT, Ai and Sawamiphakdi (1999) decomposed the solution into two parts: a smooth part and a fluctuation part. The smooth part is solved analytically and the fluctuation part is solved by using a FFT-based method. Liu et al. (2000) presented a discrete convolution and fast Fourier transform (DC-FFT) algorithm, which avoids any error associated with FFT at a cost of only doubling the domain dimensions of a contact problem. The DC-FFT algorithm has been used to calculate contact stress fields by Liu and Wang (2002) and Zhu et al. (2003).

In this chapter, the stress distributions and stress histories in the contact bodies are determined for the EHL line contact with rough surfaces. The two contacting bodies are assumed to behave elastically so that their stress states remain elastic. The stress fields are

calculated numerically. The numerical method adopted is chosen based on its accuracy and efficiency as determined in comparisons with other methods.

4.2 Mathematical Modeling

In the EHL line contact problem, the stresses may be calculated to a good approximation by considering each body as a semi-infinite elastic solid (Johnson 1985). The semi-infinite solid is shown in cross-section in Figure 4.1. The boundary surface is the x-y plane and the z-axis is directed into the solid. Surface pressure $p(x)$ and shear stress $q(x)$ act on the surface over the region from x_1 to x_2.

The pressure $p(x)$ can be obtained by solving the EHL line contact problem. The shear stress $q(x)$ is determined as part of the solution. Since the mid-plane shear stress τ_m in the film can be determined from Equation (2.9) in Chapter 2, the shear stresses existing in the film at the two surfaces are

$$\tau_1 = \tau_m + \frac{h}{2}\frac{\partial p}{\partial x} \tag{4.1a}$$

$$\tau_2 = \tau_m - \frac{h}{2}\frac{\partial p}{\partial x} \tag{4.1b}$$

where h is the film thickness. Therefore, the shear stresses acting on the solid surfaces are

$$q_1 = -\tau_1 \tag{4.2a}$$

$$q_2 = \tau_2 \tag{4.2b}$$

If direct contact occurs and no film exists at a point, the shear stress at that point is determined as the friction force of metal on metal:

$$q_1 = -\mu \cdot p \tag{4.3a}$$

$$q_2 = \mu \cdot p \tag{4.3b}$$

where μ is the coefficient of friction in direct contact. In this thesis, $\mu = 0.1$ is assumed unless specified otherwise. Different signs are added in the above equations assuming the bodies have velocities with $u_1 > u_2$.

In a line contact the semi-infinite body may be considered to be in plane strain since the thickness of the solid is large compared with the width of the contact region (Johnson 1985). Thus, the state of stress at any point is determined by the three components, σ_x, σ_z and τ_{xz}. These stress components at any subsurface point (i.e. $z > 0$) due to the complete distribution of $p(x)$ and $q(x)$ are given by (Johnson 1985)

$$\sigma_x = -\frac{2z}{\pi}\int_{x_1}^{x_2}\frac{p(s)(x-s)^2 ds}{\left[(x-s)^2+z^2\right]^2} - \frac{2}{\pi}\int_{x_1}^{x_2}\frac{q(s)(x-s)^3 ds}{\left[(x-s)^2+z^2\right]^2} \tag{4.4a}$$

$$\sigma_z = -\frac{2z^3}{\pi}\int_{x_1}^{x_2}\frac{p(s)ds}{\left[(x-s)^2+z^2\right]^2} - \frac{2z^2}{\pi}\int_{x_1}^{x_2}\frac{q(s)(x-s)ds}{\left[(x-s)^2+z^2\right]^2} \tag{4.4b}$$

$$\tau_{xz} = -\frac{2z^2}{\pi}\int_{x_1}^{x_2}\frac{p(s)(x-s)ds}{\left[(x-s)^2+z^2\right]^2} - \frac{2z}{\pi}\int_{x_1}^{x_2}\frac{q(s)(x-s)^2 ds}{\left[(x-s)^2+z^2\right]^2} \tag{4.4c}$$

On the surface (i.e. $z = 0$), the stress components can be determined by the boundary conditions, i.e.

$$\sigma_z = -p(x) \tag{4.5a}$$

$$\tau_{xz} = -q(x) \tag{4.5b}$$

σ_x on the surface can be determined by the strain and Hooke's law, i.e.

$$\begin{cases} \varepsilon_x = -\dfrac{(1-2\nu)(1+\nu)}{E}p(x) - \dfrac{2(1-\nu^2)}{\pi E}\displaystyle\int_{x_1}^{x_2}\frac{q(s)}{x-s}ds \\ \varepsilon_x = \dfrac{1}{E}\left[(1-\nu^2)\sigma_x - \nu(1+\nu)\sigma_z\right] \end{cases} \tag{4.6}$$

where E is the Young's modulus and ν is the Poisson ratio. This gives

$$\sigma_x = -p(x) - \frac{2}{\pi}\int_{x_1}^{x_2} \frac{q(s)}{x-s}\,ds \tag{4.7}$$

The integrand in Equation (4.7) has a singularity at $s = x$ and changes sign. The value required is the Cauchy principal value of the integral.

Under conditions of plane strain, $\tau_{xy} = \tau_{yz} = 0$ and

$$\sigma_y = \nu(\sigma_x + \sigma_z) \tag{4.8}$$

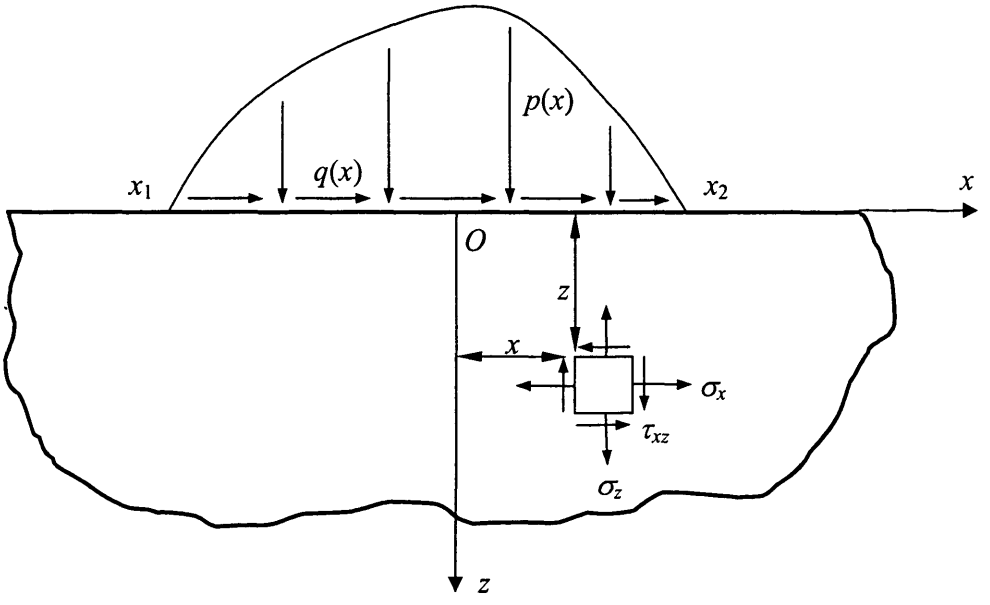


Figure 4.1: Semi-infinite solid in line contact.

4.3 Numerical Methods

Evaluating the integrals in Equation (4.4) is inherently time consuming, especially for a fatigue analysis such as is presented later that requires a transient stress analysis procedure. Therefore, seeking an efficient and accurate numerical method to determine the stress

components throughout the contacting bodies is necessary. In the following, three methods are evaluated, namely the Direct Integration Method, the Hooke-Li Method and a DC-FFT Method.

4.3.1 Direct Integration Method

In this method, the integrals in Equation (4.4) can be divided into a set of solutions for the uniform load steps p_i and q_i. It is relatively straightforward to calculate the stresses due to each step with uniform load and the actual stresses are obtained by summation of these solutions. Performing the integrations and using the notation shown in Figure 4.2, Johnson (1985) gives equations for the stresses as

$$\sigma_x = \sum_i \left\{ -\frac{p_i}{2\pi}\left[2(\theta_1-\theta_2)+(\sin 2\theta_1-\sin 2\theta_2)\right]+\frac{q_i}{2\pi}\left[4\ln(r_1/r_2)-(\cos 2\theta_1-\cos 2\theta_2)\right]\right\} \quad (4.9a)$$

$$\sigma_z = \sum_i \left\{ -\frac{p_i}{2\pi}\left[2(\theta_1-\theta_2)-(\sin 2\theta_1-\sin 2\theta_2)\right]+\frac{q_i}{2\pi}(\cos 2\theta_1-\cos 2\theta_2)\right\} \quad (4.9b)$$

$$\tau_{xz} = \sum_i \left\{ \frac{p_i}{2\pi}(\cos 2\theta_1-\cos 2\theta_2)-\frac{q_i}{2\pi}\left[2(\theta_1-\theta_2)+(\sin 2\theta_1-\sin 2\theta_2)\right]\right\} \quad (4.9c)$$

where

$$\tan\theta_{1,2} = \frac{z}{(x-x_i)\mp\frac{\Delta x}{2}} \quad (4.10a)$$

$$r_{1,2} = \sqrt{\left[(x-x_i)\mp\frac{\Delta x}{2}\right]^2+z^2} \quad (4.10b)$$

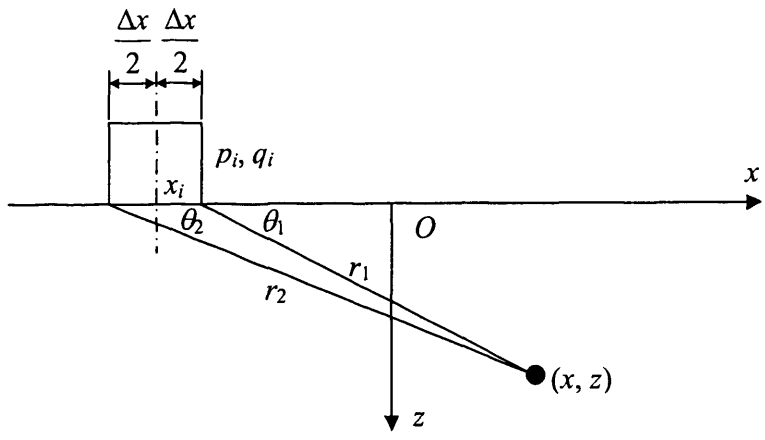


Figure 4.2: Notation used to define subsurface stresses for a uniform load at the surface where surface tractions p_i and q_i are applied at position x_i.

4.3.2 Hooke-Li Method

The Hooke-Li method (Hooke and Li 2002) is a method for evaluating elastic stresses for a semi-infinite line contact solid based on Fast Fourier Transform (FFT) technique.

$P(\omega)$ and $Q(\omega)$ are defined as the Fourier transforms of the pressure $p(x)$ and traction $q(x)$, respectively, i.e.

$$P(\omega) = \int_{-\infty}^{\infty} p(x) e^{-2\pi i x\omega} dx \tag{4.11a}$$

$$Q(\omega) = \int_{-\infty}^{\infty} q(x) e^{-2\pi i x\omega} dx \tag{4.11b}$$

For positive ω, if the Airy stress function is selected as

$$\phi = \frac{\{P(\omega) + 2\pi z\omega[P(\omega) - iQ(\omega)]\} e^{-2\pi z\omega}}{(2\pi\omega)^2} e^{2\pi i x\omega} \tag{4.12}$$

then the stress components vanish as z becomes large and satisfy the equations of equilibrium together with the compatibility equation and the boundary conditions:

$$\sigma_x = \frac{\partial^2 \phi}{\partial z^2}, \quad \sigma_z = \frac{\partial^2 \phi}{\partial x^2}, \quad \tau_{xz} = -\frac{\partial^2 \phi}{\partial x \partial z} \tag{4.13a}$$

$$\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right\}\left\{\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2}\right\} = 0 \tag{4.13b}$$

$$\sigma_z\big|_{z=0} = -P(\omega)e^{2\pi i x\omega}, \quad \tau_{xz}\big|_{z=0} = -Q(\omega)e^{2\pi i x\omega} \tag{4.13c}$$

This yields the stress components as

$$(\sigma_x + \sigma_z)/2 = -[P(\omega) - iQ(\omega)]e^{-2\pi z\omega}e^{2\pi i x\omega} = F_1(\omega)e^{2\pi i x\omega} \tag{4.14a}$$

$$(\sigma_x - \sigma_z)/2 = \{2\pi z\omega[P(\omega) - iQ(\omega)] + iQ(\omega)\}e^{-2\pi z\omega}e^{2\pi i x\omega} = F_2(\omega)e^{2\pi i x\omega} \tag{4.14b}$$

$$\tau_{xz} = i\{2\pi z\omega[P(\omega) - iQ(\omega)] + iQ(\omega)\}e^{-2\pi z\omega}e^{2\pi i x\omega} = F_3(\omega)e^{2\pi i x\omega} \tag{4.14c}$$

Thus, the actual stresses can be obtained via inverse fast Fourier transform (IFFT), i.e.

$$(\sigma_x + \sigma_z)/2 \xleftarrow{\text{IFFT}} F_1(\omega) \tag{4.15a}$$

$$(\sigma_x - \sigma_z)/2 \xleftarrow{\text{IFFT}} F_2(\omega) \tag{4.15b}$$

$$\tau_{xz} \xleftarrow{\text{IFFT}} F_3(\omega) \tag{4.15c}$$

4.3.3 DC-FFT Method

Based on Equation (4.4), we define the following influence coefficients, which are independent of the pressure distribution:

$$c_1(x,z) = -\frac{2}{\pi}\frac{x^2 z}{(x^2 + z^2)^2} \tag{4.16a}$$

$$c_2(x,z) = -\frac{2}{\pi}\frac{z^3}{(x^2+z^2)^2} \tag{4.16b}$$

$$c_3(x,z) = -\frac{2}{\pi}\frac{xz^2}{(x^2+z^2)^2} \tag{4.16c}$$

$$c_4(x,z) = -\frac{2}{\pi}\frac{x^3}{(x^2+z^2)^2} \tag{4.16d}$$

Thus, Equation (4.4) can be rewritten as

$$\sigma_x(x,z) = \int_{x_1}^{x_2} p(s)c_1(x-s,z)ds + \int_{x_1}^{x_2} q(s)c_4(x-s,z)ds \tag{4.17a}$$

$$\sigma_z(x,z) = \int_{x_1}^{x_2} p(s)c_2(x-s,z)ds + \int_{x_1}^{x_2} q(s)c_3(x-s,z)ds \tag{4.17b}$$

$$\tau_{xz}(x,z) = \int_{x_1}^{x_2} p(s)c_3(x-s,z)ds + \int_{x_1}^{x_2} q(s)c_1(x-s,z)ds \tag{4.17c}$$

It can be seen that the stress components can be obtained by calculating the convolution of normal and tangential tractions with influence coefficients. We assume that there are N points in the x domain and the mesh size is Δx. Then Equations (4.17) can be written in the following form of discrete convolution:

$$\sigma_x(i,j) = \sum_{k=0}^{N-1}\{p_k c_1(i-k,j) + q_k c_4(i-k,j)\}\Delta x \tag{4.18a}$$

$$\sigma_z(i,j) = \sum_{k=0}^{N-1}\{p_k c_2(i-k,j) + q_k c_3(i-k,j)\}\Delta x \tag{4.18b}$$

$$\tau_{xz}(i,j) = \sum_{k=0}^{N-1}\{p_k c_3(i-k,j) + q_k c_1(i-k,j)\}\Delta x \tag{4.18c}$$

It can be seen that the discrete convolution approximates the continuous convolution within the error introduced by rectangular integration. If the mesh size Δx is made sufficiently small, then the error introduced by the discrete convolution is negligible (Brigham 1988).

By taking the Fourier transform, the convolution of any two functions in the space domain x becomes the product of their individual transforms in the new domain ω. Therefore the stress components can be calculated efficiently by using FFT and inverse FFT (IFFT) as follows:

$$\sigma_x(x,z) \xleftarrow{\text{IFFT}} P(\omega)C_1(\omega,z) + Q(\omega)C_4(\omega,z) \tag{4.19a}$$

$$\sigma_z(x,z) \xleftarrow{\text{IFFT}} P(\omega)C_2(\omega,z) + Q(\omega)C_3(\omega,z) \tag{4.19b}$$

$$\tau_{xz}(x,z) \xleftarrow{\text{IFFT}} P(\omega)C_1(\omega,z) + Q(\omega)C_4(\omega,z) \tag{4.19c}$$

where

$$P(\omega) \xleftarrow{\text{FFT}} p(x) \tag{4.20a}$$

$$Q(\omega) \xleftarrow{\text{FFT}} q(x) \tag{4.20b}$$

$$C_1(\omega,z) \xleftarrow{\text{FFT}} c_1(x,z) \tag{4.20c}$$

$$C_2(\omega,z) \xleftarrow{\text{FFT}} c_2(x,z) \tag{4.20d}$$

$$C_3(\omega,z) \xleftarrow{\text{FFT}} c_3(x,z) \tag{4.20e}$$

$$C_4(\omega,z) \xleftarrow{\text{FFT}} c_4(x,z) \tag{4.20f}$$

The Discrete Convolution and FFT (DC-FFT) method stated by Liu et al. (2000) is adopted to calculate stresses in Equation (4.18). Assume there are N uniformly spaced points

in the x direction. To apply the FFT algorithm for a particular value of z, usually N is an integer of power of 2. If this is not the case, the domain is expanded to a power of 2 with zero-padding. In the z direction, a uniform distribution of mesh points is not necessary. Therefore, the z values can be selected in such a way that dense mesh points are placed in the near-surface region and sparse mesh points are placed in the deeper region. The steps of the procedure for stress calculation using DC-FFT are summarised as follows:

(1) Expand pressures $\{p\}$ and $\{q\}$ to a virtual domain of $2N$ with zero-padding only, as illustrated in Figure 4.3.

(2) Expand coefficients $\{c_1\}$, $\{c_2\}$, $\{c_3\}$ and $\{c_4\}$ to a virtual domain of $2N$ with zero-padding for the term with index $i = N$ and with wrap-around order for the terms of $i \in [N+1, 2N-1]$, i.e.

$$c_i = 0,\ i = N; \qquad c_i = c_{i-2N},\ i \in [N+1, 2N-1]$$

where c represents c_1, c_2, c_3 and c_4. Figure 4.3 shows both original and virtual domains for all the series of $\{p\}$, $\{q\}$, $\{c_1\}$, $\{c_2\}$, $\{c_3\}$ and $\{c_4\}$.

(3) Apply FFT to $\{p\}$, $\{q\}$, $\{c_1\}$, $\{c_2\}$, $\{c_3\}$ and $\{c_4\}$ as described by Equation (4.20) in the virtual domain to obtain $\{P\}$, $\{Q\}$, $\{C_1\}$, $\{C_2\}$, $\{C_3\}$ and $\{C_4\}$.

(4) Form the right hand sides of Equation (4.19) by complex multiplication of the terms $\{P\}$, $\{Q\}$, $\{C_1\}$, $\{C_2\}$, $\{C_3\}$ and $\{C_4\}$.

(5) Apply inverse FFT (IFFT) as described by Equation (4.19) and discard the terms with index $i \in [N, 2N-1]$ to obtain the stress components.

Since coefficients c_1, c_2, c_3 and c_4 are only dependent on the locations (x, z), calculations of these coefficients and their Fourier transforms only need to be executed once during a transient stress analysis and they can be stored as global data for later use.

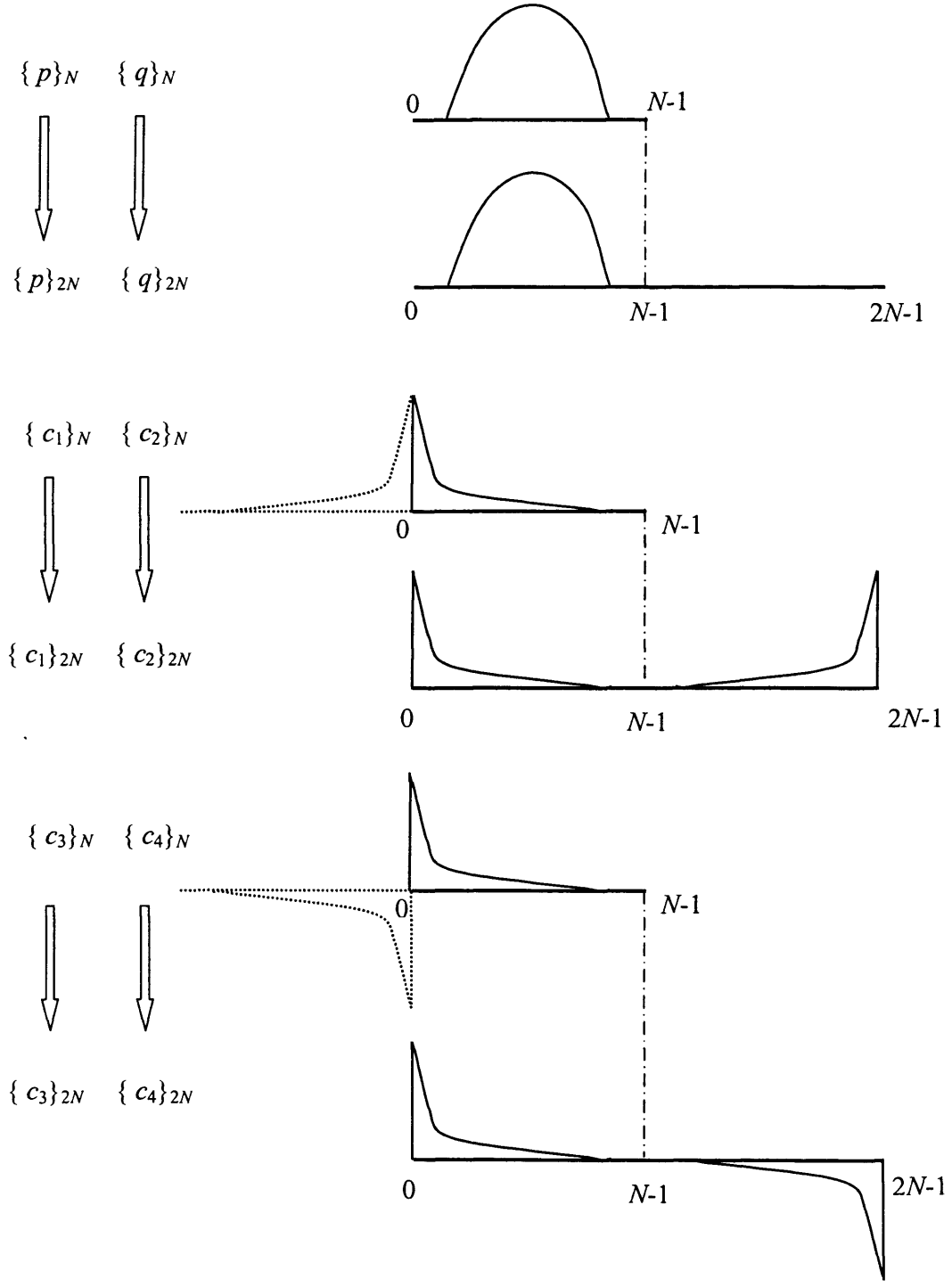


Figure 4.3: Illustration of zero-padding and wrap-around order to apply DC-FFT method.

4.3.4 Comparison of the Methods

In this section, comparisons of the above three candidate methods considered, namely the Direct Integration Method (DIM), the Hooke-Li method and the DC-FFT method, are made to demonstrate their accuracy and efficiency. In the following, the principal shear stress is investigated for comparison, which can be expressed by

$$\tau_1 = \sqrt{\left(\frac{\sigma_x - \sigma_z}{2}\right)^2 + \tau_{xz}^2} \tag{4.21}$$

4.3.4.1 Comparison between DIM and the Hooke-Li Method

Figure 4.4 shows the contours of τ_1 under Hertzian pressure obtained from the Hertzian theory, DIM and the Hooke-Li method. The mesh size is $\Delta x = \Delta z = 0.01a$ and the number of nodes is $N \times M = 301 \times 200$. The Hertzian pressure is $p_0 = 1$GPa. In implementing the Hooke-Li method, the number of pressure points is expanded to 10 times the actual number with zero padding to ensure accuracy (Hooke and Li 2002). It can be seen that the DIM is more accurate than the Hooke-Li method since the DIM only involves discretisation error whereas the Hook-Li method also involves periodic error. The CPU time of the two methods for different mesh size is summarised in Table 4.1. With finer mesh sizes, the Hooke-Li method takes less CPU time. It should be noted that finer discretisation does not improve the accuracy due to periodic error (Polonsky and Keer 2000). To improve accuracy, more zero padding is needed, but efficiency will inevitably be reduced.

In the Hook-Li method, the computational domain has to be expanded to several times the physical domain to reduce the periodic error in the target domain. To reduce the periodic error further, the real EHL pressure can be divided into two parts: Hertzian pressure and deviation. For the Hertzian pressure, the stresses can be obtained analytically. For the deviation part, the Hook-Li method is applied. The stress components can then be obtained by addition.

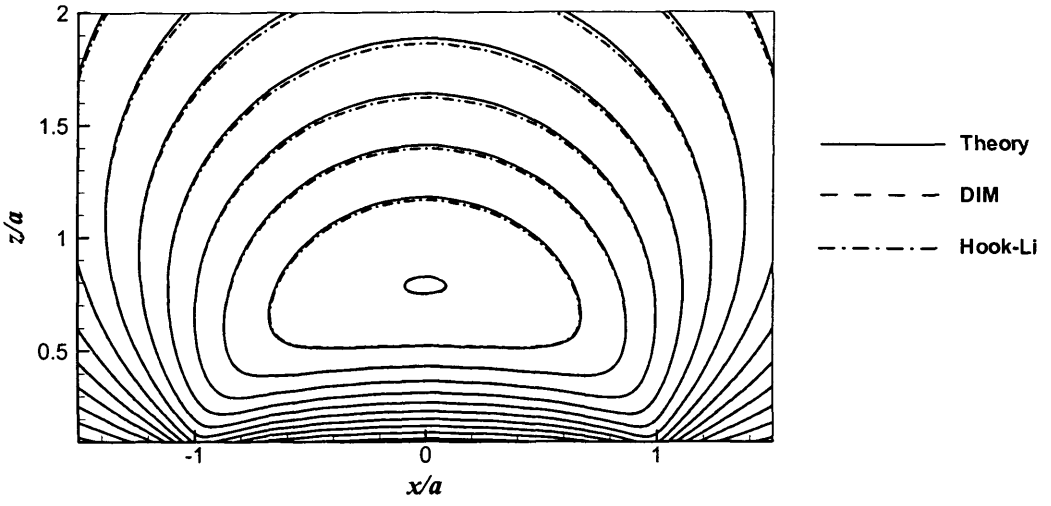


Figure 4.4: Contours of the principal shear stress under Hertzian pressure using the Hertzian theory, DIM and the Hook-Li method. (Note that solid and dashed contours are coincident.)

Table 4.1: CPU time (seconds) of DIM and Hooke-Li under Hertzian pressure.

Mesh	CPU time (s)		
	DIM (t_1)	Hooke-Li (t_2)	t_1/t_2
$\Delta x = 0.01$, $\Delta z = 0.01$, $N \times M = 301 \times 200$	7.70	11.4	0.674
$\Delta x = 0.005$, $\Delta z = 0.01$, $N \times M = 601 \times 200$	29.7	25.9	1.14
$\Delta x = 0.0025$, $\Delta z = 0.01$, $N \times M = 1201 \times 200$	118.3	54.9	2.16
$\Delta x = 0.00025$, $\Delta z = 0.01$, $N \times M = 12001 \times 200$	12768	509	25.1

Figure 4.5 shows EHL pressure distributions for smooth surface and rough surface examples. The two parts of the pressures, Hertz and deviation, are also given. The mesh size is $\Delta x = 0.02$, $\Delta z = 0.1$ and the number of nodes is $N \times M = 201 \times 20$.

Figure 4.6 is the contour of τ_1 for the smooth pressure shown in Figure 4.5. Solid lines are DIM results, and the other two are the results obtained from the summation of the Hertzian part and the deviation part to which the Hook-Li method is applied. Dashed lines are the results without domain expansion and dash-dotted lines are the results with double-sized domain.

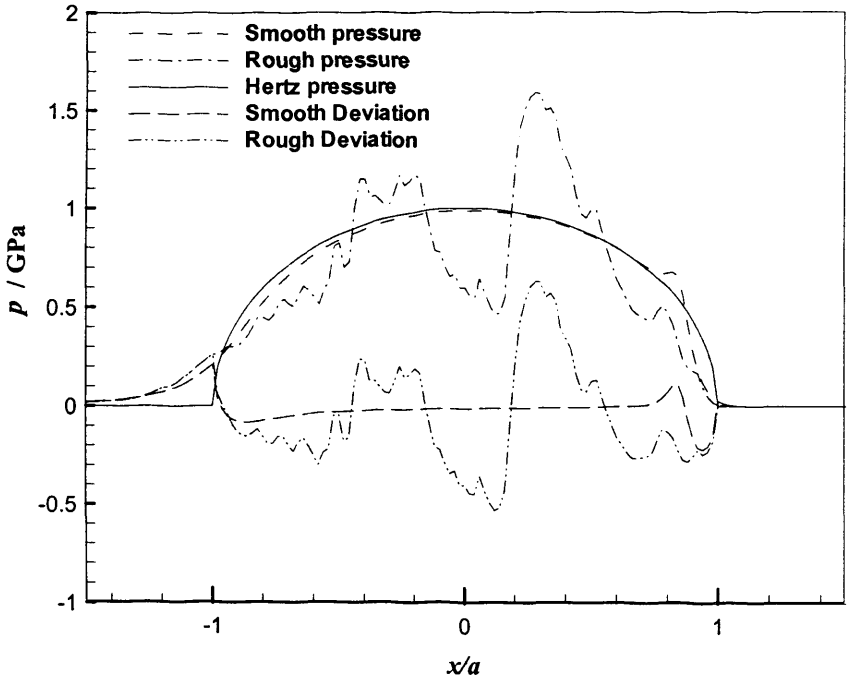


Figure 4.5: EHL pressure distributions.

Figure 4.7 is the contour of τ_1 for the rough pressure shown in Figure 4.5. Solid lines are DIM results, and the other two are the results obtained from summation of the Hertz part and the deviation part to which the Hook-Li method is applied. Dashed lines are the results

with double-sized domain expansion and dash-dotted lines are the results with 4 times expansion.

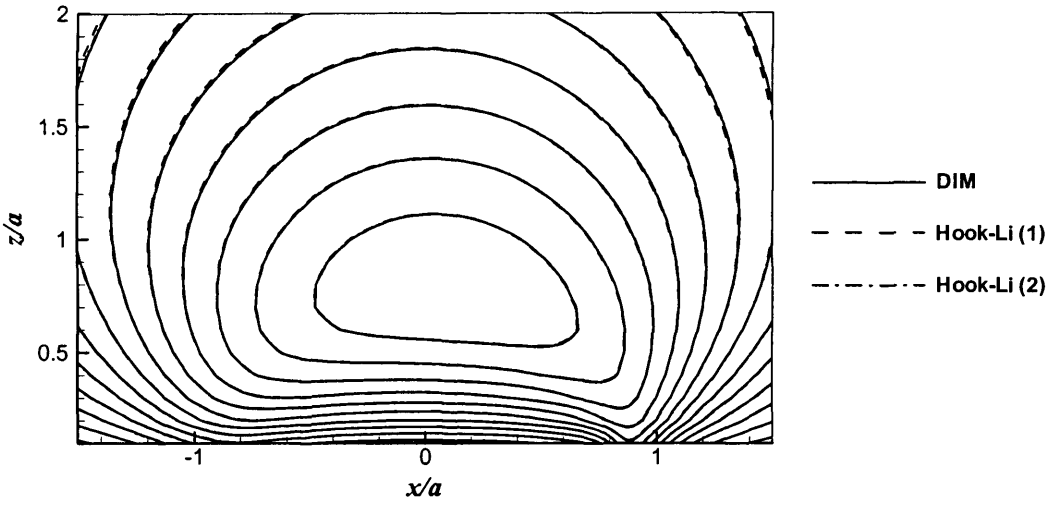


Figure 4.6: Contours of the principal shear stress under smooth pressure using DIM and the Hook-Li method (the number in the bracket is the factor of domain expansion).

The CPU time is summarised in Table 4.2. The number in parentheses indicates the size of computation domain to which the methods are applied.

Table 4.2: CPU time (seconds) of DMM and Hooke-Li with EHL pressure for case with $\Delta x = 0.02$, $\Delta z = 0.1$ and $N \times M = 201 \times 20$.

Smooth pressure			Rough pressure		
DIM ($N \times M$)	Hook-Li ($N \times M$)	Hook-Li ($2N \times M$)	DIM ($N \times M$)	Hook-Li ($2N \times M$)	Hook-Li ($4N \times M$)
0.15	0.03	0.05	0.15	0.05	0.12

From these simulation results, it can be seen that after decomposing the EHL pressure into two parts, the Hook-Li method's accuracy can be improved with less zero-padding so that the efficiency can be improved too. More zero-padding is required for rough pressures than for smooth pressures. In improving its accuracy further, the efficiency of the Hoook-Li method will be reduced, however.

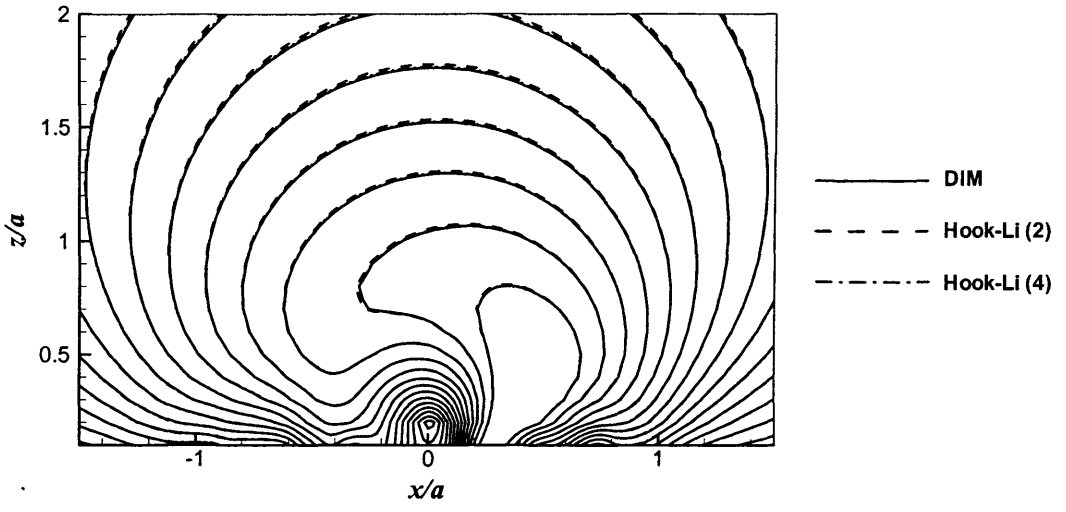


Figure 4.7: Contours of the principal shear stress under rough pressure using DIM and the Hook-Li method (the number in the bracket is the factor of domain expansion.).

4.3.4.2 Comparison between DIM and DC-FFT

Figure 4.8 shows the contours of τ_1 using DIM and DC-FFT for four cases considered: (a) Hertzian pressure, (b) EHL smooth pressure, (c) EHL rough pressure, and (d) EHL rough pressure with surface traction $q(x) = \mu \cdot p(x)$, $\mu = 0.05$. The smooth and rough pressures are those shown in Figure 4.5. For all cases, both methods have the same level of accuracy in that the contours drawn are coincident. Regarding the CPU time, cases of different mesh size, running on different machines using both methods were tested and the results are summarised in Table 4.3. It should be noted that the time listed in Table 4.3 does

not include the time for calculating the coefficients c_1, c_2, c_3 and c_4 and their Fourier transforms. This time is a fixed overhead as the coefficients need only be calculated once in a transient stress analysis. The CPU time of DIM can be 7~9 times the CPU time of DC-FFT for N = 1201. From Table 4.1, the CPU time for DIM is about twice the CPU time for the Hook-Li method for N = 1201. Therefore, DC-FFT is found to be both more accurate and more efficient than the Hook-Li method and will be applied for later transient stress analysis.

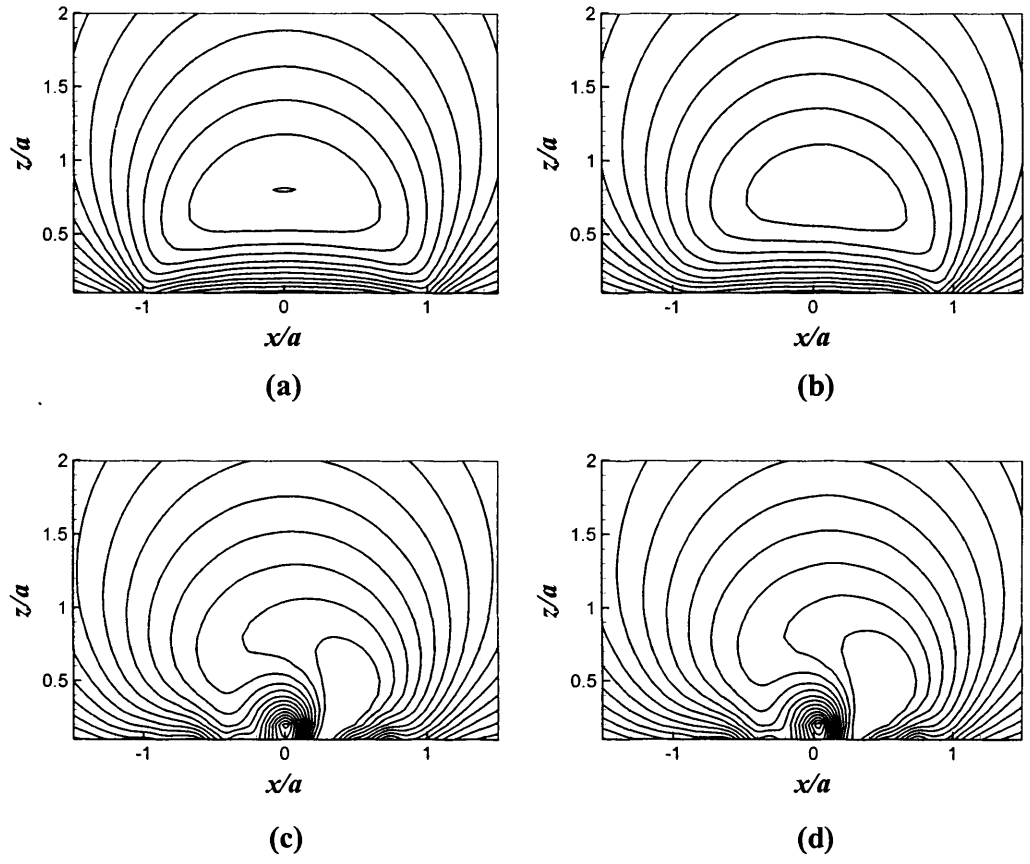


Figure 4.8: Contours of the principal shear stress using DIM (solid lines) and the DC-FFT method (dashed lines) under (a) Hertzian pressure, (b) EHL smooth pressure, (c) EHL rough pressure and (d) EHL rough pressure with friction. In all cases, the contours are co-incident.

Table 4.3: CPU time (seconds) of DIM and DC-FFT using three different computer systems.

Number of nodes	PC			Workstation 1			Workstation 2		
	DIM (t_1)	DC-FFT (t_2)	t_1/t_2	DIM (t_1)	DC-FFT (t_2)	t_1/t_2	DIM (t_1)	DC-FFT (t_2)	t_1/t_2
$N \times M =$ 201×20	0.06	0.03	2.0	0.033	0.017	1.941	0.011	0.005	2.2
$N \times M =$ 601×20	0.591	0.11	5.372	0.225	0.088	2.557	0.095	0.023	4.13
$N \times M =$ 1201×20	2.283	0.25	9.132	1.278	0.196	6.52	0.384	0.057	6.737

4.4 Transient Stress Analysis Using DC-FFT

For transient EHL line contact analysis with rough surfaces, both normal pressure p and tangential traction q are functions of time t and consequently so are the stresses. The coefficients k_1, k_2, k_3 and k_4 are not time dependent so that FFT calculations for the coefficients only need to be executed once during the transient analysis process if stored. The FFT for p and q, the convolutions and inverse FFT are calculated for each timestep.

The operating conditions and lubricant parameters adopted for transient line contact analyses are specified in Table 4.4. The speeds for the two surfaces are u_1 = 50m/s and u_2 = 25m/s, respectively. This corresponds to conditions used in scuffing tests by Patching et al. (1995) for ground and superfinished hardened steel disks operating at high sliding speeds and lubricated with a gas turbine engine oil at a temperature of 100°C. Thermal effects are not considered in stress analysis and subsequent fatigue analysis since the pressure distributions are essentially unchanged by the inclusion of the thermal analysis. Both surfaces have the rough profile (C) which is shown in Figure 2.2 of Chapter 2. It is the most aggressive of the three used by Holmes et al. (2004). The transient stress analyses are then performed and a stress state of plane strain is examined in the form of von Mises equivalent stress:

$$\sigma_{vm} = \frac{1}{\sqrt{2}}\sqrt{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2 + 6\tau_{xz}^2} \tag{4.22}$$

Table 4.4: Parameters used for transient line contact analysis.

a	0.335 mm	γ	2.27 GPa^{-1}
E_1, E_2	207 GPa	κ	63.2×10^{-6} Pa·s
p_0	1.0 GPa	λ	1.68 GPa^{-1}
R'	19.1 mm	ν_1, ν_2	0.3
W'	527 kN/m	τ_0	10 MPa
α	11.1 GPa^{-1}	χ	5.1 GPa^{-1}
η_0	0.0048 Pa·s		

4.4.1 Stress Distribution in the Contact Area

The transient pressure and shear stress at the lubricant/solid interface developed in the EHL contact are determined from transient EHL line contact analysis. The contours of the von Mises equivalent stress, σ_{vm}, for both bodies are shown in Figure 4.9. The corresponding pressure and shear stresses at a particular timestep are also shown in the figure. It can be seen that the stress distribution in the contact area of the two bodies is effectively the same at each instant in time. This is because for full film situations the shear stresses, q_1 and q_2, on the two surfaces are small compared to the pressure p and their magnitudes are about 3%-5% of p values. The only difference in loading of the two surfaces is seen in the values of these surface shear stresses. Also shown in Figure 4.9 is σ_{vm} for the corresponding smooth surface EHL solution. Comparing the plots indicates that the effects of the pressure deviations are most apparent close to the surface of the material, as might indeed be expected. Peak values for σ_{vm} of 0.97 GPa, 0.88 GPa, 0.73 GPa and 0.69 GPa are associated with the localised high pressures at x/a = -0.82, -0.22, 0.04 and 0.4, respectively.

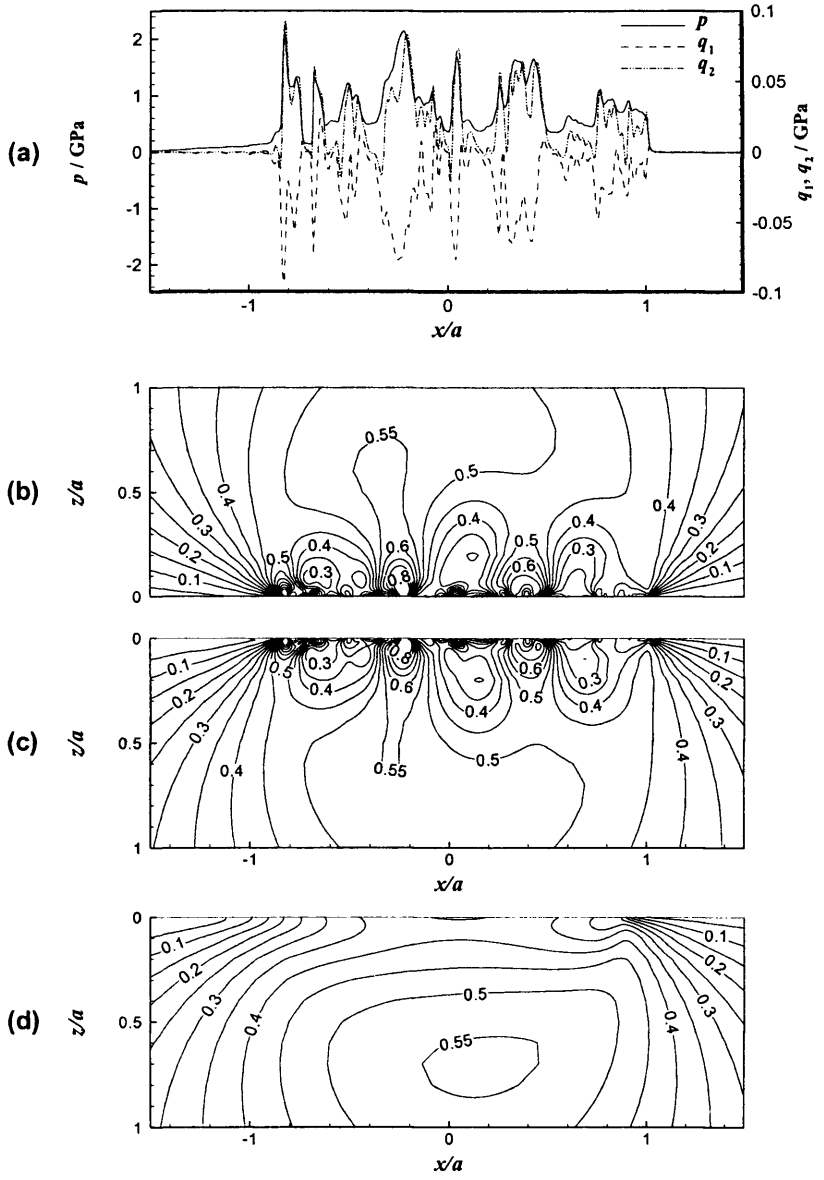


Figure 4.9: Contours of σ_{vm} / GPa for the two rough surfaces which are subject to surface loading. (a) Pressure (p) and surface shear stress (q_1 for the upper surface and q_2 for the lower surface); (b) Contours of σ_{vm} for the upper solid; (c) Contours of σ_{vm} for the lower solid; (d) Contours of σ_{vm} for corresponding smooth-smooth contact case.

4.4.2 Stress History at Fixed Points in the Contacting Materials

Information regarding stress history at fixed points within the material is very important for assessing possible failure of the material. Due to the effect of fluctuations in stress with time, the material becomes vulnerable to fatigue. Therefore the stress history at fixed points in the contacting materials is crucial in attempting to assess the damage caused by EHL contact fatigue.

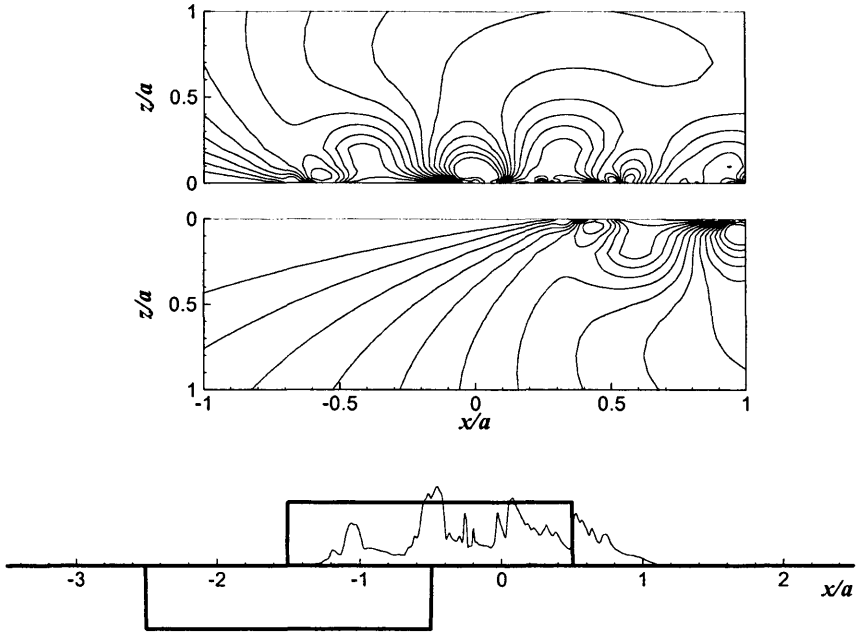
Figure 4.10 shows the variation of stress distributions within two fixed volumes of material with areas of $2a \times a$ for both solids, as they pass through the contact area. The results at four different timesteps are given. The lower part of the figure illustrates the EHL pressure distribution at the illustrated timestep together with the current positions of the two material volumes under consideration.

The stress histories at one asperity location on each body are plotted in Figure 4.11. These are the asperities within the volume considered that experience the greatest stress amplitudes. The lower part of Figure 4.11 is for the slower moving surface. It can be seen that significant cycling of σ_{vm} occurs in this example with amplitudes of at least the magnitude (0.55 GPa) of the maximum smooth surface value. The most aggressive cycling is seen to occur at and just below the surface. Traces with considerable cycling of σ_{vm} are shown at the surface of the asperity and at depths of $0.01a$ (3.4 μm) and $0.05a$ (17 μm). The cycling amplitude has diminished considerably by $0.2a$ (67 μm), and has hardly changed the stress level in comparison with the smooth surface result at depths of $0.5a$ (167 μm) and beyond.

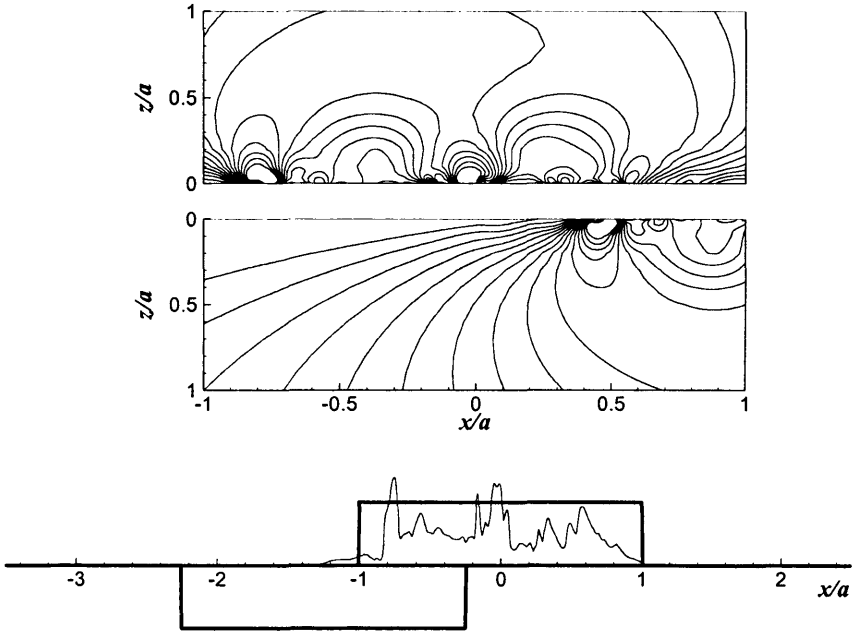
The upper part of Figure 4.11 shows the corresponding picture for the asperity selected on the faster moving surface. The cycling amplitude of σ_{vm} is the same as for the slower moving surface, which is to be expected as one of the asperity collisions whose consequences are seen in these figures takes place between the two asperities in question. Again the cycling amplitude is largest close to the material surface. The significant difference, however, is in the number of stress cycles experienced by asperities on the two

surfaces. For the slower moving surface there are eight cycles and for the faster moving surface there are four (or possibly five) cycles.

(a)



(b)



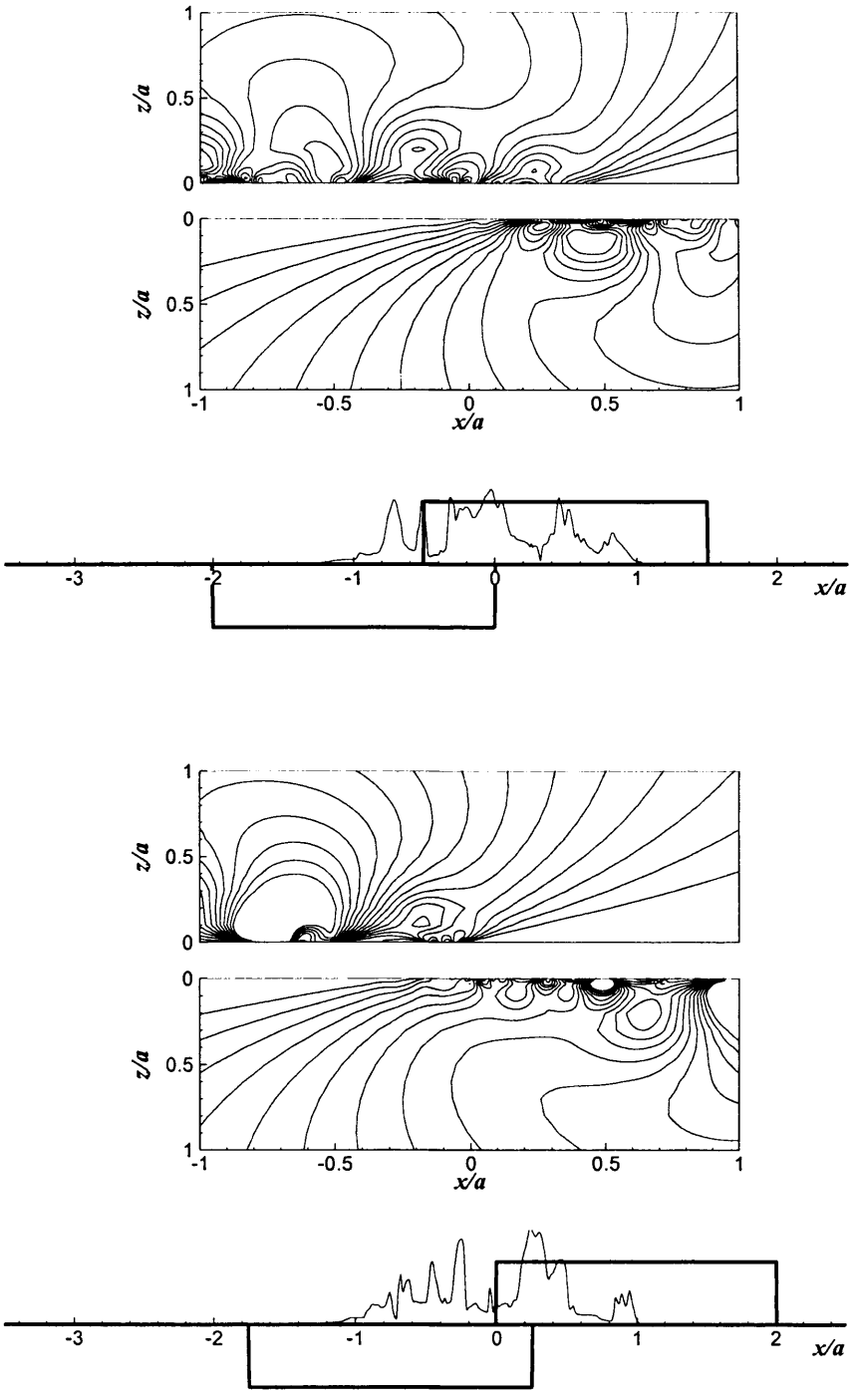


Figure 4.10: Distributions of σ_{vm} / GPa in the $2a \times a$ material volume for both the faster and slower moving surfaces at timesteps: (a) 2200, (b) 2400, (c) 2600 and (d) 2800. The corresponding pressure distribution and the position, relative to the EHL contact, of the material volumes considered are also shown in the figure.

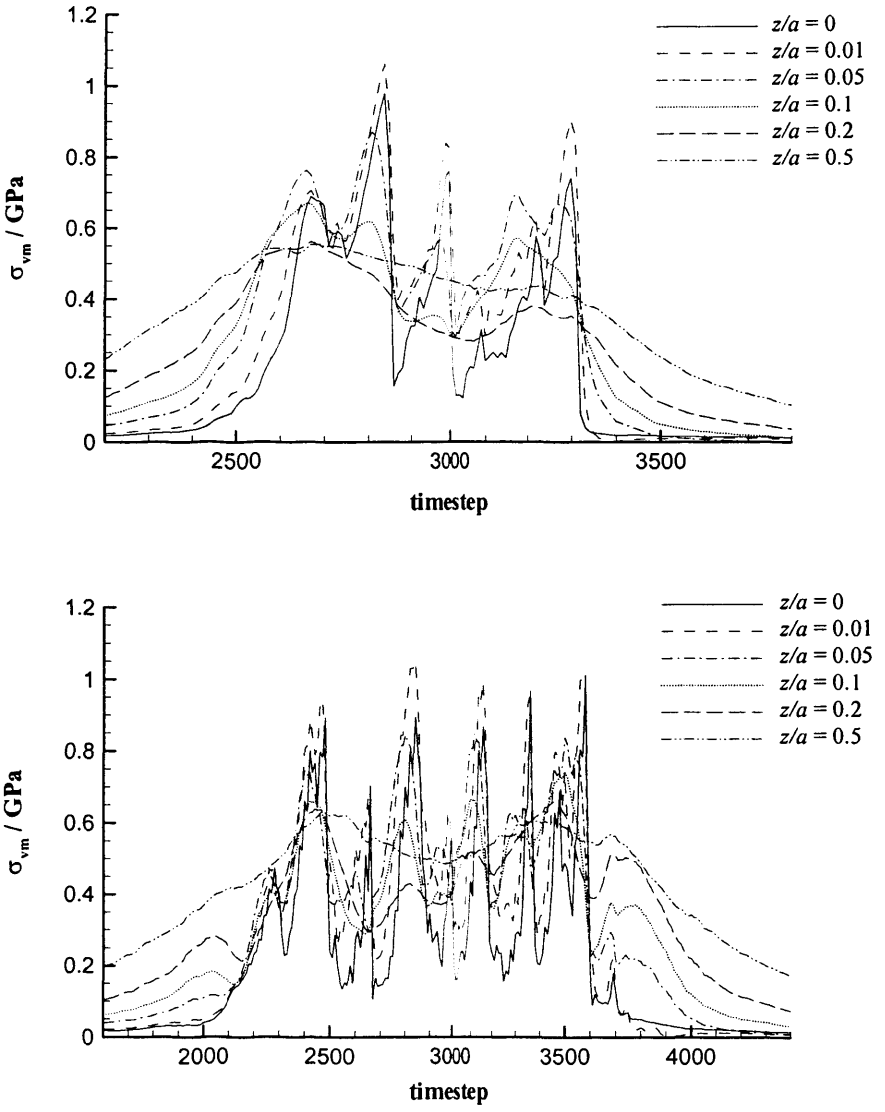


Figure 4.11: Variation of of σ_{vm} / GPa with time during transit of the contact at depths of z/a = 0, 0.01, 0.05, 0.1, 0.2 and 0.5 for two asperities from rough surfaces for contact with ξ = 0.67. The upper figure is the faster surface, and the lower figure is the slower surface.

This situation is demonstrated more clearly by considering sinusoidal surface profiles so that the results form a repeating pattern without being influenced by variation in the detailed shape and spacing of the true rough surface asperities. Figure 4.12 shows the transient variation of σ_{vm} at asperity tips for micro EHL contact between two sinusoidal surfaces with wavelength $a/4$ and amplitude 0.5 μm. The amplitude is chosen to produce significant pressure deviations from the Hertzian with peak pressures of 2.85 GPa being seen during the transient line contact analysis. The pattern described above for Figure 4.11 is seen again in Figure 4.12, albeit more distinctly because of the repetitive nature of the surface profiles. The cycling amplitude of σ_{vm} is again largest close to the surface, and the number of cycles seen by the two surfaces is nine for the slower surface and four (or five) for the faster surface.

The reason for the difference in the number of stress cycles experienced by asperities on the two surfaces may be gleaned from computing the number of times asperity features pass each other during their transit of the Hertzian contact zone. The slide roll ratio is defined as

$$\xi = \frac{u_s}{\bar{u}} = \frac{u_1 - u_2}{(u_1 + u_2)/2} \tag{4.23}$$

The time taken by each component to traverse the Hertzian contact area, i.e. a distance of $2a$, is

$$t_1 = \frac{2a}{u_1}, \qquad t_2 = \frac{2a}{u_2} \tag{4.24}$$

The distance moved relative to the other surface during this time is given by the product of the traverse time and the sliding velocity

$$d_1 = \frac{2au_s}{u_1}, \qquad d_2 = \frac{2au_s}{u_2} \tag{4.25}$$

If λ_1 and λ_2 are the (average) spacing of asperities of micro-EHL significance on the two surfaces, then the number of asperity collisions to be expected during traverse of the contact area will be

$$N_1 = \frac{2au_s}{u_1\lambda_2}, \qquad N_2 = \frac{2au_s}{u_2\lambda_1} \tag{4.26}$$

Written in terms of slide roll ratio, the expressions for N_1 and N_2 become

$$N_1 = \frac{4\xi a}{(2+\xi)\lambda_2}, \qquad N_2 = \frac{4\xi a}{(2-\xi)\lambda_1} \tag{4.27}$$

If these ideas are applied to contacts between sinusoidal surfaces for the case of Figure 4.12 where the slide roll ratio is $\xi = 0.67$, the calculated values are $N_1 = 4$, $N_2 = 8$, which corresponds with the pattern seen in the figure. If different wavelengths are considered for the two surfaces, for example $a/4$ and $a/6$, the time variations of σ_{vm} are shown in Figures 4.13 and 4.14. Figure 4.13 shows the case where $\lambda_1 = a/4$ and $\lambda_2 = a/6$, and the case where $\lambda_1 = a/6$ and $\lambda_2 = a/4$ is shown in Figure 4.14. If a different slide roll ratio is considered, for example $\xi = 0.4$, the variation of σ_{vm} for the case where $\lambda_1 = a/8$ and $\lambda_2 = a/4$ is shown in Figure 4.15. Also shown on the figures are the corresponding values of N_1 and N_2 calculated from Equations (4.27). Table 4.5 summarises the outcome of a number of transient stress analyses of EHL contact between sinusoidal surfaces. The values of N_1 and N_2 calculated from Equations (4.27) and observed in the transient analyses are given in the table, and may be compared to confirm the validity of the equations. A + sign is added to the observed cycles where an additional low amplitude embryonic cycle can be seen in the contact inlet. The change of the entrainment velocity (with the same value of ξ) is found to have little effect on the stress cycle rate, and is limited to the + symbols becoming full cycles.

The good correlation between calculated and observed cycle numbers confirms the applicability of the argument leading to Equations (4.27). In a real rough surface situation the appropriate asperity spacing wavelength depends on the separation between surface asperity features that generate a significant micro-EHL pressure response.

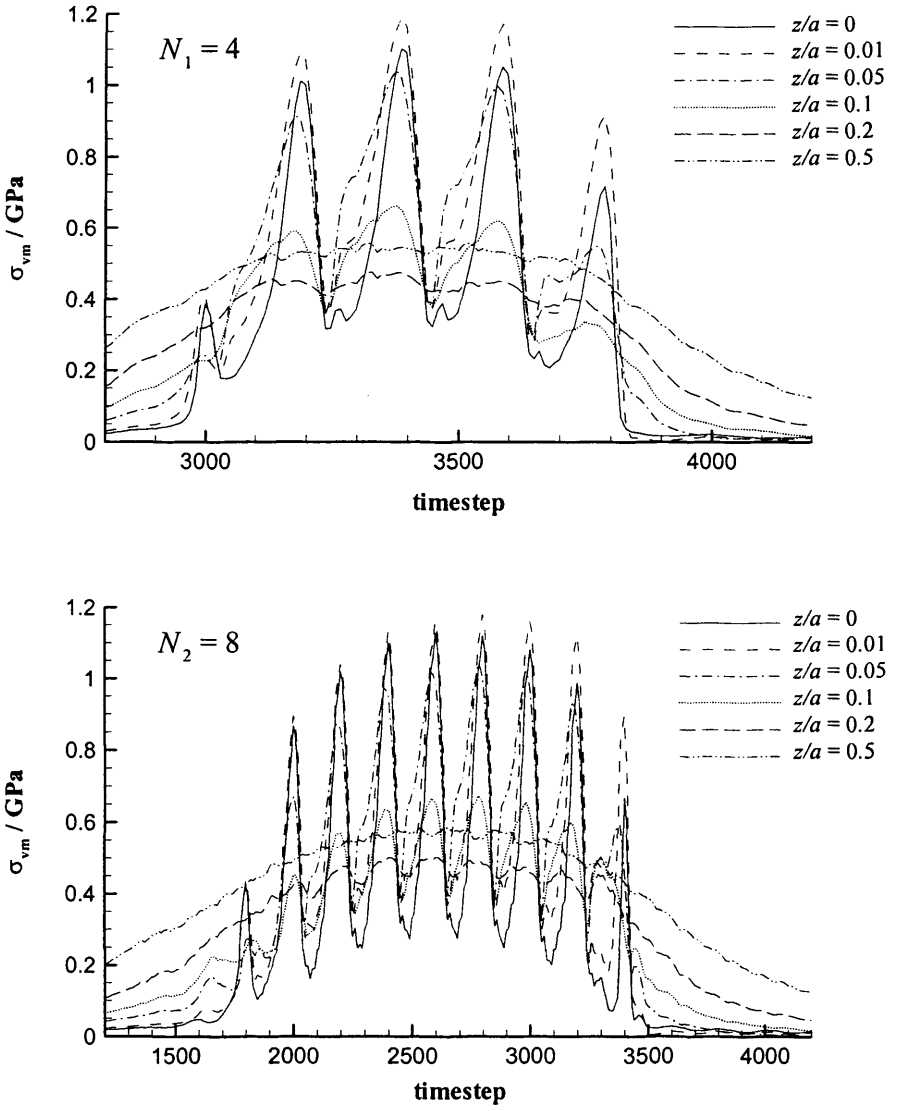


Figure 4.12: Variation of of σ_{vm} / GPa with time during transit of the contact at depths of z/a = 0, 0.01, 0.05, 0.1, 0.2 and 0.5 for two asperities from sinusoidal surfaces with $\lambda_1 = \lambda_2 = a/4$ and $\xi = 0.67$. The upper figure is for the faster surface, and the lower figure is for the slower surface.

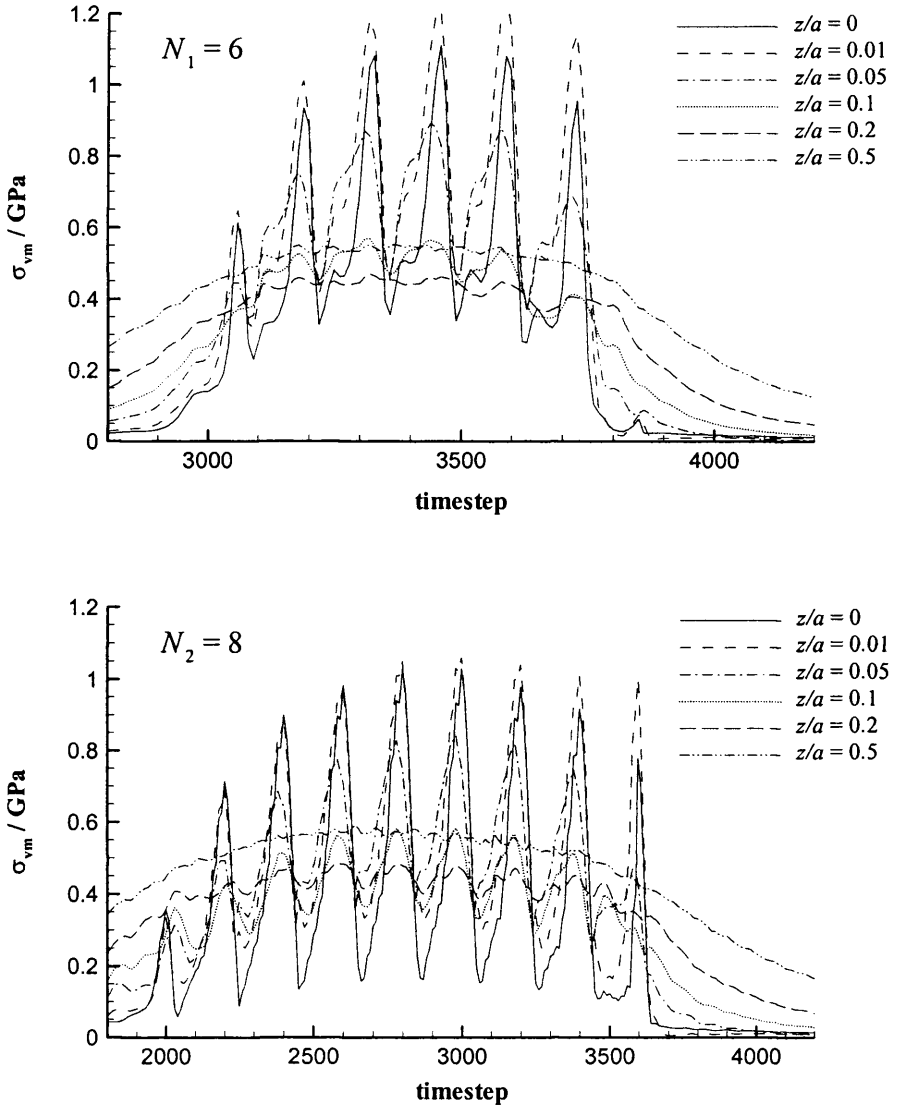


Figure 4.13: Variation of of σ_{vm} / GPa with time during transit of the contact at depths of z/a = 0, 0.01, 0.05, 0.1, 0.2 and 0.5 for two asperities from sinusoidal surfaces with $\lambda_1 = a/4$, $\lambda_2 = a/6$ and $\xi = 0.67$. The upper figure is for the faster surface, and the lower figure is for the slower surface.

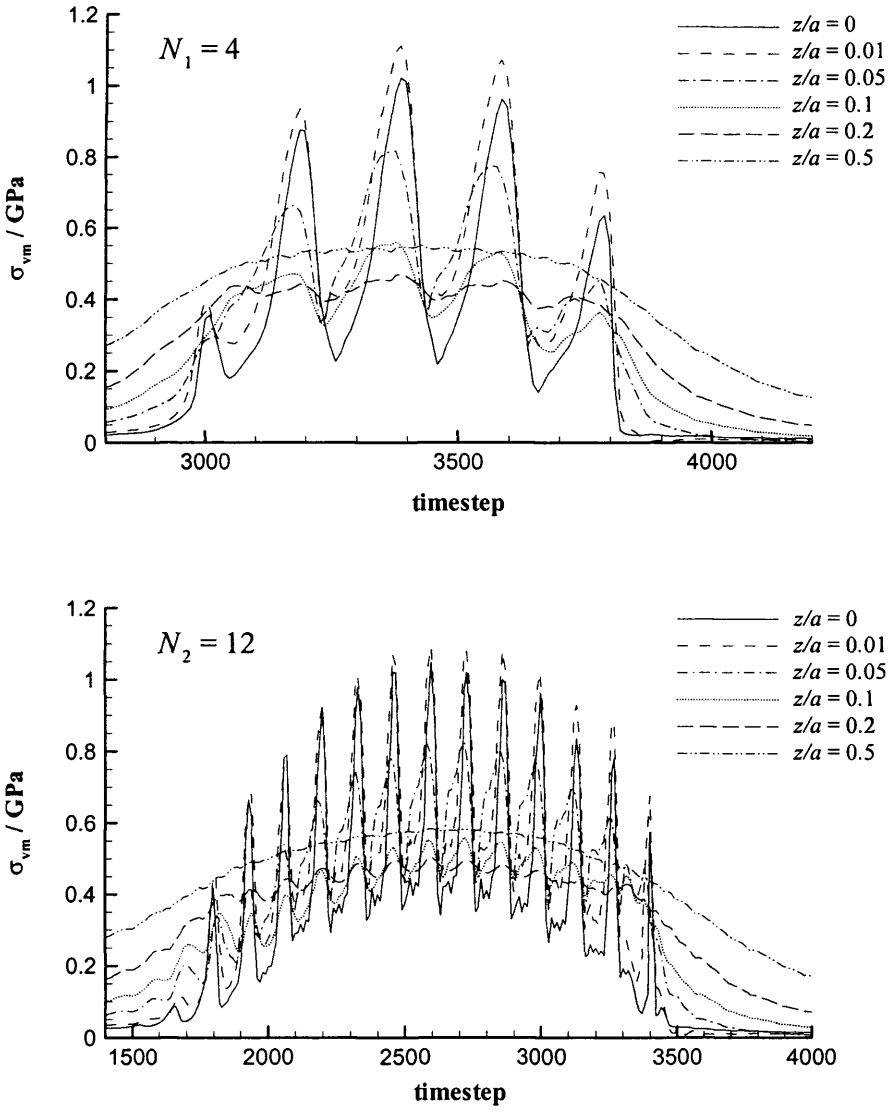


Figure 4.14: Variation of of σ_{vm} / GPa with time during transit of the contact at depths of z/a = 0, 0.01, 0.05, 0.1, 0.2 and 0.5 for two asperities from sinusoidal surfaces with $\lambda_1 = a/6$, $\lambda_2 = a/4$ and $\xi = 0.67$. The upper figure is for the faster surface, and the lower figure is for the slower surface.

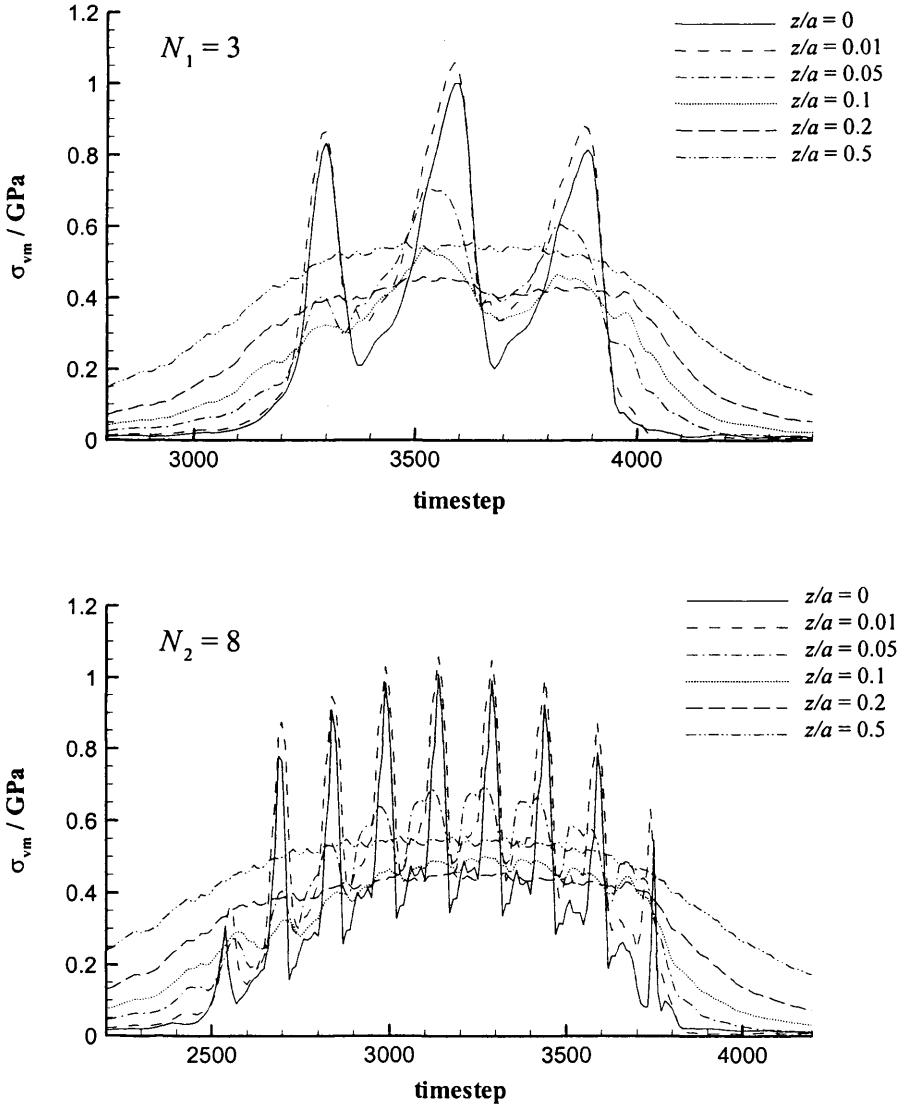


Figure 4.15: Variation of of σ_{vm} / GPa with time during transit of the contact at depths of z/a = 0, 0.01, 0.05, 0.1, 0.2 and 0.5 for two asperities from sinusoidal surfaces with $\lambda_1 = a/8$, $\lambda_2 = a/4$ and $\xi = 0.4$. The upper figure is for the faster surface, and the lower figure is for the slower surface.

Table 4.5: Calculated and observed numbers of stress cycles.

ξ	λ_1	λ_2	Calculated		Observed	
			N_1	N_2	N_1	N_2
0.67 (u_1=50, u_2=25)	$a/4$	$a/4$	4	8	4 +	8 +
	$a/6$	$a/4$	4	12	4 +	12 +
	$a/4$	$a/6$	6	8	6	8 +
	$a/8$	$a/4$	4	16	4 +	16 +
	$a/4$	$a/8$	8	8	8	8
0.67 (u_1=25, u_2=12.5)	$a/4$	$a/4$	4	8	5	9
	$a/6$	$a/4$	4	12	4 +	13
	$a/4$	$a/6$	6	8	6 +	9
	$a/8$	$a/4$	4	16	5	17
	$a/4$	$a/8$	8	8	9	9
0.4 (u_1=45, u_2=30)	$a/4$	$a/4$	2.67	4	3	4
	$a/6$	$a/4$	2.67	6	3	6
	$a/4$	$a/6$	4	4	4 +	4 +
	$a/8$	$a/4$	2.67	8	3	8 +
	$a/4$	$a/8$	5.33	4	5	4

4.5 Summary

In this chapter, an elastic stress analysis is performed for EHL line contact with rough surfaces. Accuracy and efficiency of the computer program using Discrete Convolution and FFT (DC-FFT) have been tested by comparison with other numerical methods. Results for the transient variation of von Mises equivalent stress at surface asperities are presented.

Transient variation in the stress pattern experienced by contacting rough surfaces in rolling/sliding contact as they pass through the load bearing area shows significant cycling. The effect is associated with asperity "collisions" and the cycle rate experienced can be

deduced from the kinematics, load and asperity spacing. The stress cycling can have large amplitudes and when this occurs the effect is greatest close to the surface. Such stress cycling may be expected to be closely related to contact fatigue and subsequent failure by micropitting, pitting, or spalling.

Chapter 5

Theories of Fatigue – A Review

5.1 Introduction

Rolling contact fatigue is a common mechanism for failure of the surfaces in an EHL system and is seen as the limiting factor controlling the operational life of rolling element bearings. Conventional pitting is usually associated with failure beneath the surface of a rolling contact and in the past has been linked with the occurrence of the maximum shear stress that occurs in the classical Hertzian solution for smooth surfaces. Another example of fatigue in rolling/sliding contacts is micropitting which is seen in gears and may well be viewed as rolling contact fatigue on the scale of surface asperities. Although the severity of the EHL contact can be assessed qualitatively by pressure and film thickness within the contact, it is desirable to evaluate the fatigue life quantitatively using a fatigue model which takes account of the true surface loading, particularly when roughness is present.

This chapter opens with an introduction to some of the fundamentals of fatigue. Advanced topics such as mean stress effects, multiaxial fatigue and variable amplitude fatigue are then discussed. Finally, theories of rolling contact fatigue are presented in the form of a brief literature review of the models developed. The fatigue analysis carried out for EHL line contact with rough surfaces will be reported in the following two chapters.

5.2 Fundamentals of Fatigue

5.2.1 Fatigue of Materials

Fatigue is a progressive failure that occurs in a material subjected to repeated, cyclic or fluctuating loading. Fatigue failure may occur at a nominal stress level well below the ultimate strength of the material. There are three stages in the process of fatigue: crack initiation, crack propagation, and fracture (Boardman 1990). The process may be described as the creation of microscopic cracks due to repeated local plasticity, growth of microscopic cracks and propagation of dominant cracks until a critical size is reached, whereupon failure occurs.

Fatigue behaviour is dependent upon a wide range of factors under service conditions. Studies have shown that mechanical properties, composition, microstructure and macrostructure significantly influence the fatigue life of a material subjected to cyclic stressing. In addition to material properties, fatigue life is also affected by other factors such as the type of the applied loading (bending or torsion), loading pattern (constant or variable amplitude loading), stress concentration, manufacturing process, surface roughness, operating temperature and environment, material imperfections, etc. (Boardman 1990). With so many factors involved, predicting the fatigue life of a given component is very complicated. A thorough understanding of these factors is essential to prevention of fatigue failure.

5.2.2 Cyclic Material Behaviour

In a fatigue test, the specimen is loaded so that the stress is cycled between maximum and minimum values, as shown schematically in Figure 5.1. The stress range, the stress amplitude and the mean stress, respectively, are defined as

$$\Delta\sigma = \sigma_{max} - \sigma_{min} \tag{5.1a}$$

$$\sigma_a = \frac{\Delta\sigma}{2} = \frac{\sigma_{\max} - \sigma_{\min}}{2} \tag{5.1b}$$

$$\sigma_m = \frac{\sigma_{\max} + \sigma_{\min}}{2} \tag{5.1c}$$

Ratios of two specified stress values in one cycle are sometimes used. Two commonly used ratios are the stress ratio R and the amplitude ratio A:

$$R = \frac{\sigma_{\min}}{\sigma_{\max}} \tag{5.2a}$$

$$A = \frac{\sigma_a}{\sigma_m} \tag{5.2b}$$

With this definition, there are five conditions that R can take ranging from -1 to $+1$ (Boardman 1990):

- Fully reversed loading, $R = -1$
- Partially reversed loading, $-1 < R < 0$
- Zero-to-tension loading, $R = 0$
- 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. As shown in Figure 5.2 (Hertzberg 1976), stress and strain controlled tests are used to show both the cycle hardening and cycle softening responses. The purpose of these tests is to produce a stabilized stress-strain hysteresis loop which is shown in Figure 5.3 (Dowling 1998). The size of the hysteresis loop is defined by the stress range $\Delta\sigma$ and the strain range $\Delta\varepsilon$. The cyclic strain range $\Delta\varepsilon$ comprises two components, the elastic strain range $\Delta\varepsilon_e$ and the plastic strain range $\Delta\varepsilon_p$:

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{\Delta\sigma}{E} + \Delta\varepsilon_p \tag{5.3}$$

where E is the elastic modulus.

Hysteresis loops from tests at several different strain amplitudes can be plotted in a common stress-strain diagram, as shown in Figure 5.4. A line connecting the tips of 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. If the resulting cyclic stress-strain curve is above the monotonic stress-strain curve, the material is the one that cyclically hardens, and vice versa. The Ranberg-Osgood equation is commonly used to describe cyclic stress-strain curves upon summation of the elastic strain amplitude ε_{ea} and plastic strain amplitude ε_{pa} (Dowling 1998):

$$\varepsilon_a = \varepsilon_{ea} + \varepsilon_{pa} = \frac{\sigma_a}{E} + \left(\frac{\sigma_a}{H'}\right)^{\frac{1}{n'}} \tag{5.4}$$

where H' is the cyclic strength coefficient and n' is the cyclic strain hardening exponent.

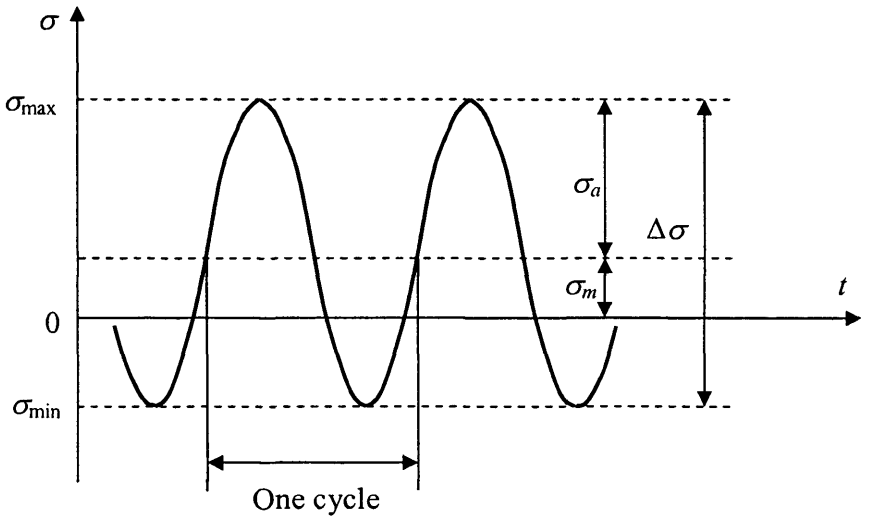


Figure 5.1: Nomenclature to describe cyclic stress. The variation of stress σ with time t is shown.

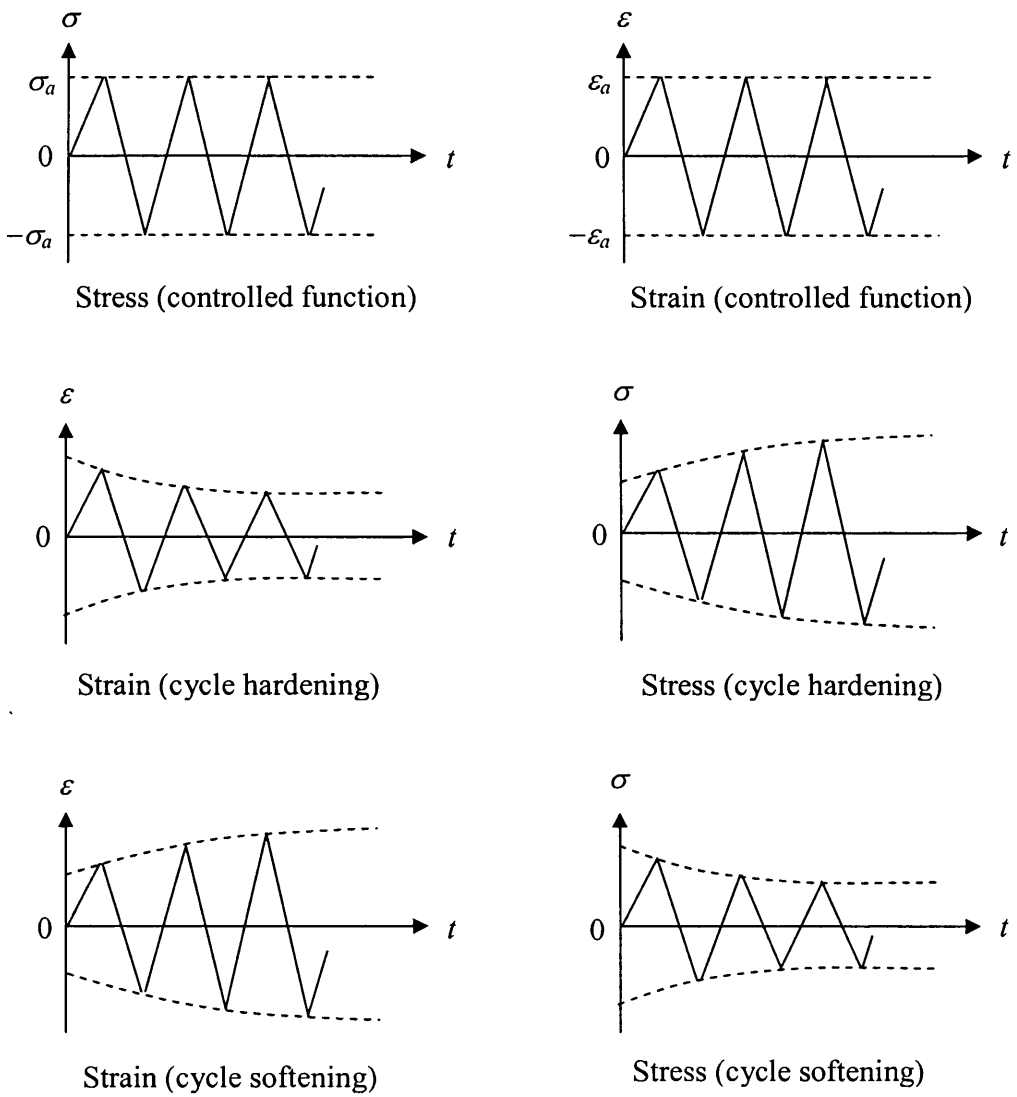


Figure 5.2: Cyclic material behaviour under (a) stress control and (b) strain control.

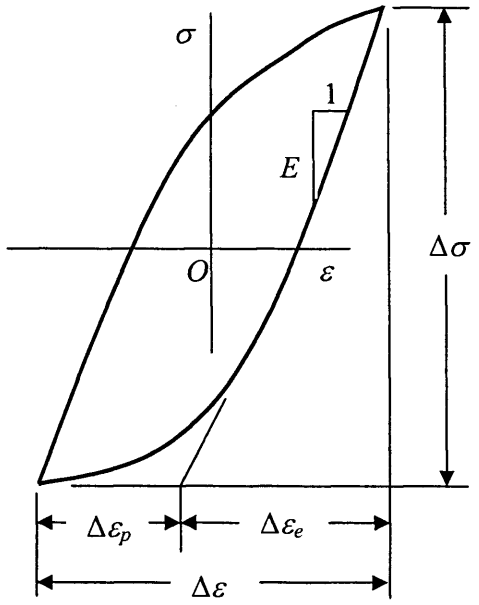


Figure 5.3: Stable stress-strain hysteresis loop.

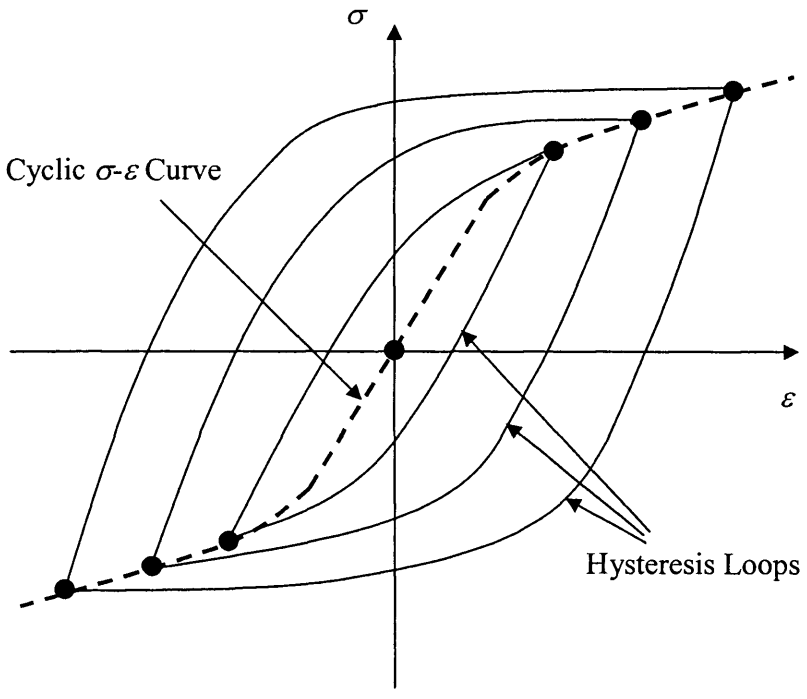


Figure 5.4: Cyclic stress-strain curve.

5.2.3 Fatigue Life

Fatigue life is the total number of stress cycles required for a fatigue crack to initiate and grow large enough to produce catastrophic failure, that is, separation into two pieces (Boardman 1990). In this thesis, the fatigue life is expressed in terms of total life.

Fatigue data are usually obtained from a fully reversed, constant amplitude fatigue test and presented in the form of *S-N* curves (also referred to as Wöhler's curves), where applied stress (*S*) is plotted against the total cycles to failure (*N*). A complete *S-N* curve may be divided into two portions, the high-cycle (low stress) range and the low-cycle (high stress) range, although they are not completely separable. Consequently, fatigue problems can be conveniently divided into two categories, namely high-cycle fatigue (HCF) and low-cycle fatigue (LCF).

In HCF, a large number of cycles are required to cause failure and macroscopic plasticity or yielding does not dominate the behaviour. Since under these conditions the usual elastic relationship between stress and strain applies, HCF curves are usually represented by stress amplitude (σ_a) as the ordinate, and number of cycles to failure (N_f) as the abscissa, as indicated in Figure 5.5. This stress-life relationship can be described by the Basquin relation as (Dowling 1998, Suresh 1998)

$$\sigma_a = \sigma'_f (2N_f)^b \tag{5.5}$$

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

In Figure 5.5, the symbol σ_e is called the fatigue limit or endurance limit. Below this stress level, the specimen may be cycled indefinitely without failure. For materials which do not exhibit such behaviour, a fatigue limit is often defined as a specific long life, say 10^7 or 10^8 cycles (Dowling 1998). Another term is fatigue strength, which is used to specify a stress amplitude at a particular life of interest. For example, the fatigue strength at 10^5 cycles is the stress amplitude corresponding to $N_f = 10^5$.

In LCF, the lifetime is limited to a relatively small number of cycles and the material is subjected to stress levels that produce considerable macroscopic plastic strains. Under these circumstances, the linear relationship between stress and strain is invalid, and LCF curves are usually represented by the total strain amplitude (ε_a) as the ordinate and number of cycles to failure (N_f) as the abscissa as indicated in Figure 5.6, where both the ordinate and the abscissa are on logarithmic scales. The elastic strain amplitude is related to the fatigue life by the Basquin relationship:

$$\varepsilon_{ea} = \frac{\sigma_a}{E} = \frac{\sigma'_f}{E}(2N_f)^b \tag{5.6}$$

The plastic strain amplitude is related to the fatigue life by the Coffin-Manson relationship (Dowling 1998, Suresh 1998):

$$\varepsilon_{pa} = \varepsilon'_f(2N_f)^c \tag{5.7}$$

where ε'_f is the fatigue ductility coefficient, c is the fatigue ductility exponent. (Note that powers b and c are intrinsically negative.) Combining Equations (5.6) and (5.7) gives an equation relating the total strain amplitude and life (Dowling 1998, Suresh 1998):

$$\varepsilon_a = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \tag{5.8}$$

From Figure 5.6, it can be seen that the total strain is asymptotic to the plastic strain component in the LCF region but asymptotic to the elastic strain component in the HCF region. The crossing point of the elastic and plastic strain lines is identified as the transition fatigue life, N_t, which can be obtained by using $\varepsilon_{ea} = \varepsilon_{pa}$:

$$N_t = \frac{1}{2}\left(\frac{\sigma'_f}{\varepsilon'_f E}\right)^{\frac{1}{c-b}} \tag{5.9}$$

The value of N_t may be the most logical point for separating low-cycle and high-cycle fatigue (Dowling 1998).

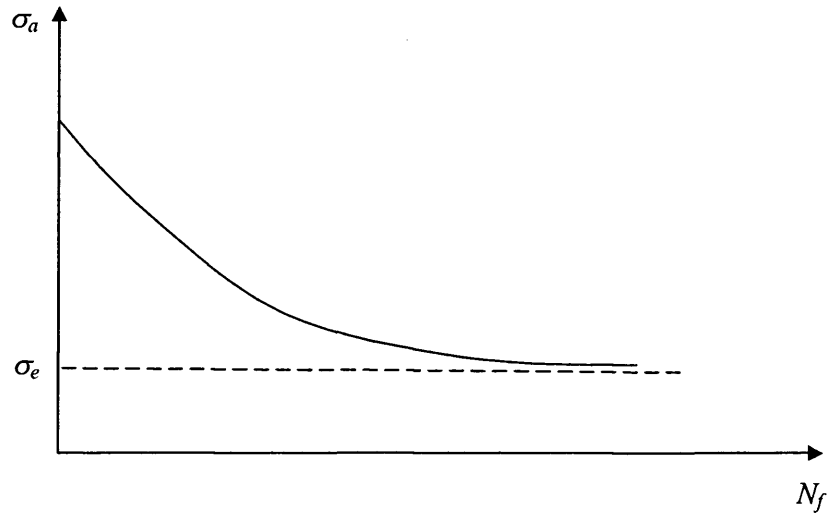


Figure 5.5: Stress-life relationship.

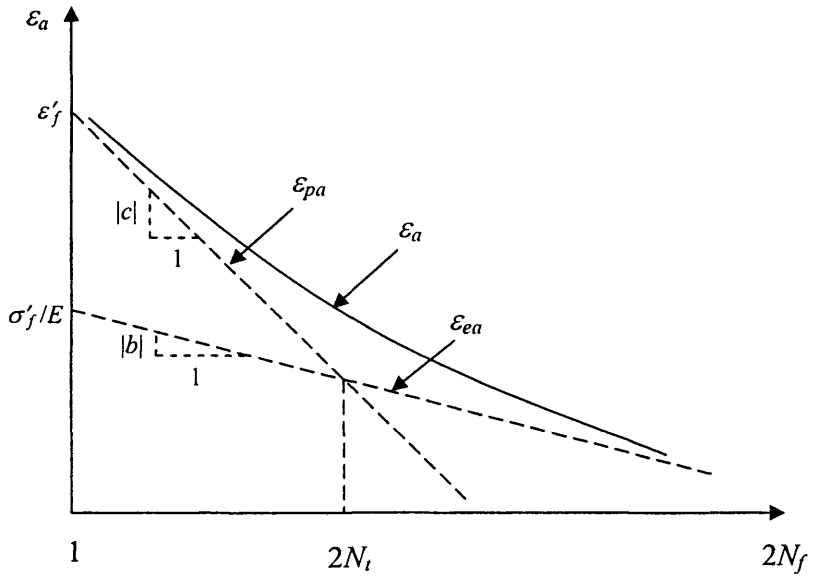


Figure 5.6: Strain-life relationship (Note that both ε_a and N_f are in log scales, and powers b and c are negative and their lines are accordingly illustrated).

5.2.4 Estimating Fatigue Parameters

The fatigue parameters (σ'_f, b, ε'_f, c) can be determined experimentally. The values obtained for several materials are available in references (Boardman 1990, Hertzberg 1976, Zahavi & Torbilo 1996). In the absence of experimentally determined values, these parameters can be related to other material properties and estimated.

From Equation (5.4), the plastic strain term of the cyclic stress-strain curve gives

$$\sigma_a = H'\varepsilon_{pa}^{n'} \tag{5.10}$$

Eliminating N_f between Equation (5.5) and Equation (5.7) and combining with Equation (5.10) yields relationships for the constants of the strain-life curve and those of the cyclic stress-strain curve (Dowling 1998):

$$n' = \frac{b}{c} \tag{5.11a}$$

$$H' = \frac{\sigma'_f}{\left(\varepsilon'_f\right)^{n'}} \tag{5.11b}$$

However, the above relationships are satisfied only approximately. In some situations where the data do not fit the assumed mathematical form very well, there may be large inconsistencies among these parameters.

The fatigue strength coefficient for steels with hardness below 500HB can be approximated in MPa using the ultimate strength of the material, σ_u, i.e. (Boardman 1990)

$$\sigma'_f = \sigma_u + 345 \tag{5.12}$$

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

$$\varepsilon'_f \approx \varepsilon_f = \ln\left(\frac{100}{100 - \%RA}\right) \tag{5.13}$$

where $\%RA$ is the percentage of the reduction in area at fracture.

For steels the fatigue strength exponent b is usually about -0.085 and the fatigue ductility exponent c is usually about -0.6 (Boardman 1990). For most metals, b is in the range of -0.05 to -0.12 and c is in the range of -0.5 to -0.7 (Suresh 1998). Based on an energy argument, Morrow determined b and c as (Hertzberg 1976)

$$b = \frac{-n'}{1+5n'} \tag{5.14a}$$

$$c = \frac{-1}{1+5n'} \tag{5.14b}$$

5.2.5 Approaches to Fatigue Design

There are two major approaches used for fatigue life estimation, namely total-life approach and defect-tolerant approach (Suresh 1998).

The total-life approach is the classical approach to fatigue design. It may be the stress-based method for the high-cycle fatigue, or the strain-based method for the low-cycle fatigue. The strain-based method can also be used for high-cycle fatigue design where elastic strains dominate. Since the crack initiation life constitutes a major component of the total life, the stress-based and strain-based methods represent, in many cases, design against fatigue crack initiation. A drawback of the total-life approach is that the choice of a definition for crack initiation is ambiguous.

The defect-tolerant approach uses the methods of fracture mechanics to calculate the number of cycles required to propagate a crack to a critical size. In this approach, materials are considered as inherently flawed and the size of the pre-existing flaw has to be detected. However, inspection of cracks, especially small ones, is an expensive process. If no cracks are detected, an initial crack size is estimated. However, the results are sensitive to the choice of the initial size.

5.3 Mean Stress Effects

The descriptions of fatigue life given above are obtained from fully reversed fatigue loads where the mean stress $\sigma_m = 0$. Fatigue tests show that the *S-N* curve is generally a strong function of the applied mean stress and increasing the mean stress in the tensile direction is accompanied by a decreasing fatigue life. To take into account the mean stress effects, a normalized amplitude-mean diagram (Dowling 1998) can be plotted using different combinations of the stress amplitude and mean stress providing a constant fatigue life. Let σ_{ar} be the fully reversed stress amplitude with zero mean stress, σ_o and σ_u the yield strength and ultimate strength of the material. Three well known models, the Soderberg relation, the modified Goodman relation and the Gerber parabola relation, are described, respectively, by the following expressions (Dowling 1998, Suresh 1998):

$$\frac{\sigma_a}{\sigma_{ar}} + \frac{\sigma_m}{\sigma_o} = 1 \tag{5.15a}$$

$$\frac{\sigma_a}{\sigma_{ar}} + \frac{\sigma_m}{\sigma_u} = 1 \tag{5.15b}$$

$$\frac{\sigma_a}{\sigma_{ar}} + \left(\frac{\sigma_m}{\sigma_u}\right)^2 = 1 \tag{5.15c}$$

and are illustrated in Figure 5.7.

The discussions of the advantages and disadvantages of these models can be found in the references (Zahavi & Torbilo 1996, Dowling 1998, Suresh 1998). As a general rule-of-thumb, the Soderberg line is overly conservative whereas the Gerber curve is hazardous. The Goodman line intercepted by the yield line (i.e. line ACB in Figure 5.7) is considered to be the most suitable for design purposes (Zahavi & Torbilo 1996).

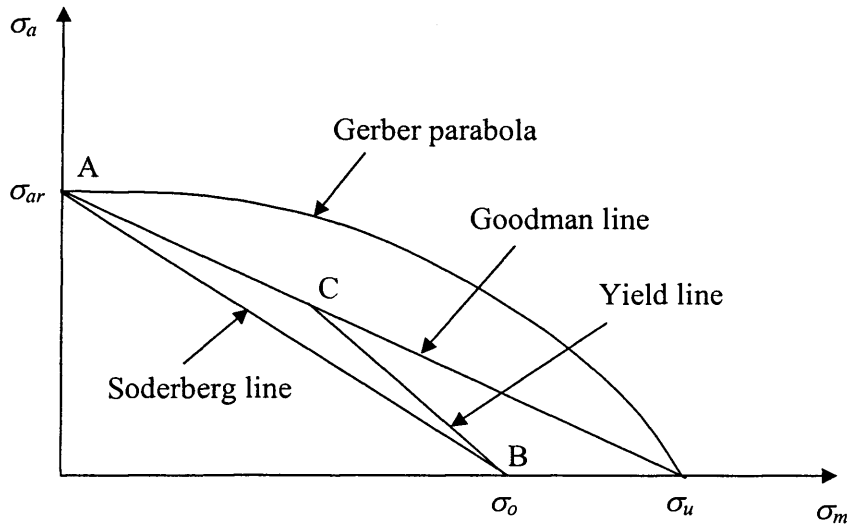


Figure 5.7: Normalized amplitude-mean diagram.

Morrow proposed a modification of the Goodman line by replacing σ_u with σ'_f in Equation (5.15b) (Dowling 1998):

$$\frac{\sigma_a}{\sigma_{ar}} + \frac{\sigma_m}{\sigma'_f} = 1 \tag{5.16}$$

Since fully reversed tests are generally employed to obtain the constants, the Basquin equation needs to be written as

$$\sigma_{ar} = \sigma'_f (2N_f)^b \tag{5.17}$$

Substituting for σ_{ar} in Equation (5.16) yields the Morrow equation which accounts for mean stress effects:

$$\sigma_a = (\sigma'_f - \sigma_m)(2N_f)^b \tag{5.18}$$

The relationship between the plastic strain amplitude and the fatigue life can be obtained by using Equation (5.10), (5.11) and (5.18):

$$\begin{aligned}\varepsilon_{pa} &= \left(\frac{\sigma_a}{H'}\right)^{\frac{1}{n'}} = \varepsilon_f'\left(\frac{\sigma_a}{\sigma_f'}\right)^{\frac{c}{b}} \\ &= \varepsilon_f'\left[\frac{\sigma_f' - \sigma_m}{\sigma_f'}(2N_f)^b\right]^{\frac{c}{b}} \\ &= \varepsilon_f'\left(\frac{\sigma_f' - \sigma_m}{\sigma_f'}\right)^{\frac{c}{b}}(2N_f)^c\end{aligned} \tag{5.19}$$

Thus, from Equation (5.4) the strain-life relation for nonzero mean stresses can be written as (Dowling 1998)

$$\varepsilon_a = \frac{\sigma_f' - \sigma_m}{E}(2N_f)^b + \varepsilon_f'\left(1 - \frac{\sigma_m}{\sigma_f'}\right)^{\frac{c}{b}}(2N_f)^c \tag{5.20}$$

A modified version of Equation (5.20) is often used (Dowling 1998):

$$\varepsilon_a = \frac{\sigma_f' - \sigma_m}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \tag{5.21}$$

In Equation (5.21), the effect of the mean stress is restricted to the elastic strain domain.

An additional relationship is the Smith, Watson, and Topper (SWT) equation (Dowling 1998):

$$\sigma_{ar} = \sqrt{\sigma_{\max}\sigma_a} \quad (\sigma_{\max} > 0) \tag{5.22}$$

It predicts that no damage will occur when $\sigma_{\max} < 0$, where by definition $\sigma_{\max} = \sigma_a + \sigma_m$. For the strain-based method, it assumes the life for any situation of mean stress depends on the product $\sigma_{\max}\varepsilon_a$. Let σ_{ar} and ε_{ar} be the fully reversed stress and strain amplitudes, and note that for $\sigma_m = 0$, $\sigma_{\max} = \sigma_{ar}$. Thus,

$$\begin{aligned}\sigma_{\max}\varepsilon_a &= \sigma_{ar}\varepsilon_{ar} \\ &= \sigma'_f(2N_f)^b\left[\frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c\right] \\ &= \frac{(\sigma'_f)^2}{E}(2N_f)^{2b} + \sigma'_f\varepsilon'_f(2N_f)^{b+c}\end{aligned} \tag{5.23}$$

5.4 Multiaxial Fatigue

The discussion up to this point has dealt with the fatigue for loading on the basis of uniaxial cyclic stresses, such as pure axial tension and compression, pure bending or torsion. However, in many practical applications components are subjected to multiaxial cyclic stresses due to complicated geometries and loadings. In this section, the multiaxial fatigue theories are discussed for both the stress-based approach and the strain-based approach.

5.4.1 Stress-based Approach

5.4.1.1 Effective Stress Approach

Similar to the von Mises yield criterion, the amplitude of the effective stress can be written as (Dowling 1998)

$$\bar{\sigma}_a = \frac{1}{\sqrt{2}}\sqrt{(\sigma_{1a}-\sigma_{2a})^2 + (\sigma_{2a}-\sigma_{3a})^2 + (\sigma_{3a}-\sigma_{1a})^2} \tag{5.24}$$

where σ_{ia} (i = 1, 2, 3) are the amplitudes of principal stresses. In terms of the amplitudes of the stress components, the effective stress amplitude can also be calculated by (Dowling 1998)

$$\bar{\sigma}_a == \frac{1}{\sqrt{2}}\sqrt{(\sigma_{xa}-\sigma_{ya})^2 + (\sigma_{ya}-\sigma_{za})^2 + (\sigma_{za}-\sigma_{xa})^2 + 6(\tau^2_{xya}+\tau^2_{yza}+\tau^2_{zxa})} \tag{5.25}$$

The effective mean stress can be calculated from three principal mean stresses:

$$\bar{\sigma}_m = \sigma_{1m} + \sigma_{2m} + \sigma_{3m} = \sigma_{xm} + \sigma_{ym} + \sigma_{zm} \tag{5.26}$$

An alternative form of the effective mean stress is based on the octahedral shear stress

$$\bar{\sigma}_m = \frac{1}{\sqrt{2}}\sqrt{(\sigma_{1m} - \sigma_{2m})^2 + (\sigma_{2m} - \sigma_{3m})^2 + (\sigma_{3m} - \sigma_{1m})^2} \tag{5.27}$$

With the use of effective stresses, the multiaxial stress situation is transformed into an equivalent fully reversed uniaxial stress and Equation (5.18) can be written as (Dowling 1998)

$$\bar{\sigma}_a = (\sigma'_f - \bar{\sigma}_m)(2N_f)^b \tag{5.28}$$

This effective stress approach is limited to proportional (in-phase) loading (Fatemi and Socie 1988). Other drawbacks are that the differing effects of axial tension and compression mean stresses may not be accurately captured and the orientation of fatigue cracks is not quantitatively determined.

5.4.1.2 Multiaxial Fatigue Criteria

Like the effective stress approach, multiaxial fatigue criteria can be applied to reduce a given multiaxial stress state to an equivalent effective uniaxial stress condition, which can be compared to uniaxial fatigue data. Unlike the effective stress approach which uses the stress-life relationship, multiaxial fatigue criteria compare the stress state to the fatigue limit or the stress level for a fixed fatigue life. Of the numerous criteria that have been proposed (You & Lee 1996, Papadopoulos et al. 1997), two groups of stress-based criteria (i.e. criteria for high-cycle fatigue) are briefly discussed here.

5.4.1.2.1 Stress Invariant Based Criteria

The ingredients of these fatigue criteria are the hydrostatic stress and the second invariant of the stress deviator which have the same magnitude for all directions. The result is that these criteria are easy to use. However, the orientation of the potential fatigue damage is by no means specified.

In terms of stress components, the hydrostatic stress, σ_h, is defined as

$$\sigma_h = \frac{1}{3}(\sigma_x + \sigma_y + \sigma_z) \tag{5.29}$$

The square root of the second invariant of the stress deviator is defined as

$$\sqrt{J_2} = \frac{1}{\sqrt{6}}\sqrt{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2 + 6(\tau_{xy}^{\ 2} + \tau_{yz}^{\ 2} + \tau_{zx}^{\ 2})} \tag{5.30}$$

Sines Criterion

This criterion is based on the amplitude of the second invariant of the stress deviator and the mean value of the hydrostatic stress, written as:

$$\sqrt{J_{2,a}} + \bar{\kappa}\sigma_{h,m} \le \bar{\lambda} \tag{5.31}$$

Fatigue failure occurs if this inequality is not fulfilled. The parameters $\bar{\kappa}$ and $\bar{\lambda}$ can be obtained from a torsion test (i.e. $\sqrt{J_{2,a}} = \tau_{af}$, $\sigma_{h,m} = 0$ and τ_{af} is the fatigue limit for pure torsion) and a repeated bending test (i.e. $\sqrt{J_{2,a}} = \sigma_{ab}/\sqrt{3}$, $\sigma_{h,m} = \sigma_{ab}/3$, and σ_{ab} is the repeated bending fatigue limit):

$$\bar{\kappa} = \frac{3\tau_{af}}{\sigma_{ab}} - \sqrt{3} \tag{5.32a}$$

$$\bar{\lambda} = \tau_{af} \tag{5.32b}$$

Since the limit σ_{ab} is seldom available, σ_{ab} is usually obtained from the fatigue limit on the fully reversed bending σ_{af} and the ultimate tensile strength σ_u using the Goodman relation (for the repeated bending, $\sigma_m = \sigma_{ab}$) (Papadopoulos et al. 1997):

$$\sigma_{ab} = \frac{\sigma_{af}}{1 + \sigma_{af}/\sigma_u} \tag{5.33}$$

Crossland Criterion

The Crossland criterion differs from the Sines criterion only concerning the influence of the hydrostatic stress, which appears in the Crossland criterion formula in terms of its maximum value:

$$\sqrt{J_{2,a}} + \bar{\kappa}\sigma_{h,\max} \le \bar{\lambda} \tag{5.34}$$

The parameters $\bar{\kappa}$ and $\bar{\lambda}$ can easily be obtained from a torsion test (i.e. $\sqrt{J_{2,a}} = \tau_{af}$, $\sigma_{h,\max} = 0$) and a fully reversed bending test (i.e. $\sqrt{J_{2,a}} = \sigma_{af}/\sqrt{3}$, $\sigma_{h,\max} = \sigma_{af}/3$). Application of the criterion in these two cases provides:

$$\bar{\kappa} = \frac{3\tau_{af}}{\sigma_{af}} - \sqrt{3} \tag{5.35a}$$

$$\bar{\lambda} = \tau_{af} \tag{5.35b}$$

5.4.1.2.2 Critical Plane Based Criteria

In these criteria, the fatigue failure assessment is performed in a plane called the "critical plane" where the amplitude or the value of some stress components or a combination of them attains its maximum. The orientation of the initiated crack coincides with orientation of the critical plane.

The orientation of the critical plane is defined by the spherical coordinates (φ, θ) of its unit normal vector **n**. θ is the angle between **n** and the z axis and φ is the angle between the projection of **n** onto the x-y plane and the x axis (Figure 5.8). The reference coordinate axes, x, y and z, are marked on the figure with the axis of the fatigue testing specimen as the x axis.

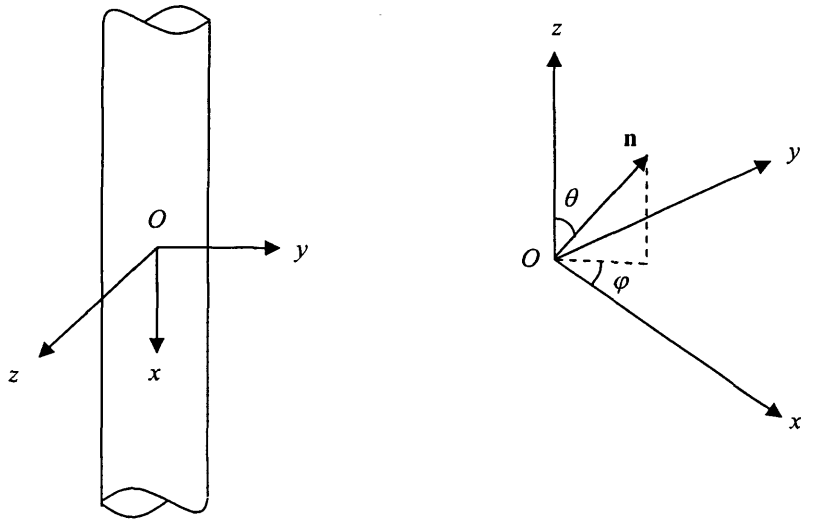


Figure 5.8: Coordinates for definition of critical plane.

Findley Criterion

The Findley criterion determines the critical plane (φ_c, θ_c) by maximizing a linear combination of τ_a (shear stress amplitude) and $\sigma_{n,\max}$ (maximum value of the normal stress):

$$(\varphi_c, \theta_c): \max_{(\varphi,\theta)} \left[\tau_a(\varphi,\theta) + \bar{\kappa}\sigma_{n,\max}(\varphi,\theta)\right] \tag{5.36}$$

Fatigue failure occurs if the following expression is not fulfilled:

$$\tau_a(\varphi_c,\theta_c) + \bar{\kappa}\sigma_{n,\max}(\varphi_c,\theta_c) \le \bar{\lambda} \tag{5.37}$$

where $\bar{\kappa}$ and $\bar{\lambda}$ are determined from the fatigue limits on fully reversed bending and torsion tests (Carpinteri & Spagnoli 2001):

$$\bar{\kappa} = \frac{2 - \sigma_{af}/\tau_{af}}{2\sqrt{\sigma_{af}/\tau_{af} - 1}} \tag{5.38a}$$

$$\bar{\lambda} = \frac{\sigma_{af}}{2\sqrt{\sigma_{af}/\tau_{af} - 1}} \tag{5.38b}$$

Matake Criterion

According to the Matake criterion, the critical plane is the plane on which the amplitude of shear stress reaches its maximum, that is:

$$(\varphi_c, \theta_c): \quad \max_{(\varphi,\theta)}[\tau_a(\varphi,\theta)] \tag{5.39}$$

Then the criterion is written in the same form as the Findley criterion:

$$\tau_a(\varphi_c,\theta_c) + \bar{\kappa}\sigma_{n,\max}(\varphi_c,\theta_c) \le \bar{\lambda} \tag{5.40}$$

where $\bar{\kappa}$ and $\bar{\lambda}$ are determined by applying the equation to two limit cases: fully reversed torsion (i.e. $\tau_a(\varphi_c,\theta_c) = \tau_{af}$ and $\sigma_{n,\max}(\varphi_c,\theta_c) = 0$) and fully reversed bending (i.e. $\tau_a(\varphi_c,\theta_c) = \sigma_{af}/2$ and $\sigma_{n,\max}(\varphi_c,\theta_c) = \sigma_{af}/2$). This yields:

$$\bar{\kappa} = \frac{2\tau_{af}}{\sigma_{af}} - 1 \tag{5.41a}$$

$$\bar{\lambda} = \tau_{af} \tag{5.41b}$$

McDiarmid Criterion

The McDiarmid criterion uses the same definitions of critical plane and criterion inequality as the Matake criterion, i.e. Equations (5.39) and (5.40). But the material parameters $\bar{\kappa}$ and $\bar{\lambda}$ are determined in a different way by employing the concept of case A and case B cracks. Case A cracks propagate along the surface of the specimen, whereas case B cracks propagate inwards from the surface. $\bar{\kappa}$ and $\bar{\lambda}$ can be expressed as

$$\bar{\kappa} = \frac{\tau_{A,B}}{2\sigma_u} \tag{5.42a}$$

$$\bar{\lambda} = \tau_{A,B} \tag{5.42b}$$

where the shear stress fatigue limit, $\tau_{A,B}$, is equal to τ_A or τ_B, depending on which case the critical plane corresponds to. For combined bending and torsion, case A cracks develop with the critical plane of $\theta_c = 90°$ and therefore $\tau_{A,B} = \tau_A = \tau_{af}$.

Dang Van Criterion

The Dang Van criterion (Dang Van et al. 1989) has a different form from the previous critical plane criteria in that the current stress value is considered. If the mean shear stress τ_m is determined on a plane, then the current shear stress amplitude on the same plane is defined as (Ekberg 1997)

$$\tau_a(t) = \left|\tau(t) - \tau_m\right| \tag{5.43}$$

where $\tau(t)$ is the shear stress at time t. Then the Dang Van criterion is written as the combination of the current amplitude $\tau_a(t)$ and current hydrostatic stress $\sigma_h(t)$ on the plane:

$$\tau_a(t) + \bar{\kappa}\sigma_h(t) \le \bar{\lambda} \tag{5.44}$$

where $\bar{\kappa}$ and $\bar{\lambda}$ are determined by fully reversed torsion tests and fully reversed bending tests as (Ciavarella & Maitournam 2004)

$$\bar{\kappa} = 3\frac{\tau_{af}}{\sigma_{af}} - \frac{3}{2} \tag{5.45a}$$

$$\bar{\lambda} = \tau_{af} \tag{5.45b}$$

5.4.2 Strain-based Approach

5.4.2.1 Effective Strain Approach

The effective strain amplitude can be written as the summation of the elastic and plastic parts:

$$\bar{\varepsilon}_a = \frac{\bar{\sigma}_a}{E} + \bar{\varepsilon}_{pa} \tag{5.46}$$

where the effective stress amplitude is defined by Equation (5.24) and the effective plastic strain amplitude is defined in terms of amplitudes of principal plastic strains (Dowling 1998) as

$$\bar{\varepsilon}_{pa} = \sqrt{\frac{2}{3}\left[\left(\frac{\Delta\varepsilon_{p1}}{2}\right)^2 + \left(\frac{\Delta\varepsilon_{p2}}{2}\right)^2 + \left(\frac{\Delta\varepsilon_{p3}}{2}\right)^2\right]} \tag{5.47}$$

where ε_{pi} (i = 1, 2, 3) are the principal plastic strains. Thus, the strain-based expression based on the effective strain becomes

$$\bar{\varepsilon}_a = \frac{\sigma'_f - \bar{\sigma}_m}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \tag{5.48}$$

The application of the effective strain approach is limited to combined loadings that are in-phase or 180° out-of-phase (Dowling 1998). If the loading is nonproportional to a significant degree, a critical plane approach is needed.

5.4.2.2 Critical Plane Approach

In this approach, the stresses and strains acting on the critical plane are used to assess fatigue damage. There are two distinct modes of crack initiation and early growth, namely growth on planes of high tensile stress (mode I) and growth on planes of high shear stress (mode II) (Dowling 1998). For tensile cracking, the SWT relationship, i.e. Equation (5.23) can be used. The quantity ε_a is the amplitude of normal strain for the critical plane and σ_{max} is the maximum normal stress on the same plane as ε_a. For shear cracking, a relationship suggested by Fatemi and Socie (1988) can be used:

$$\gamma_a\left(1 + K\frac{\sigma_{max}}{\sigma'_o}\right) = \frac{\tau'_f}{G}(2N_f)^b + \gamma'_f(2N_f)^c \tag{5.49}$$

where γ_a is the amplitude of shear strain on the critical plane, and σ_{max} is the maximum tensile stress normal to the plane of γ_a. K is a material constant, K = 0.6 to 1.0, G is the shear modulus, and σ'_o is the yield strength for the cyclic stress-strain curve. The shear fatigue strength, τ'_f, and shear fatigue ductility coefficient, γ'_f, are determined from fully reversed tests in pure shear.

5.5 Variable Amplitude Fatigue

Under service conditions the fatigue load usually varies randomly, hence the variable amplitude fatigue behaviour should be addressed to achieve accurate lifetime predictions. To deal with random loading, several major problems have to be solved. One is how to identify complete stress cycles as induced by the load spectra. An effective cycle counting method is required to determine the stress cycles and corresponding stress levels. Another problem is how to sum up partial damages resulting from these different stress levels. To achieve this, a cumulative damage theory must be applied. If multiaxial fatigue must be dealt with, the problem of random loading becomes extremely complicated.

5.5.1 Cycle Counting

A cycle counting scheme is employed in fatigue analysis to reduce a complex irregular loading history to some simple form of cycle (i.e. constant amplitude cycle). Various counting methods have been developed to achieve this, such as Level Crossing Counting, Peak Counting, Simple Range Counting, Rainflow Counting, etc. (Society of Automotive Engineers 1988). Amongst these methods, the rainflow method has been widely used and may be the best approach for cycle counting (Dowling 1998).

The classical rainflow counting method is based on the cyclic material behaviour and counts cycles by identifying closed hysteresis loops in the stress-strain response. This is illustrated by the example given in Figure 5.9 (Anthes 1997). The stress-time history at the left leads to the stress-strain path in the middle that forms the closed hysteresis loops shown at the right. The cycles identified for this example are (3-4), (5-6) and (1-2).

Since the rainflow counting method was introduced by Endo and colleagues in 1968 (Dowling 1998), several variants of the method have been developed (Downing 1982, Glinka & Kam 1987, Nie 1991, Zhu et al. 1993, Anthes 1997). The source codes of computer programs written in FORTRAN can also be found in the literature (Downing 1982, Glinka & Kam 1987). Wide applications of the rainflow method led some researchers to optimize the procedures and offer standards regarding the method (Amzallag et al. 1994).

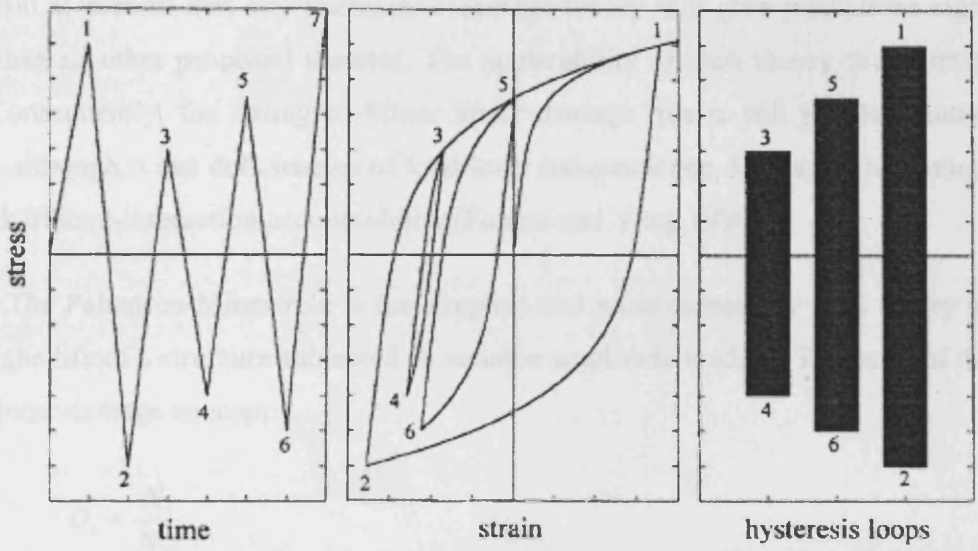


Figure 5.9: Example of the classical rainflow cycle counting method.

5.5.2 Cumulative Damage

Fatigue damage increases with applied cycles in a cumulative manner which may lead to fracture. The problem of summing partial damages resulting from different load cycles is dealt with by cumulative damage theories. Since the introduction of the damage accumulation concept by Palmgren, a variety of cumulative damage theories have been developed (Fatemi & Yang 1998, Rao et al. 2001). A thorough, comprehensive literature review of cumulative damage theories was given by Fatemi and Yang (1998) who define six categories:

- Linear damage rules
- Nonlinear damage curve and two-stage linearization approaches
- Life curve modification methods
- Approaches based on crack growth concepts
- Continuum damage mechanics models
- Energy-based theories

Though many damage theories have been developed, unfortunately, there is no indication at present that any cumulative damage theory will give predictions significantly better than all other proposed theories. The applicability of each theory varies from case to case. Consequently, the Palmgren-Miner linear damage rule is still predominantly used in design, although it has deficiencies of load-level independence, load-sequence independence and lack of load-interaction accountability (Fatemi and Yang 1998).

The Palmgren-Miner rule is the simplest and most commonly used theory to predict the fatigue life of a structure subjected to variable amplitude loading. The basis of this theory is the linear damage concept:

$$D_i = \frac{N_i}{N_{fi}} \tag{5.50}$$

That is, if a stress amplitude σ_{ai} gives a life of N_{fi} cycles and the structure is subjected to N_i cycles, it suffers a damage fraction D_i. D_i has a value less than or equal to unity. Failure is then predicted to take place when the total damage

$$D = \sum D_i = \sum \frac{N_i}{N_{fi}} \geq 1 \tag{5.51}$$

5.6 Rolling Contact Fatigue

Rolling contact fatigue is the principal mode of failure of rolling surfaces when the surfaces repeatedly touch each other (Johnson 1989). Some prominent failure modes, as

described in Chapter 1, include pitting, micropitting, spalling and various other types of surface distress.

5.6.1 Repeated Contacts and Shakedown

The stress situation in contact is more complex than in standard bending or axial fatigue. The material response to repeated contacts can be in any of the four possible ways as illustrated in Figure 5.10 (Johnson 1989):

(a) *Perfectly elastic* If the maximum stress does not exceed the yield stress of the material, the behaviour will be perfectly elastic.

(b) *Elastic shakedown* The elastic limit is exceeded and thus the plastic deformation will take place and lead to the build-up of residual stresses during early cycles. Generally such residual stresses are protective and make plastic deformation in subsequent cycles less likely. Consequently, the steady state of cyclic deformation is purely elastic.

(c) *Plastic shakedown* (*cyclic plasticity*) The steady cyclic state consists of a closed plastic stress-strain loop without any accumulation of plastic strains.

(d) *Ratcheting* (*cyclic creep* or *incremental collapse*) The steady state consists of an open cycle of plastic stress and strain and plastic strain is accumulated with each cycle of stress.

In terms of the Tresca yield criterion, the elastic shakedown limit for Hertzian line contact without friction is (Johnson 1985)

$$\frac{p_0}{k} = 4 \tag{5.52}$$

where p_0 is the maximum Hertzian pressure and k is the shear yield stress. For the case involving both rolling and frictional sliding, the shakedown limit is a function of the friction coefficient, μ. For a low friction coefficient ($\mu < 0.3$), yielding occurs beneath the surface. For higher values of friction ($\mu > 0.3$), the yield point moves to the surface and the shakedown is found to be (Johnson 1989, Suresh 1998)

$$\frac{p_0}{k} = \frac{1}{\mu} \tag{5.53}$$

Thus, pure rolling has a shakedown limit which is 4μ times larger than the shakedown limit for tractive rolling with $\mu > 0.3$. This behaviour can be illustrated in a shakedown map, which has been given by Johnson (1989) for a perfectly plastic and a kinematically hardening material in line contact. Morales-Espejel et al. (1999) analysed the shakedown process for both dry (with different friction coefficients) and rough elastohydrodynamically lubricated contacts. The results showed that the roughness produces a reduction of the required yield load.

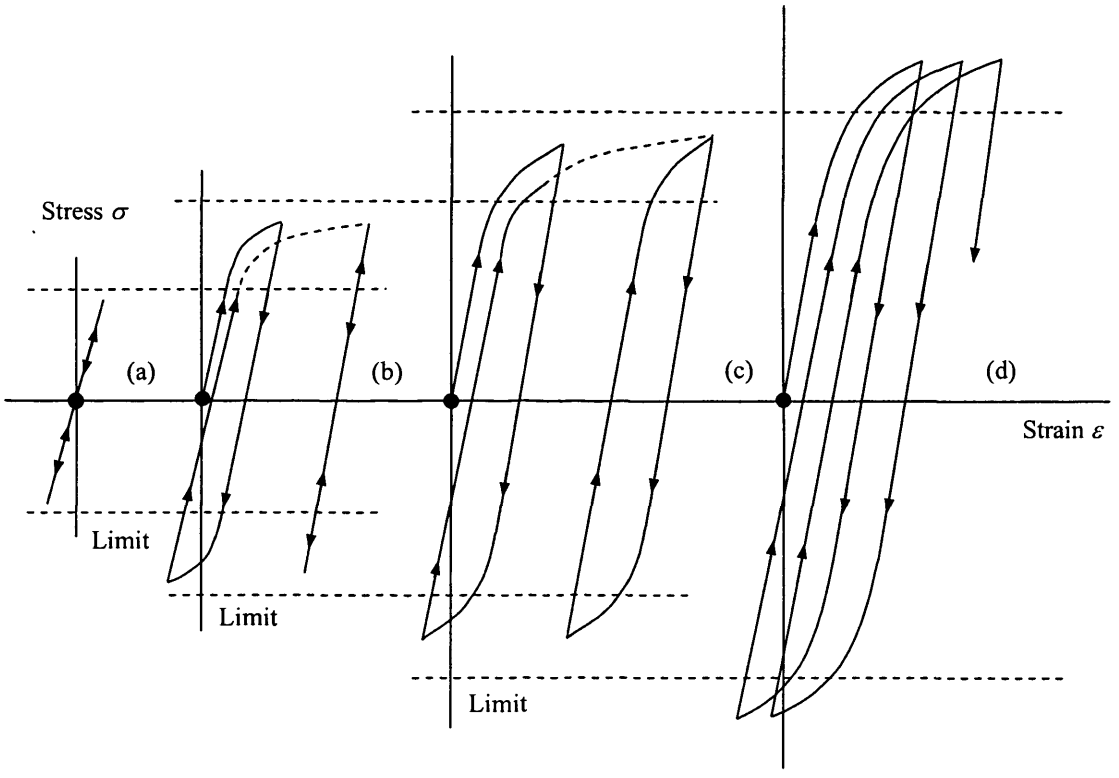


Figure 10: Material response to cyclic loading: (a) Perfectly elastic; (b) Elastic shakedown; (c) Plastic shakedown; (d) rachetting.

5.6.2 Models for Rolling Contact Fatigue

Rolling contact has a wide range of engineering applications such as the rail/wheel, rolling bearings, traction drives, gears and cams (Johnson 1989). Therefore, the understanding of rolling contact fatigue has become a crucial problem for reasons of both safety and economy. It is no surprise that it has attracted the attention of many researchers for many years in both experimental and theoretical studies.

The rolling contact fatigue process is influenced by a large number of parameters and tests have been performed to evaluate the roles of these parameters. For example, surface treatments such as shot-peening (Girish et al. 1997, Guagliano et al. 2002), surface hardening (Wilkinson & Olver 1999), improved surface finish (Krantz et al. 2001), etc. can improve pitting fatigue life of gears significantly. Widmark and Melander (1999) studied experimentally how a number of factors, including the effect of carburizing steel alloy composition, the sequence order of grinding and carburizing and the effect of shot peening, influence the contact fatigue resistance. Stewart and Ahmed (2002) reviewed investigations of various surface coatings on the rolling contact fatigue performance. Other important parameters which influence rolling contact fatigue include relative slip (Kimura et al. 2002), lubricant viscosities and additives (Song et al. 1996, Fernandez Rico et al. 2003, Krantz & Kahraman 2004), inclusions and operating conditions (Nelias et al. 1999), thermal effects (Hohn & Michaelis 2004, Seireg 2001), etc. With so many parameters involved, design work to determine rolling contact fatigue life is extremely difficult.

So far, many mathematical models for predicting rolling contact fatigue have been published including, as classified by Tallian (1992), engineering models and research models. The engineering models are practical but empirical, presuming that several material and environmental parameters are known. To establish these parameters, a large number of life tests are required. These models try to calibrate the final life equations to the experimental data and their applications may be limited. Tallian (1992) and Kudish & Burris (2000) reviewed tens of engineering models. Most reviewed engineering models are mainly adopted for rolling bearings. Representative engineering models are those by Ioannides and Harris (1985), Tallian (1992, 1996, 1999), Zaretsky (1994), etc. These models grew out of

the Weibull and Lundberg-Palmgren theories and were originally developed for rolling bearings. In the Ioannides-Harris model, the survival probability is expressed as an integral over the stressed material volume and the integrand includes the fatigue life and the effective stress above a fatigue limit. The Tallian model attempted to include many of the factors influencing life, such as contact geometry, material fatigue parameters, defect severity, applied normal and frictional stresses, Λ ratio, etc. In the Zaretsky model, the Weibull equation was rewritten by making the stress exponent independent of the Weibull slope. Like the Ioannides-Harris model, Zaretsky (1987) and August and Zaretsky (1993) developed a generalized methodology for system life prediction that uses an elemental stressed volume approach.

The research models are based on theories of material fatigue failures and therefore often require complete information of stress-strain behaviour of the material under repeated rolling contact. Most fatigue models for general loading, as described in previous sections, can be applied to rolling contact fatigue problems. For example, Ekberg (1997) developed a fatigue life model for rolling contact with application to rail/wheel damage based on the Dang Van criterion; Kim et al. (2003) calculated the maximum Hertz pressure corresponding to the material fatigue limits and the site of crack initiation using the Matake, the Crossland and the Dang Van fatigue criterion; Sraml et al. (2003) analysed crack initiation with moving Hertzian contact pressure and Coulomb friction using the Coffin-Manson relation, the Morrow equation and the Smith-Watson-Topper (SWT) equation, and later Dang Van criterion (Sraml et al. 2004); Ringsberg et al. (2000) assessed the initiation of cracks using the Dang Van criterion, the Coffin-Manson relation, the SWT relation, and a ratchetting failure model for elastic shakedown, plastic shakedown and ratchetting material responses; Ringsberg (2001) presented a strategy for rolling contact fatigue crack initiation based on the failure model proposed by Jiang and Sehitoglu (1999), which is a mixed 'shear' and 'normal' crack model; etc. These models are for crack initiation, assuming that nearly all the life consists of the initiation phase. For crack propagation, a number of solutions to simulation of surface pitting have been obtained based on fracture mechanics, such as the work by Glodez et al. (1998), Flasker et al. (2001), Datsyshyn and Panasyuk (2001), Kudish (2002), Kudish and Burris (2004), etc.

The study of EHL with rough surfaces leads naturally to a better understanding of micro-EHL or mixed lubrication. Mixed lubrication is a particularly challenging and important regime to understand as, in truth, almost all failures of surfaces in rolling/sliding contact with an EHL lubricant film occur in this regime. The Λ ratio has been widely recognized as a dominant factor in determining the fatigue life for this situation. Some experimental work has been performed to predict the relation between the fatigue life and the Λ ratio for gears (Gao et al. 2005). In the low Λ regime, however, fatigue life depends more strongly on surface roughness than on film thickness and the occurrence of micropitting is more load dependent than the Λ ratio (Chiu 1997). In recent years theoretical analyses have begun to emerge for mixed lubrication and some recent developments seek to extend the evaluation of pressure variation to include their effect on fatigue life due to roughness and cyclic loading. Ioannides et al. (1989) applied the Ioannides-Harris model to study the effects of lubricant films, roughness, contaminants and residual stress on fatigue life. The same model was used by Kim and Olver (1997, 1998) to study the fatigue failure in rolling sliding contact with rough surfaces. Ai (1998) used material's S-N-P curve with an effective stress based on damage accumulation theory to evaluate the surface roughness effect on contact fatigue life under lubricated conditions. Tallian (1999) extended his rolling bearing life prediction model to cover variable operating conditions including the Λ ratio. Epstein et al. (2003a, 2003b) calculated fatigue life in mixed lubrication using the Zaretsky fatigue model. The effects of surface texture on fatigue life were presented by a variety of surfaces with different finishes.

5.7 Summary

In this chapter, various aspects of the theory of fatigue are presented. Models for rolling contact fatigue have been reviewed. For mixed lubrication problems, only a limited number of papers has been published to predict the effects of surface roughness on fatigue life. Most of the models used for the evaluation of fatigue life are engineering models. For a comprehensive understanding of the mechanisms of fatigue under mixed lubrication conditions, it is necessary to apply both engineering and research models.

As stated in this chapter, a number of models have been proposed for the prediction of fatigue damage. Since many influencing factors exist in fatigue analysis, there is as yet no universally accepted method (Fatemi & Socie 1988). However, it has been reported that better predictions are often achieved using a critical plane approach (Chu 1995, You & Lee 1996). Therefore, the critical plane approach will be used in the next two chapters for analysing fatigue damage in mixed lubrication line contacts.

Chapter 6

Fatigue Analysis – Applying Multiaxial Fatigue Criteria

6.1 Introduction

As is well-known, surface contact fatigue is the most common failure mode of surfaces subjected to repeated contacts. A typical example is fatigue failures in gears (Fernandes & McDuling 1997). As stated in Chapter 5, rolling and sliding contact involves complex time-varying loadings and like many engineering components and structures stress states are multiaxial rather than uniaxial. However, it is usual to rely on uniaxial fatigue test data for life prediction in practice. Consequently, several multiaxial fatigue criteria have been proposed to reduce the multiaxial stress state to an equivalent uniaxial stress condition. Some of these criteria are based on the critical plane approach, which often gives better fatigue predictions than other approaches. In the following, multiaxial fatigue criteria based on the critical plane approach are used to perform fatigue analysis for the EHL line contact with rough surfaces.

6.2 Multiaxial Fatigue Criteria Based on Critical Plane Approach

The criteria based on the critical plane approach, such as the Findley criterion, the Matake criterion, and the Dang Van criterion, are applied to the EHL line contact fatigue

analysis. The review of these criteria was given in Chapter 5. These criteria can be written in a general form as

$$A + \bar{\kappa} B \leq \bar{\lambda} \tag{6.1}$$

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. $\bar{\kappa}$ and $\bar{\lambda}$ take different values for different criteria, and can be determined from the endurance limits under fully reversed bending and fully reversed torsion tests. A summary of these parameters is given in Table 6.1, 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 Dang Van criterion uses the current values of the shear stress amplitude on the critical plane, $\tau_a(t)$, and hydrostatic stress, $\sigma_h(t)$, and t is time.

Table 6.1 Expressions of parameters for criteria considered.

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

The analysis procedure to use the critical plane approach is divided into two steps: (1) to determine the critical plane based on parameters A and B; (2) to assess the fatigue behavior using Equation (6.1). From Table 6.1, it can be seen that calculation of the amplitude and mean value of the shear stress acting on the critical plane is required to apply these fatigue criteria in the first step.

In a general three-dimensional non-proportional loading case, the evaluation of the amplitude and mean value of the shear stress on a plane is not a trivial problem. The definition of τ_a and τ_m has been given by Papadopoulos (1998), in what is called a minimum-circumscribed circle proposal. It is based on the construction of the minimum circle circumscribed to a plane curve described by the shear stress vector on the plane. The radius of the circle is defined to be the shear stress amplitude and its centre defines the tip of the mean shear stress vector. Algorithms to determine τ_a and τ_m , such as the points combination algorithm, the incremental method and the optimization method, have been presented by Bernasconi (2002).

Since the EHL line contact is a plane problem, only the x-z plane should be examined. This simplifies the calculation of τ_a and τ_m because the shear stress on the critical plane becomes a scalar. As shown in Figure 6.1, the normal and shear stress history on a plane can be calculated from its directional components by (Timoshenko & Goodier 1970)

$$\sigma(t) = \sigma_x \cos^2\theta + \sigma_z \sin^2\theta + 2\tau_{xz} \sin\theta\cos\theta \tag{6.2a}$$

$$\tau(t) = \tau_{xz}(\cos^2\theta - \sin^2\theta) + (\sigma_z - \sigma_x)\sin\theta\cos\theta \tag{6.2b}$$

where θ is the angle between the normal of the plane and the x-axis. Thus, the amplitude and mean value of the shear stress on the plane are determined by

$$\tau_a = \frac{\tau_{\max} - \tau_{\min}}{2} \tag{6.3a}$$

$$\tau_m = \frac{\tau_{\max} + \tau_{\min}}{2} \tag{6.3b}$$

where $\tau_{\max}$ and $\tau_{\min}$ are the maximum and minimum values of the shear stress on the plane, respectively. The Dang Van criterion uses the current shear stress amplitude, which is calculated as (Ekberg 1997)

$$\tau_a(t) = \left|\tau(t) - \tau_m\right| \tag{6.4}$$

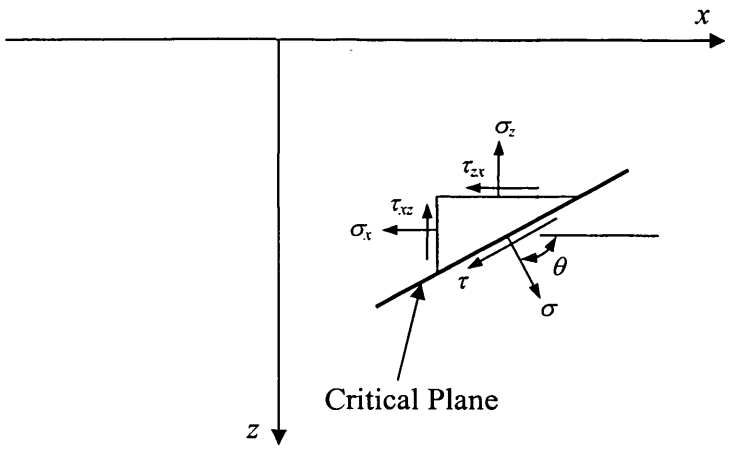


Figure 6.1: Definition of a critical plane.

Once parameters A and B are obtained, the next step is to evaluate the fatigue failure. A fatigue parameter (FP) is defined from Equation (6.1) as

$$FP = \frac{A + \bar{\kappa}B}{\bar{\lambda}} \tag{6.5}$$

With this definition, if FP is greater than unity, fatigue failure is expected to occur. For the Dang Van criterion, since both A and B use current stress values, the fatigue parameter is defined as

$$FP = \max_{t}\left(\frac{A + \bar{\kappa}B}{\bar{\lambda}}\right) \tag{6.6}$$

where the maximisation is carried out over time (t).

Since the critical plane through a given material point is not known before the analysis, FP has to be examined on all the candidate planes in order to identify the critical plane. Further, the above analysis procedure has to be repeated for all the material points to be examined.

6.3 Estimation of Failure Probability

The fatigue model developed by Ioannides and Harris (1985) is used to estimate the failure probability. The model can be written as

$$\ln\frac{1}{S} = \overline{A}N^{e}\int_{V}\frac{(\tau_e - \tau_u)^{c}}{z'^{h}}dV \tag{6.7}$$

where S is the probability of survival, N is the fatigue life, τ_e is equivalent stress, τ_u is the threshold fatigue limit stress below which no failure would occur, z' is the stress weighted average depth, and V is stressed volume. $\overline{A}$, e, c, and h are empirical constants to be determined from experimental data. The depth factor and the fatigue limit stress have been discussed by many researchers (Zaretsky et al. 1996, Kim & Olver 1997, Epstein et al. 2003b) and some of these authors prefer to ignore these parameters. Thus, the following form will be used in this study:

$$\ln\frac{1}{S} = \overline{A}N^{e}\int_{V}\tau_e{}^{c}dV \tag{6.8}$$

The equivalent stress τ_e can be chosen based on multiaxial fatigue criteria in different ways. Ioannides et al. (1989) listed some examples of criteria including maximum Hertzian pressure, orthogonal shear stress, maximum shear stress, maximum surface traction, von Mises stress and maximum shear amplitude regardless of plane and direction, and employed a Dang Van type modified shear stress in their fatigue analysis. This Dang Van type equivalent stress was also used by Kim and Olver (1997, 1998). Epstein et al. (2003a, 2003b) used the orthogonal shear stress in their fatigue analysis and stated that using different forms of equivalent stress had only a minor effect on the final results. In this study, the equivalent stress is calculated based on the Dang Van criterion as

$$\tau_e = FP \times \overline{\lambda} = \max_{t}\left[\tau_a(t) + \overline{\kappa}\sigma_h(t)\right] \tag{6.9}$$

6.4 Numerical Results

The critical plane based multiaxial fatigue criteria are evaluated for the EHL line contact between two rough surfaces each with the profile (C), which is the most aggressive of the three shown in Figure 2.2 of Chapter 2. The parameters used to specify the EHL problem are given in Table 6.2.

Table 6.2: Parameters used for transient EHL line contact analysis.

a	0.335 mm	γ	2.27 GPa^{-1}
E_1, E_2	200 GPa	κ	63.2×10^{-6} Pa.s
p_0	1.0 GPa	λ	1.68 GPa^{-1}
R'	0.0191 m	ν_1, ν_2	0.3
w'	527 kN/m	τ_0	10 MPa
α	11.1 GPa^{-1}	χ	5.1 GPa^{-1}

As shown in Figure 6.2, sections of the two surfaces of size $2a\times a$ (shaded) are selected for fatigue analysis, where a is the Hertzian semi-contact width. The selected material is located at $-4 \le x/a \le -2$ at the start of the analysis and then passes the contact area and moves to $2 \le x/a \le 4$. The stress history of the material during its traverse of the contact area is determined using the DC-FFT method as described in Chapter 4. The mesh is set up for stress and fatigue analysis in such a way that dense grids are used near the surface to capture the possible high stress gradients due to asperity contact. In this study, 20 depth levels are examined, i.e. z/a = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. In the x direction, a uniform mesh with 201 grid points is used. Thus, the total number of material points for stress and fatigue analysis is 4020.

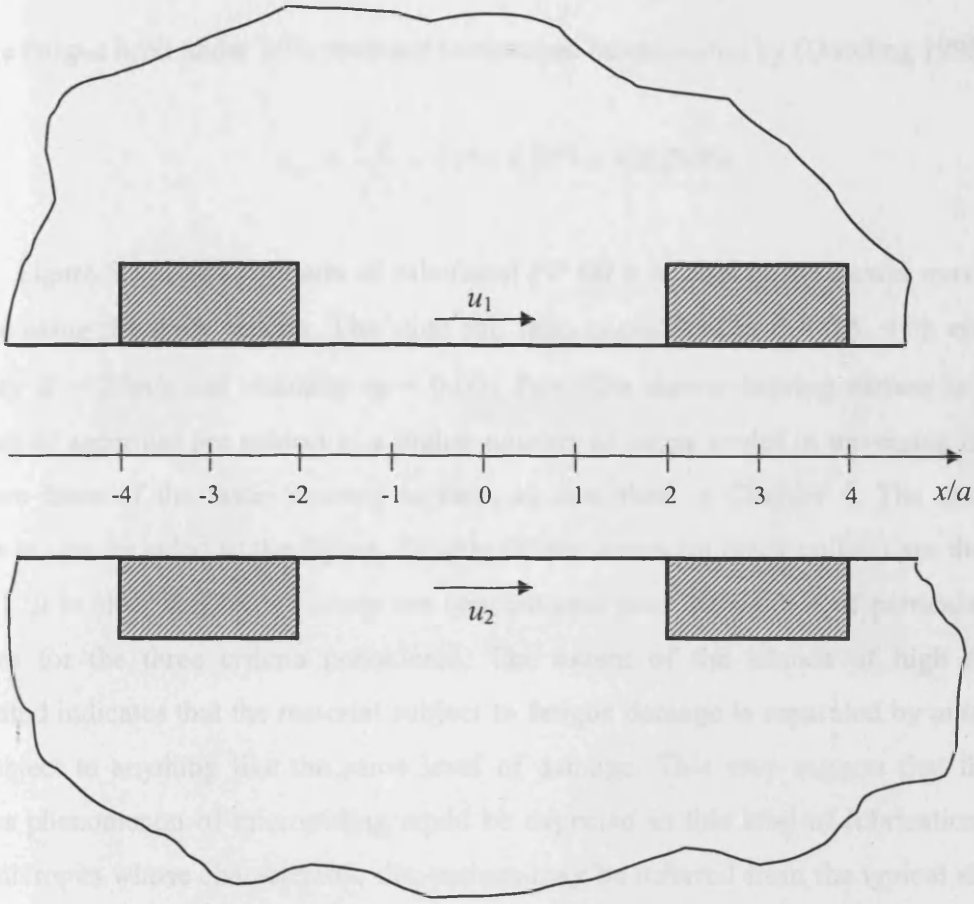


Figure 6.2: Volumes of materials for fatigue analysis.

The critical plane is defined by θ, which is the angle between the normal to the critical plane and the x-axis, as shown in Figure 6.1. The fatigue damage is computed for $0° \leq \theta \leq 180°$ with an interval of 1° to search for the critical plane.

The material assumed for the two solids is SAE4340 (BS970:En24) whose hardness is HB = 409 and ultimate tensile strength is σ_u = 1470MPa (Zahavi & Torbilo 1996). Its fatigue limit at $N = 10^7$ cycles under fully reversed bending can be approximated by (Ciavarella & Maitournam 2004)

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

and the fatigue limit under fully reversed torsion can be estimated by (Dowling 1998)

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

Figure 6.3 shows contours of calculated *FP* for a section of the slower moving rough surface using the three criteria. The slide roll ratio considered is ξ = 0.5, with entrainment velocity $\bar{u}$ = 25m/s and viscosity η_0 = 0.001 Pa·s. The slower moving surface is examined because its asperities are subject to a higher number of stress cycles in traversing the contact than are those of the faster moving surface, as described in Chapter 4. The shape of the profile is also included in the figure. Fatigue failure zones (in black colour) are those where $FP \geq 1$. It is clear that failure zones are concentrated near the surface of particular asperity features for the three criteria considered. The extent of the islands of high *FP* values calculated indicates that the material subject to fatigue damage is separated by areas that are not subject to anything like the same level of damage. This may suggest that the surface distress phenomenon of micropitting could be expected in this kind of lubrication situation with micropits whose characteristic dimensions may be inferred from the typical sizes of the islands of high *FP* values. For the current example which uses a rough surface profile taken from FZG gear testing this dimension would be of the order 50 μm. A summary of the application of these criteria to this analysis is presented in Table 6.3, which records ten asperities which have highest *FP* values. The ten asperities are labelled with letters of A to J and identified by the distinctive arrows in Figure 6.3. It can be seen that the Findley criterion and the Matake criterion give almost the same results. Using the Dang Van criterion, the calculated *FP* values are slightly lower for this case study and the sequence of asperities with decreasing *FP* values is reordered.

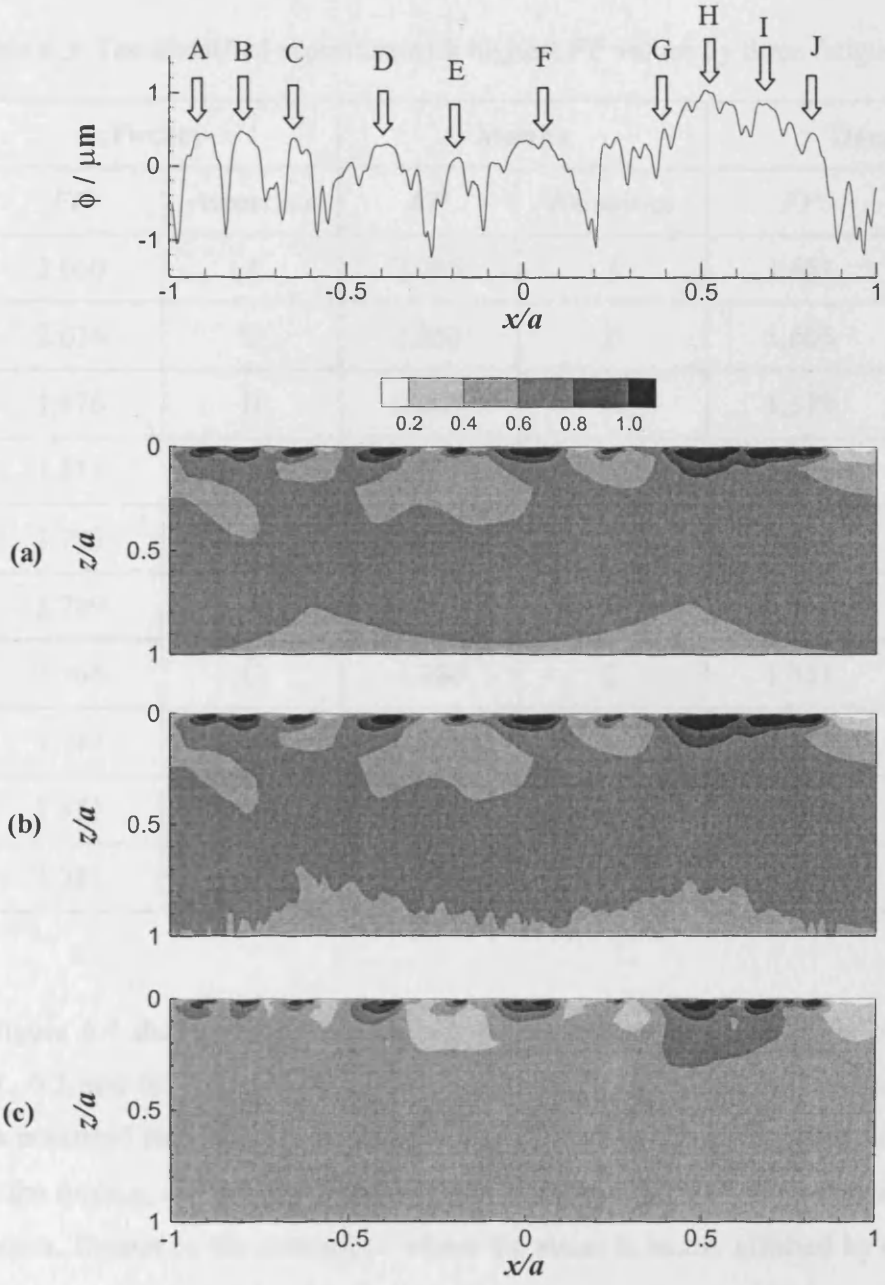


Figure 6.3: Contours of fatigue parameter (*FP*) for fatigue at 10^7 cycles determined by using (a) Findley, (b) Matake, and (c) Dang Van criteria. The top plot is the surface profile.

Table 6.3: Ten identified asperities with highest *FP* values by three fatigue criteria.

Index	Findley		Matake		Dang Van	
	FP	Asperities	*FP*	Asperities	*FP*	Asperities
1	2.060	J	2.085	J	1.651	H
2	2.036	D	2.053	D	1.605	J
3	1.976	H	2.000	H	1.579	G
4	1.811	I	1.825	I	1.468	D
5	1.797	E	1.810	E	1.444	I
6	1.789	A	1.804	A	1.399	C
7	1.766	C	1.780	C	1.331	E
8	1.747	B	1.767	B	1.327	A
9	1.551	F	1.570	F	1.215	B
10	1.321	G	1.294	G	1.212	F

Figure 6.4 shows the *FP* values plotted against x/a at depth levels z/a = 0.0, 0.01, 0.05, 0.1, 0.2 and 0.5 using these criteria. As the *FP* exceeds unity at a position, fatigue failure is predicted to occur. It can be seen that *FP* values have their greatest fluctuation at and near the surface, such as depths of $0.01a$ (3.4μm) and $0.05a$ (17μm), due to the influence of roughness. Deeper in the subsurface where the stress is hardly affected by the roughness, for example at depths of $0.2a$ (67μm) and $0.5a$ (169μm), *FP* lines become flat and lower. The highest damage values occur at the depth level of $0.01a$ (3.4μm), which is just below the surface.

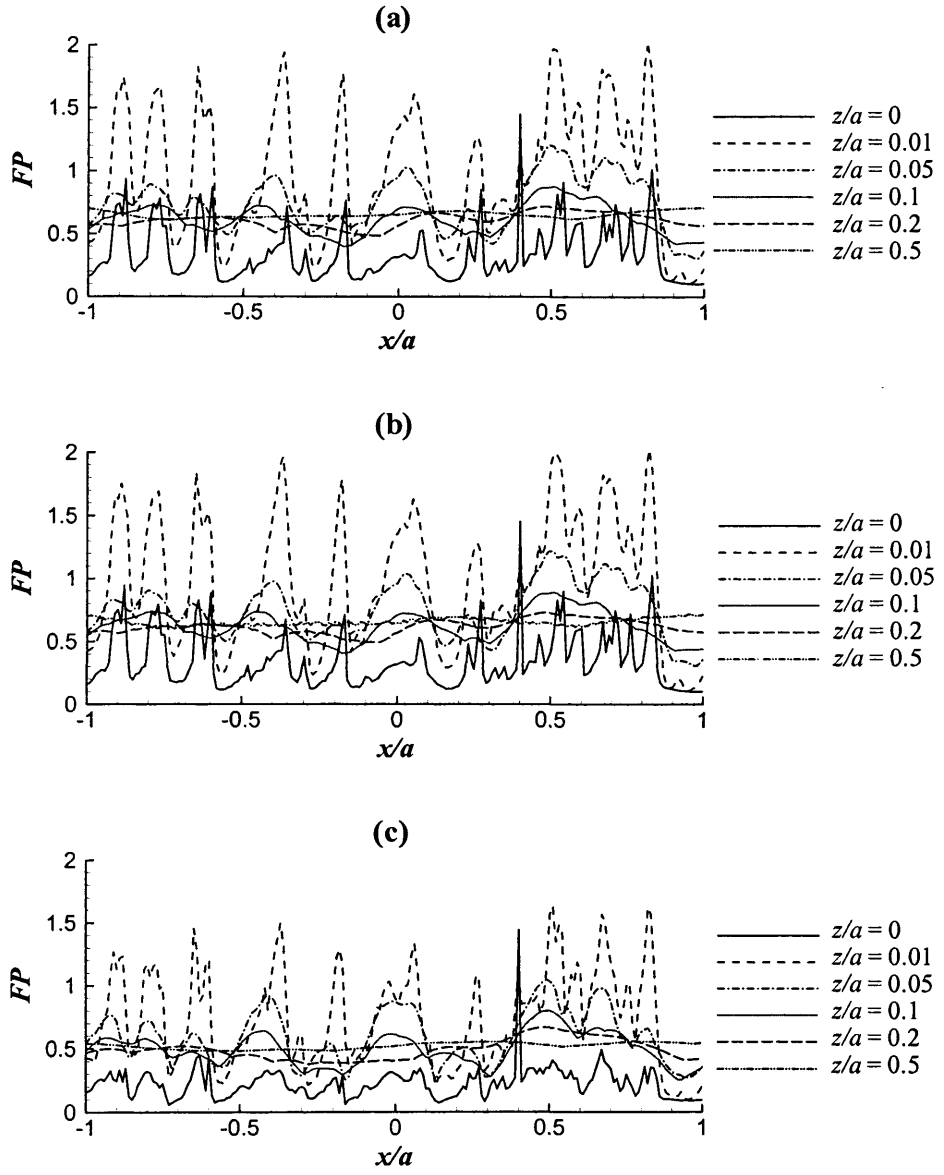


Figure 6.4: Distributions of fatigue parameter (*FP*) at 10^7 cycles at different depth levels by using (a) Findley, (b) Matake, and (c) Dang Van criteria.

To assess the fatigue damage at a particular depth level, the probability of failure F is determined by the Ioannides-Harris model, where $F = 1 - S$. The exponents, e and c, take the values adopted by Ioannides and Harris (1985): $e = 10/9$ and $c = 31/3$. Following the strategy used by Kim and Olver (1997), the parameter $\overline{A}$ is assumed to be related to the tensile strength of the material as follows:

$$\overline{A} = \frac{1}{\sigma_u{}^c V_{ref}} \tag{6.10}$$

where V_{ref} is the reference volume, which can be determined experimentally for a specific material. For comparative study of fatigue failures under different contact conditions, the actual value of V_{ref} is not important and is taken as the material volume considered in this study, i.e. $V_{ref} = 2a \times a$. The element volume for a stressed point at a depth level is chosen as $\Delta V = \Delta x \times z_{ref}$, where Δx is the mesh size in the x direction and z_{ref} is a reference spacing in the z direction for a depth level. A small value of $z_{ref} = 10^{-4}a$ is chosen in this study by a trial and error approach. If z_{ref} is too large, the variation of the failure probability across the near surface layer, which has greatest stress gradients, may be overlooked. On the other hand, if z_{ref} is too small, then the calculated failure probability may be far smaller than unity. Therefore, the calculated F values obtained in this way are only stated for comparison. To predict actual fatigue life in practical application, experiments are required to determine these parameters in order to use the Ioannides-Harris model.

Figure 6.5 shows the probability of failure F against the depth level z/a for the case shown in Figure 6.3 based on the Dang Van equivalent stress. (Note the logarithmic scale for F in this and similar figures). It can be seen that F increases drastically starting from the surface and reaches the highest value in the near surface layer and drops sharply with deeper depth. Figure 6.6 compares the probability of failure F for four contact conditions: (i) smooth-smooth, (ii) smooth-rough, (iii) rough-smooth, and (iv) rough-rough. (The first surface has a higher speed than the second one. For example, smooth-rough implies the faster surface is smooth and the slower surface is rough.) The same slide roll ratio, velocity and viscosity are used for this comparison, i.e. $\xi = 0.5$, $\overline{u} = 25$m/s and $\eta_0 = 0.001$ Pa·s. It can be seen that fatigue risks, in order of severity, are those of rough-rough, rough-smooth, smooth-

rough and smooth-smooth. Their corresponding maximum *FP* values are summarised in Table 6.4. It is very interesting that two surfaces for case (ii) and case (iii) are the same but reversed, but they may have significantly different fatigue performance. This phenomenon has already been observed by Kim and Olver (1998) by examining the stress history and they concluded that sliding has no influence on case (ii) but a significant influence on case (iii). Except for the smooth-smooth contact, the fatigue risks for the other three cases are highest near the surface due to the influence of roughness and drop quickly deeper in the subsurface. For the smooth-smooth case, the highest fatigue risk is located near the Hertzian maximum shear stress zone as expected. All the four curves approach the same probability of failure in the zone where the roughness effect is no longer active.

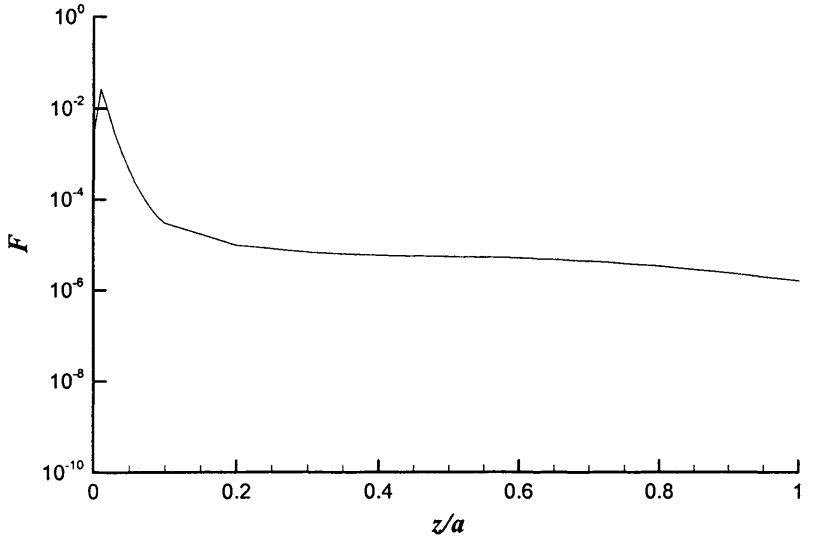


Figure 6.5: Fatigue failure probabilities versus depth for the slower moving surface.

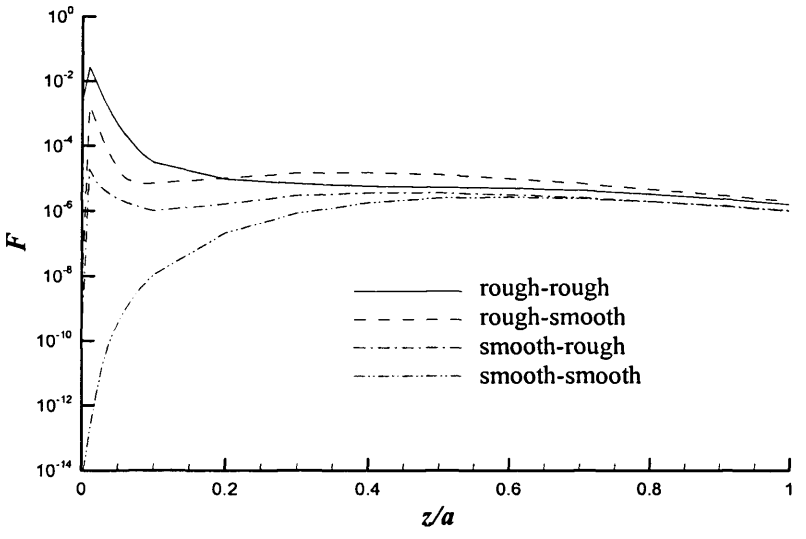


Figure 6.6: Comparisons of fatigue failure probabilities versus depth for four contact conditions.

Table 6.4: Maximum *FP* values obtained using the three fatigue criteria for four contact conditions.

	Findley	Matake	Dang Van
Smooth-smooth	0.644	0.647	0.489
Smooth-rough	0.940	0.951	0.873
Rough-smooth	1.762	1.774	1.261
Rough-rough	2.060	2.085	1.651

The effect of slide roll ratio on fatigue is studied with ξ values of 0.25, 0.5, 0.75 and 1.0 and the same entrainment velocity $\overline{u}$ = 25m/s and the viscosity η_0 = 0.001 Pa·s. The contours of calculated *FP* for the slower moving surface using the Dang Van criterion are shown in Figure 6.7. The volume of failure zones (black colour) is seen to increase with increased ξ values and the location of these failure zones is unchanged. Their corresponding maximum *FP* values obtained using the three fatigue criteria are summarised in Table 6.5. Figure 6.8 compares the corresponding *F* curves across the depth for the four slide roll ratios. It can be seen that higher sliding generally leads to greater failure probabilities. In this comparison it is necessary to point out that the "collisions" experienced by each asperity on the surface will differ according to the specified slide roll ratio. The number of "collisions" during a traverse of the contact area will increase with slide roll ratio due to the increase in relative velocity of the surfaces, as shown and discussed in Chapter 4. Also the part of the opposing surface passing through the contact area in direct interaction with the section considered will change according to the sliding speed specified. Therefore, it is not surprising that the maximum *FP* value obtained does not strictly increase with increasing slide roll ratio, as observed in Table 6.5.

Table 6.5: Maximum *FP* values obtained using the three fatigue criteria for different slide roll ratios.

Slide roll ratio (ξ)	Findley	Matake	Dang Van
0.25	2.187	2.214	1.553
0.5	2.060	2.085	1.651
0.75	2.138	2.155	1.420
1.0	2.328	2.280	2.810

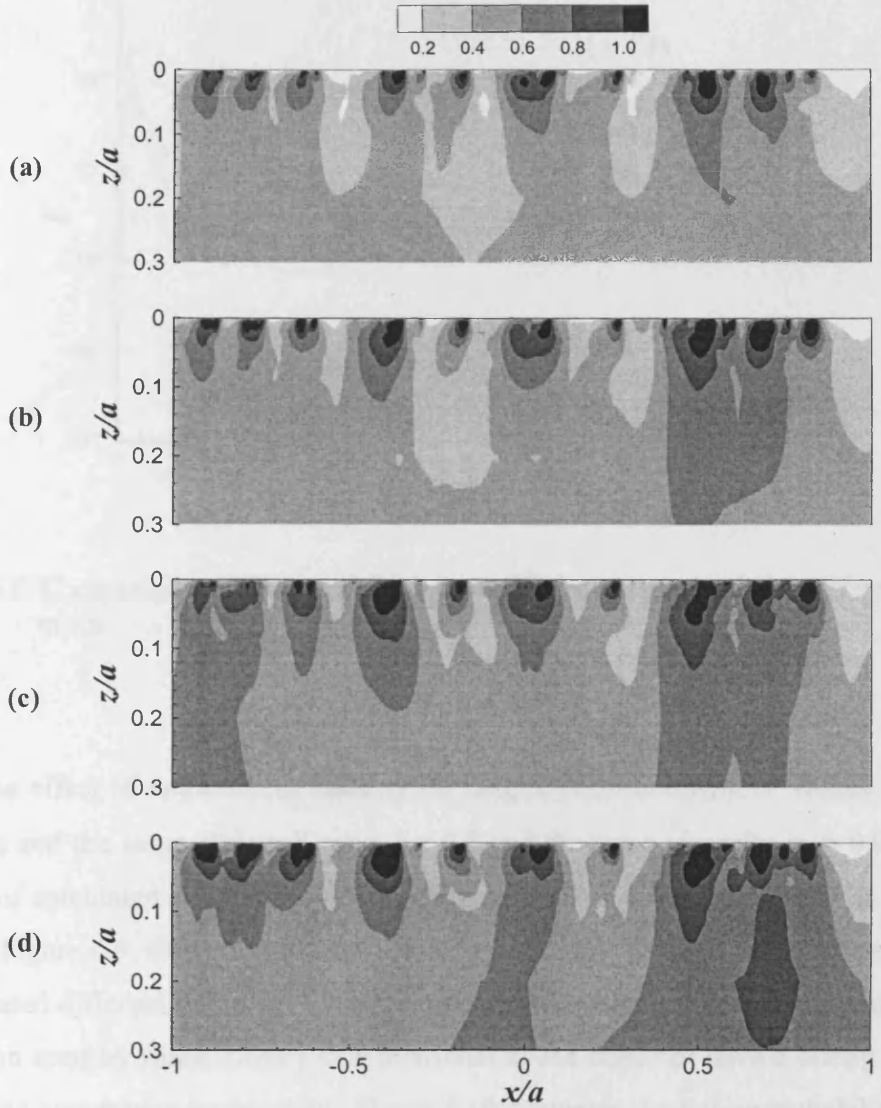


Figure 6.7: Contours of fatigue parameter (*FP*) at 10^7 cycles obtained using the Dang Van criterion for the section of the slower moving surface illustrated in Figure 6.3 with different slide roll ratio: (a) $\xi = 0.25$, (b) $\xi = 0.5$, (c) $\xi = 0.75$, and (d) $\xi = 1.0$.

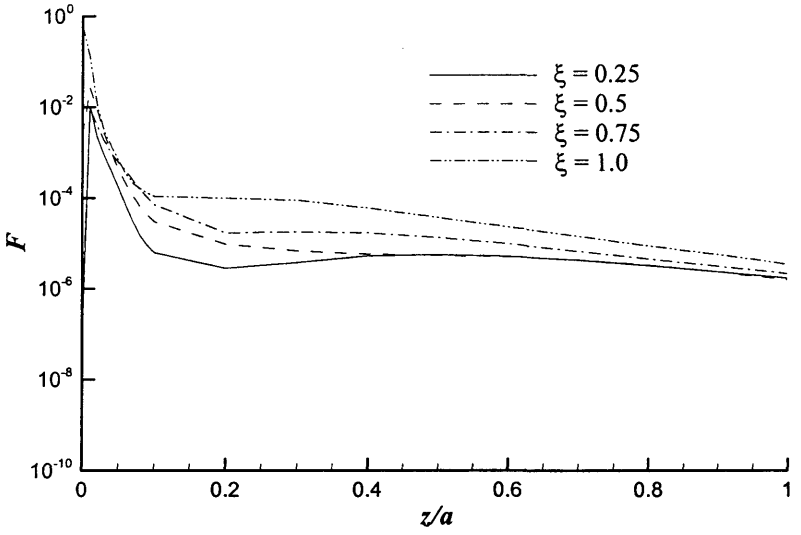


Figure 6.8: Comparisons of fatigue failure probabilities versus depth for different slide roll ratios.

The effect of entrainment velocity on fatigue is studied with $\bar{u}$ values of 15, 20, 25 and 30m/s and the same slide roll ratio $\xi = 0.5$ and the same viscosity $\eta_0 = 0.001$ Pa·s. The contours of calculated *FP* for the slower moving surface using the Dang Van criterion are shown in Figure 6.9. Comparing these contours shows that by using different velocity values the calculated differences in the *FP* values appear to be minor. This is in agreement with the observation seen by Shaw (2003) that rotational speed does not have a strong influence on micropitting occurrence in gear tests. Figure 6.10 compares the failure probability against the depth for these four cases. Only minor differences are seen between four curves once again indicating velocity has no strong influence on fatigue. The largest discrepancies are seen at the near surface layers ($z/a < 0.05$) and the failure probability tends to increase with decreasing velocity. This can be explained by the fact that increasing velocity will lead to increasing film thickness and decreasing asperity interactions. Since the Ioannides-Harris model is very sensitive to the stress term due to the high value of exponent c (= 31/3), a single high stress value may lead to high failure probability. Table 6.6 lists the corresponding maximum *FP* values obtained using the three fatigue criteria.

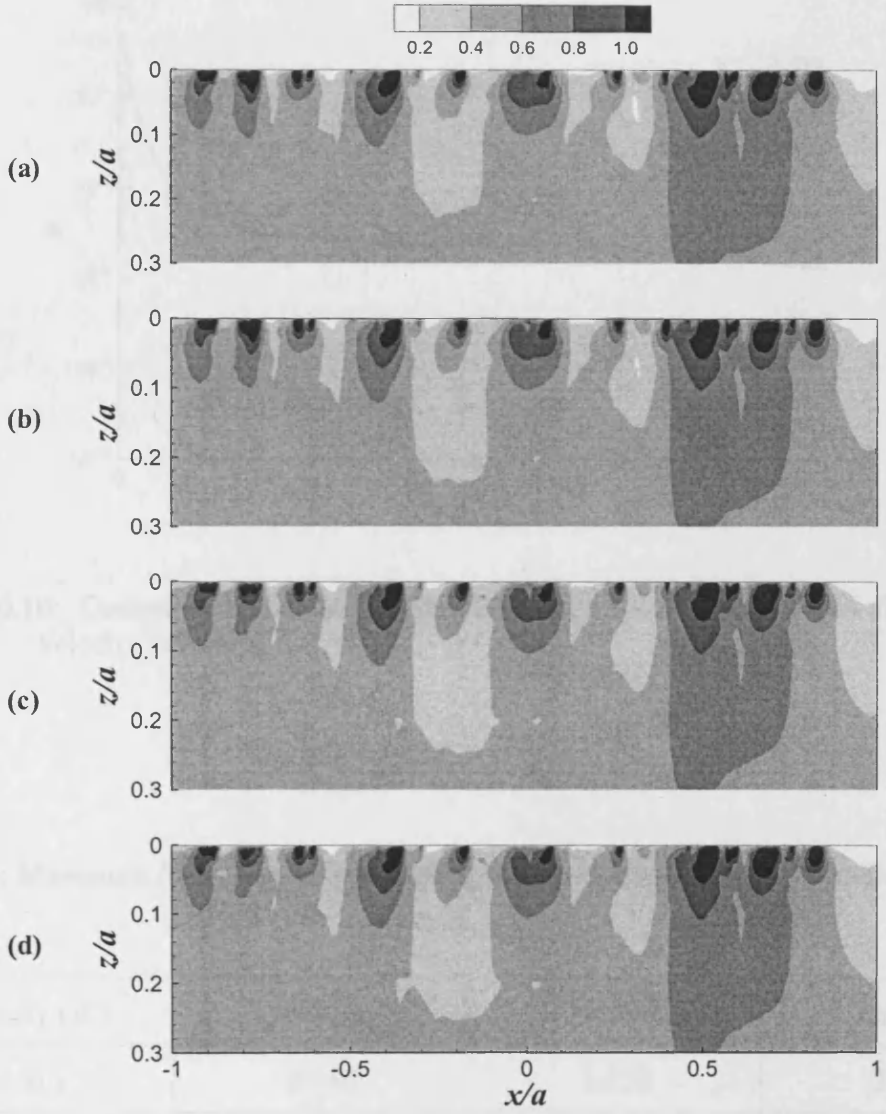


Figure 6.9: Contours of fatigue parameter (*FP*) at 10^7 cycles using the Dang Van criterion for the section of the slower moving surface illustrated in Figure 6.3 with different entrainment velocity: (a) $\overline{u}$ = 15, (b) $\overline{u}$ = 20, (c) $\overline{u}$ = 25, and (d) $\overline{u}$ = 30m/s.

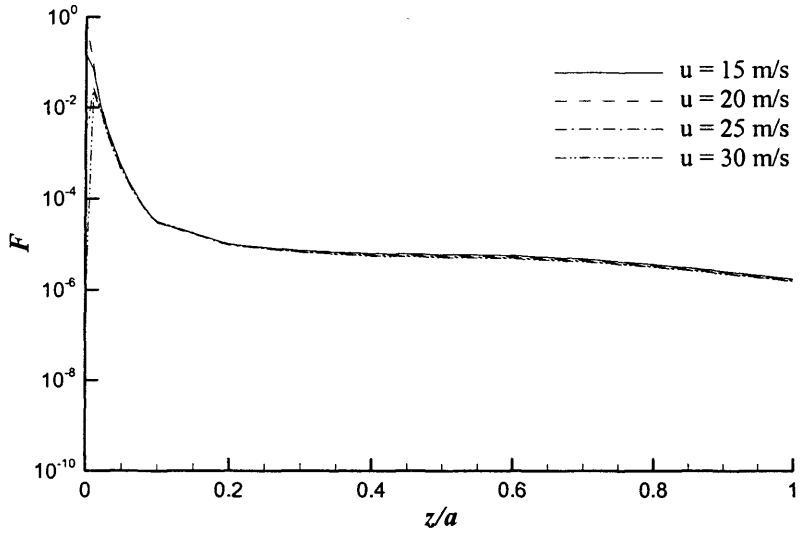


Figure 6.10: Comparisons of fatigue failure probabilities versus depth for different velocity values.

Table 6.6: Maximum *FP* values obtained using the three fatigue criteria for different velocity values.

Velocity ($\bar{u}$)	Findley	Matake	Dang Van
15 m/s	2.140	2.159	2.364
20 m/s	2.627	2.611	3.218
25 m/s	2.060	2.085	1.651
30 m/s	2.045	2.064	1.630

The effect of viscosity on fatigue is studied with η_0 values of 0.001, 0.002, 0.004 and 0.008 Pa·s and the same slide roll ratio $\xi = 0.5$ and the same velocity $\overline{u} = 25$m/s. Figure 6.11 shows the *FP* contours calculated for the slower surface using the Dang Van criterion. The volume of failure zones (black colour) is seen to decrease with increased η_0. The maximum *FP* values obtained using the three fatigue criteria are summarised in Table 6.7. The Λ ratios are also listed in the table for comparison. It is clear that increasing viscosity will improve the fatigue performance. At η_0 = 0.008 Pa·s, the three fatigue criteria may give different answers; that is, the Findley criterion and the Matake criterion predict fatigue failures, whereas the Dang Van criterion predicts no fatigue failures, but the trend with increasing viscosity is the same for each model. The corresponding curves of failure probability against the depth level are shown in Figure 6.12. It can be seen that fatigue failure probability decreases with increasing viscosity. This is due to the fact that greater viscosity generally leads to an increase in film thickness; as a result asperity interaction decreases and thus the micro-EHL phenomenon becomes weak. This observation is in agreement with the experimental results presented by Fernandez Rico et al. (2003).

Table 6.7: Maximum *FP* values obtained using the three fatigue criteria for different viscosity values.

Viscosity (η_0)	Λ ratio	Findley	Matake	Dang Van
0.001 Pa·s	0.384	2.060	2.085	1.651
0.002 Pa·s	0.611	1.927	1.947	1.554
0.004 Pa·s	0.973	1.627	1.645	1.336
0.008 Pa·s	1.548	1.170	1.184	0.974

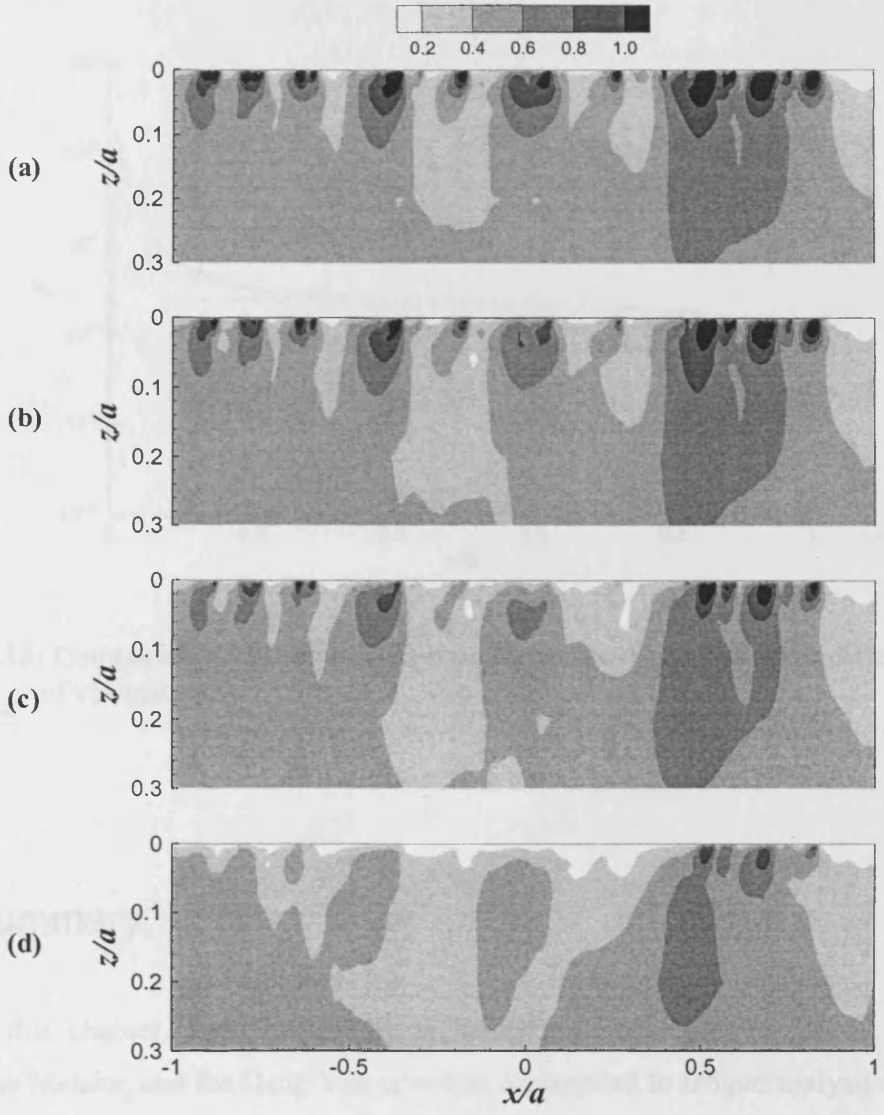


Figure 6.11: Contours of fatigue parameter (*FP*) at 10^7 cycles using the Dang Van criterion for the section of the slower moving surface illustrated in Figure 6.3 with different viscosity: (a) $\eta_0 = 0.001$, (b) $\eta_0 = 0.002$, (c) $\eta_0 = 0.004$, and (d) $\eta_0 = 0.008$Pa·s.

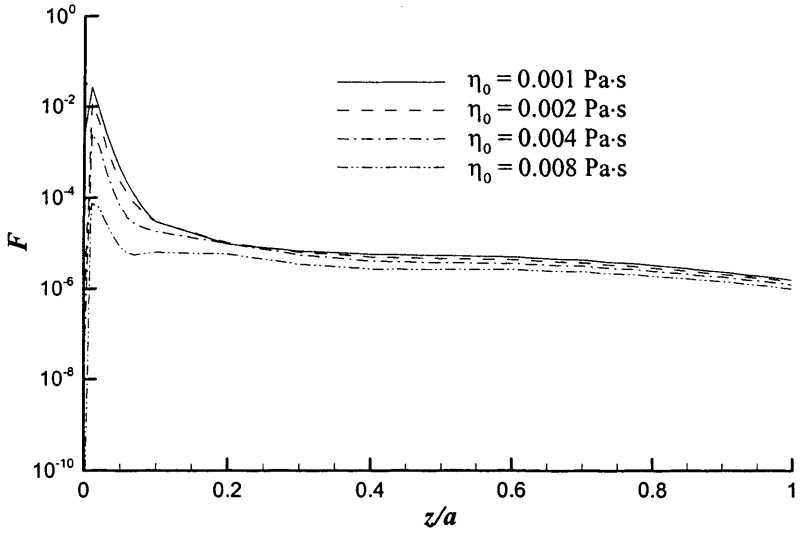


Figure 6.12: Comparisons of fatigue failure probabilities versus depth for different values of viscosity.

6.5 Summary

In this chapter, three critical plane based multiaxial fatigue criteria, namely the Findley, the Matake, and the Dang Van criterion, are applied to fatigue analysis for EHL line contact of two rough surfaces. The influence of roughness on predicted fatigue damage is investigated. From the numerical results it can be seen that the fatigue damage increases and moves to the near-surface layer when roughness is present. Prominent asperities are, not surprisingly, subject to greater damage risks than those of similar shape that are at a lower level than their neighbours. The investigation has shown that the various multiaxial fatigue criteria considered give similar results for this particular application.

The effects of slide roll ratio, entrainment velocity and lubricant viscosity on fatigue are investigated using the three criteria considered together with the Ioannides-Harris fatigue

model. The results have shown that slide roll ratio and viscosity have a stronger influence than velocity on fatigue. Decreasing slide roll ratio and increasing viscosity can improve fatigue performance.

Chapter 7

Fatigue Analysis – Applying Variable Amplitude Multiaxial Fatigue Theories

7.1 Introduction

In Chapter 6, several critical plane based multiaxial fatigue criteria were applied to fatigue analysis for EHL line contact problems. These criteria concern a specific fatigue life and only deal with the boundary between safety and failure for that life.

For the micro or mixed lubrication problem, the stress history shows that severe stress cycling with variable amplitude occurs in the material during the traverse of the contact area, especially close to the surface. Consequently, it is reasonable to apply variable amplitude multiaxial fatigue theories to analyse the effects of rough surfaces on fatigue. This chapter presents a variable amplitude multiaxial fatigue damage estimation procedure based on a critical plane approach. Numerical simulations are performed and results are presented for several cases of line contact of rough surfaces.

7.2 A Shear Strain Based Critical Plane Approach

7.2.1 Fatigue Models

As stated in Chapter 5, a number of fatigue models have been proposed for the prediction of fatigue life of components subjected to a complex loading. Although there is no universally accepted approach to fatigue analysis, it has been widely recognized that better predictions are often achieved by critical plane models. Critical plane models are based on a physical interpretation of the fatigue process whereby cracks form and grow on critical

planes (Bannantine & Socie 1992). In such an approach, stresses and strains are often evaluated for a number of planes at the same spatial position to determine the critical plane.

Cracking behaviour is dependent on material type and loading magnitude (Bannantine & Socie 1992, Jiang 2000). As shown in Figure 7.1, materials generally exhibit shear, tensile or mixed cracking behaviour (Jiang 2000). For tensile cracks, cracks are initiated in the direction of maximum normal amplitude and a fatigue model based on the Smith-Watson-Topper (SWT) parameter can be used to estimate the fatigue life:

$$\sigma_{\max}\varepsilon_a = \frac{(\sigma_f')^2}{E}(2N_f)^{2b} + \sigma_f'\varepsilon_f'(2N_f)^{b+c} \tag{7.1}$$

where ε_a is the amplitude of normal strain on the critical plane and $\sigma_{\max}$ is the maximum normal stress on the same plane as ε_a. For shear dominated fatigue damage, cracks are formed in the direction of maximum shear amplitude and a fatigue model proposed by Fatemi and Socie (1988) resulted in better correlation of fatigue lives:

$$\gamma_a\left(1 + K\frac{\sigma_{\max}}{\sigma_o'}\right) = \frac{\tau_f'}{G}(2N_f)^b + \gamma_f'(2N_f)^c \tag{7.2}$$

where γ_a is the amplitude of shear strain on the critical plane, and $\sigma_{\max}$ is the maximum tensile stress normal to that plane. For the material with mixed cracking behaviour, cracks are initiated on the maximum normal planes for tension loading but on the maximum shear planes for torsion loading. A model combining both shear and normal components was developed to consider the different cracking behaviour (Chu 1995, Jiang & Sehitoglu 1999):

$$\sigma_{\max}\varepsilon_a + J\Delta\tau\Delta\gamma = \frac{(\sigma_f')^2}{E}(2N_f)^{2b} + \sigma_f'\varepsilon_f'(2N_f)^{b+c} \tag{7.3}$$

where $\sigma_{\max}$ and ε_a are the maximum stress and the strain amplitude, respectively, normal to the crack plane; $\Delta\gamma$ and $\Delta\tau$ are the shear strain range and the shear stress range, respectively, on the crack plane. J is a material constant and should be obtained from tension/torsion tests. If $J = 0$, Equation (7.3) is reduced to the Equation (7.1) for tensile cracks. Equation (7.3) predicts shear cracking if J takes on a large value.

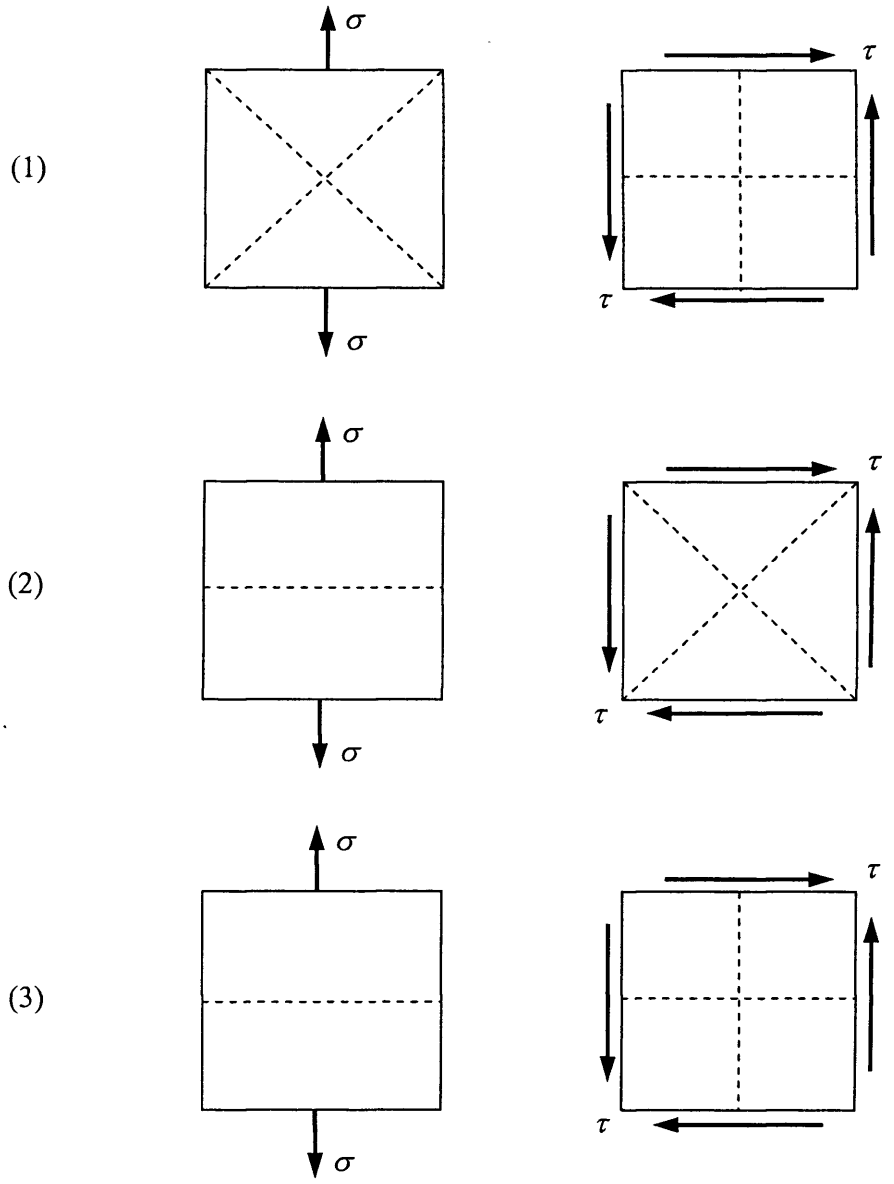


Figure 7.1: Fatigue cracking behaviour under tension and torsion loading: (1) shear cracking, (2) tensile cracking and (3) mixed cracking.

For rolling contact fatigue initiation, there is evidence that cracks take place in shear under hydrostatic compression (Johnson 1989). Therefore, the Fatemi and Socie (1988) shear strain model will be used to predict the variable amplitude multiaxial fatigue damage in this chapter.

7.2.2 Numerical Procedure

The procedure to estimate the fatigue damage of a material subjected to variable amplitude multiaxial loading is presented in Figure 7.2. This procedure is based on the Fatemi and Socie (FS) shear strain based critical plane model. The critical plane can be defined as the plane of maximum shear strain range, $\Delta\gamma_{max}$. Alternatively, Bannantine and Socie (1992) defined the critical plane as the plane experiencing maximum damage, D_{max}. Comparing the predicted and experimental lives, Kim and Park (1999) concluded that either the $\Delta\gamma_{max}$ plane or the D_{max} plane can be accepted as the critical plane. Since the latter definition accounts for the cumulative damage effect for variable amplitude loading, it is followed in this chapter.

As shown in Figure 7.2, for each material point and each plane, the solution procedure can be divided into the following steps:

- Establish stress and strain histories for the material point.
- Rotate stresses and strains to the plane analysed and obtain shear and normal stress and strain histories.
- Count cycles based on the shear strain history on the plane analysed.
- Estimate the damage for each identified cycle, using the Fatemi and Socie shear strain fatigue model.
- Calculate the total damage by summing the damages of all identified cycles, using the Palmgren-Miner cumulative damage theory.

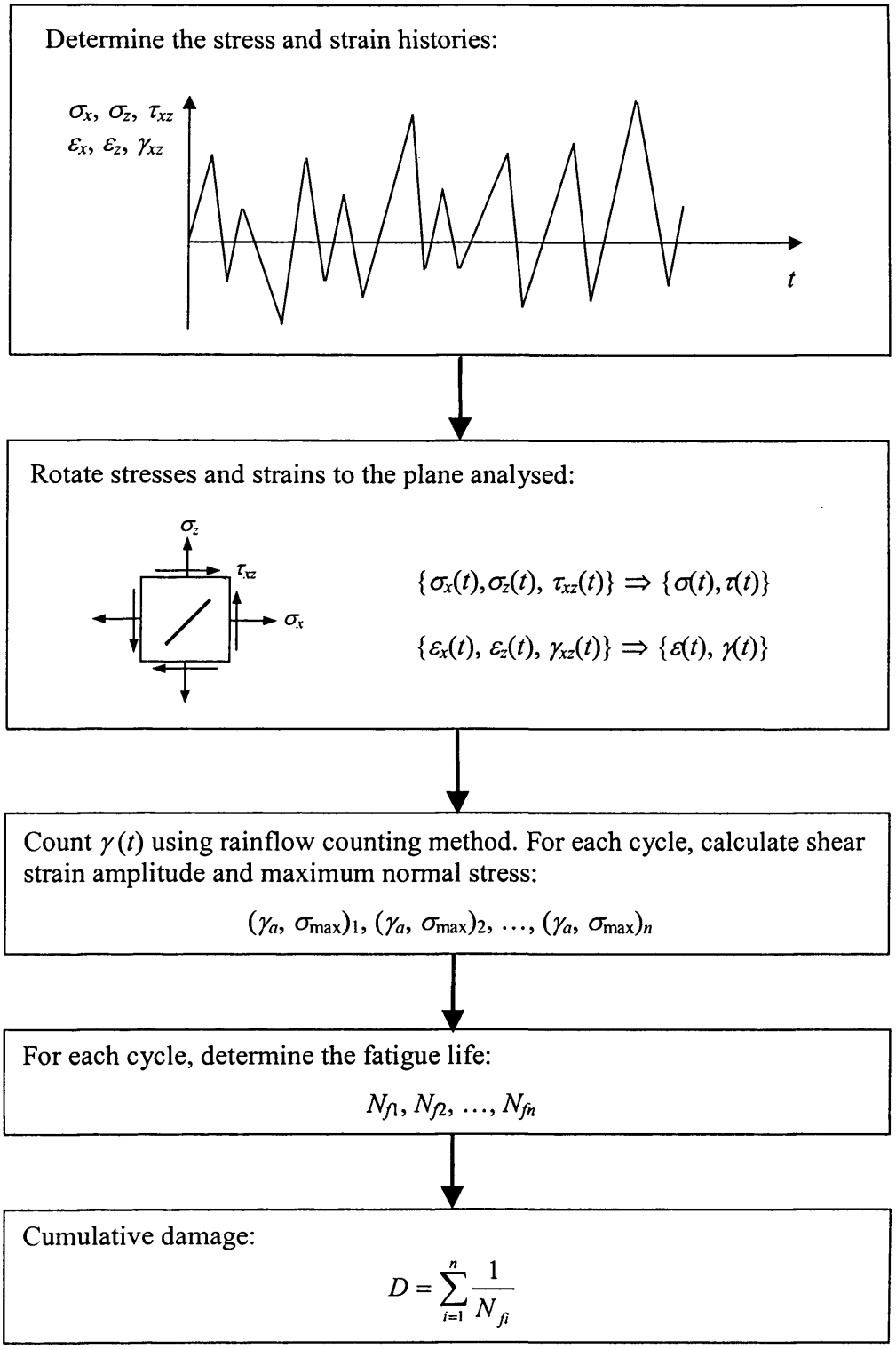


Figure 7.2: Procedure of fatigue analysis for EHL line contact.

7.2.3 Cyclic Stress and Strain Fields

To apply the FS fatigue model, it is necessary to determine the subsurface stress and strain fields developed under the contact in the region of interest throughout the load cycle. The stresses are evaluated in Chapter 4 assuming the stress field remains elastic (or experiences elastic shakedown under repeated contact if the contact pressure doesn't exceed the shakedown limit). Of course, plastic deformation is inevitable at the asperity and microstructural scale even under the lightest loads, especially during early cycles of load. This is also the case in most fatigue tests to determine the fatigue properties for materials. Therefore, the effect of plasticity at the asperity scale is assumed to have already been taken into account by fatigue parameters.

Since the stresses remain elastic at the macroscopic scale, the Hooke's law is used to evaluate the strains. For plane strain, they can be written as

$$\varepsilon_x = \frac{1}{E}\left\{(1-\nu^2)\sigma_x - \nu(1+\nu)\sigma_z\right\} \tag{7.4a}$$

$$\varepsilon_z = \frac{1}{E}\left\{(1-\nu^2)\sigma_z - \nu(1+\nu)\sigma_x\right\} \tag{7.4b}$$

$$\gamma_{xz} = \frac{\tau_{xz}}{G} = \frac{2(1+\nu)}{E}\tau_{xz} \tag{7.4c}$$

Since the critical plane is not known before the analysis, all the candidate planes have to be examined in order to identify the critical plane. The effect of the applied loading on a plane is determined by a tensor rotation of the stresses and strains to this plane. Since this is a plane strain problem, only planes within x-z plane need to be examined. The normal and shear stress on a plane is thus obtained by (Timoshenko & Goodier 1970)

$$\sigma = \sigma_x \cos^2\theta + \sigma_z \sin^2\theta + 2\tau_{xz}\sin\theta\cos\theta \tag{7.5a}$$

$$\tau = \tau_{xz}(\cos^2\theta - \sin^2\theta) + (\sigma_z - \sigma_x)\sin\theta\cos\theta \tag{7.5b}$$

where θ is the angle between the normal of the plane and the x-axis. In the same way, the normal and shear strain on the same plane is determined by

$$\varepsilon = \varepsilon_x \cos^2\theta + \varepsilon_z \sin^2\theta + \gamma_{xz} \sin\theta\cos\theta \tag{7.6a}$$

$$\frac{\gamma}{2} = \frac{\gamma_{xz}}{2}(\cos^2\theta - \sin^2\theta) + (\varepsilon_z - \varepsilon_x)\sin\theta\cos\theta \tag{7.6b}$$

7.2.4 Cycle Counting and Damage Cumulation

The rainflow counting method developed by Amzallag et al. (1994) is employed on the shear strain history $\gamma(t)$ to extract loading cycles. Application of the rainflow counting method requires a preliminary treatment of the shear strain history to extract extrema. A simple method is used to determine if a point in the history is a peak or valley point. For example, three successive points denoted by i-1, i, i+1 have shear strain values γ_{i-1}, γ_i and γ_{i+1}, respectively. If $(\gamma_i - \gamma_{i-1})(\gamma_{i+1} - \gamma_i) \geq 0$, then the ith point is not an extreme and will be removed from the strain history. Otherwise, it is an extreme and kept. In such a way, the strain history contains only extrema.

According to Amzallag et al. (1994), there are two possible types of cycle that can occur for four successive extrema in the load history as illustrated in Figure 7.3. The strain history (the left hand side) can be transformed in the form of stress-strain relationship (the right hand side) to see if a cycle is formed. The four extrema are labeled by 1, 2, 3 and 4, and the corresponding strain values are γ_1, γ_2, γ_3 and γ_4. Strain ranges are determined by $\Delta\gamma_1 = |\gamma_2 - \gamma_1|$, $\Delta\gamma_2 = |\gamma_3 - \gamma_2|$, and $\Delta\gamma_3 = |\gamma_4 - \gamma_3|$. If $\Delta\gamma_2 \leq \Delta\gamma_1$ and $\Delta\gamma_2 \leq \Delta\gamma_3$, then the cycle represented by its extreme values γ_2 and γ_3 is counted.

Once a cycle is counted by the rainflow counting method, the shear strain amplitude and the maximum normal stress can be determined for that cycle. If the cycle is represented by its extreme values, $\gamma(t_1)$ and $\gamma(t_2)$, at two instants t_1 and t_2, then the shear strain amplitude is determined by

$$\gamma_a = \frac{1}{2}|\gamma(t_1) - \gamma(t_2)| \tag{7.7}$$

and the maximum normal stress during that cycle is

$$\sigma_{\max} = \max_{t_1 \leq t \leq t_2}\{\sigma(t)\} \tag{7.8}$$

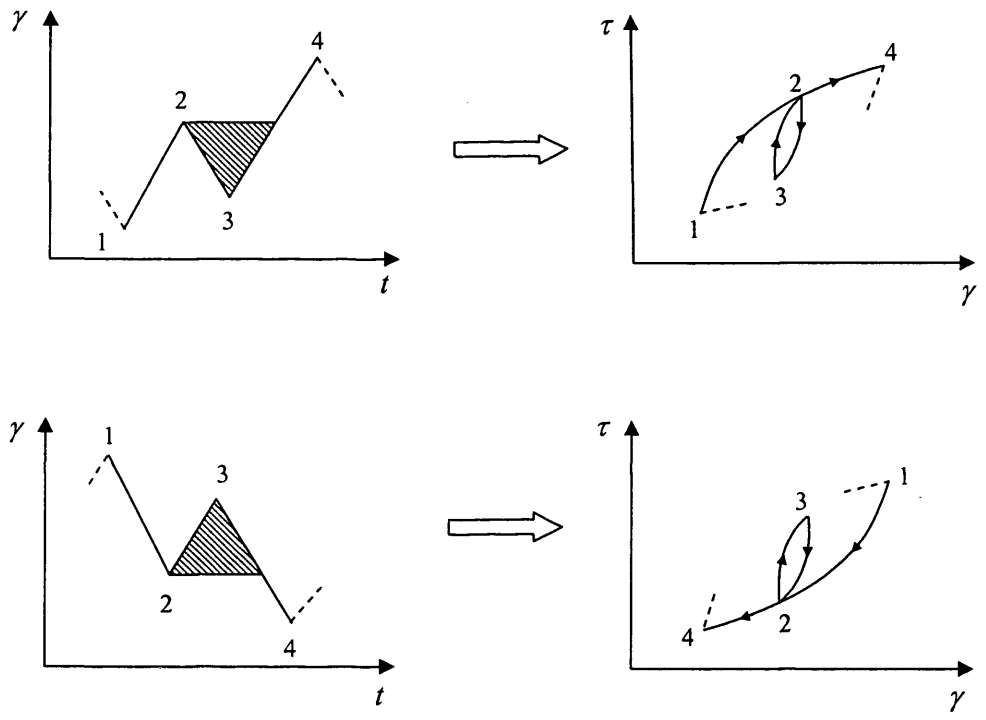


Figure 7.3: Principle of cycle extraction.

Thus the fatigue life N_f, or fatigue damage, $1/N_f$, associated with each cycle of shear strain acting on the plane being analysed is determined using Equation (7.2). Iterative techniques must be employed to solve Equation (7.2). In this thesis, a simple bisection method is applied assuming N_f lies between 1 and 10^{20}. At each step of the bisection algorithm, the right-hand side is evaluated at the logarithmic midpoint on the interval $[N_1, N_2]$ which contains the solution, i.e. $\ln N_f = (\ln N_1 + \ln N_2)/2$. N_f is regarded as the solution if the relative difference between the left-hand side (*LHS*) and right-hand side (*RHS*) of Equation (7.2) is less than 10^{-4}, i.e. $|LHS - RHS| < LHS \times 10^{-4}$.

The total damage is obtained using the Palmgren-Miner cumulative damage rule so that the damage, D, is

$$D = \sum_{i=1}^{n} \frac{1}{N_{fi}} \tag{7.9}$$

where n is the number of cycles and N_{fi} is the fatigue life for the ith cycle. The critical plane is identified as the plane with the maximum damage, i.e.

$$D_{\max} = \max_{0^\circ \le \theta \le 180^\circ} \{D(\theta)\} \tag{7.10}$$

A value of $D_{\max}$ equal to unity then corresponds to fatigue failure. In the following, the calculated damage D refers to the damage on the critical plane.

7.3 Damage Distribution by Weibull Analysis

To account for the effect of damage randomness due to randomness of rough surface asperities, a statistical approach can be used to give a damage distribution for a particular depth beneath the surface in the materials based on the calculated damage values for all the material points at that depth. The damage probability density function is naturally skewed towards low damage and the Weibull distribution is thus an appropriate stochastic model to use for this purpose. The damage distribution expressed in the 3-parameter Weibull cumulative density function (CDF) has the form (Meeker & Escobar 1998)

$$F(D) = 1 - \exp\left[-\left(\frac{D-\mu}{\eta}\right)^{\beta}\right], \quad 0 \le \mu \le D, \quad \eta, \beta > 0 \tag{7.11}$$

where η is the scale parameter, β is the shape parameter and μ is the location parameter. If μ = 0, the 3-parameter Weibull distribution becomes the 2-parameter Weibull distribution. These two distributions are identical except that the 3-parameter Weibull distribution is shifted by a distance of μ.

To fit the calculated damage data to a Weibull distribution, the method of median rank regression is applied in this thesis to estimate all three parameters. Equation (7.11) can be transformed to a simple form for a line:

$$Y = mX + b \tag{7.12}$$

where

$$Y = \ln\left[\ln\left(\frac{1}{1-F(D)}\right)\right] \tag{7.13a}$$

$$X = \ln(D-\mu) \tag{7.13b}$$

$$m = \beta \tag{7.13c}$$

$$b = -\beta \ln \eta \tag{7.13d}$$

If n is the sample size, rearrange the n calculated damage values in an ascending order, i.e. $D_1 \le D_2 \le \ldots \le D_i \le \ldots \le D_n$, where D_i denotes the ith smallest damage. $F(D_i)$ can be approximated by Benard's median rank (Abernethy 1996):

$$F(D_i) = \frac{i-0.3}{n+0.4} \tag{7.14}$$

If the location parameter μ is known, then a corresponding ordered set of (X_i, Y_i) can be obtained from Equations (7.13a) and (7.13b). The two unknowns m and b can be evaluated using a traditional least square method (LSM) by solving the two simultaneous equations as follows (Ayyub & McCuen 1996):

$$m = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2} \tag{7.15a}$$

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - m\sum_{i=1}^{n} X_i\right) \tag{7.15b}$$

Then, the shape and scale parameters are calculated as follows:

$$\beta = m \tag{7.16a}$$

$$\eta = \exp\left[-\left(\frac{b}{\beta}\right)\right] \tag{7.16b}$$

The goodness of the fit can be assessed by the correlation coefficient (R), which measures the degree to which the measured and predicted values agree. The square of the correlation coefficient (R^2), which is called the coefficient of determination, represents the fraction of the variance of the output explained by the regression model and is given by (Ayyub & McCuen 1996):

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{7.17}$$

where $\hat{Y}_i$ denotes the estimate of the observations on Y_i obtained from the regression model and $\bar{Y}$ is the mean of Y_i. The closer R^2 is to unity, the better is the model performance.

The complexity of the above method to estimate the Weibull parameters arises from the fact that the location parameter μ is also unknown. In order to find μ that corresponds to the best fit, the least square method above is used to fit the data for all values of μ in the interval [0, D_1], where D_1 is the smallest calculated damage and is selected as the upper limit to make Equation (7.13b) applicable. The correlation coefficient R is used as the criterion to determine the best fit value of μ. The value of μ that corresponds to the maximum R^2 is taken as the estimated location parameter. This is obtained by a trial and error approach.

7.4 Fatigue Analysis Results and Discussion

In this section, the application of the shear strain based critical plane approach for damage calculations is demonstrated in the line contact situation. The parameters used to specify the EHL problem are the same as given by Table 6.2 in Chapter 6 unless otherwise specified. The material used for the damage modelling is SAE4340 steel (BS970:En24) whose material properties can be found in Zahavi and Torbilo (1996) and are listed in Table 7.1. The shear fatigue strength and shear fatigue ductility coefficient in Equation (7.2) can be estimated by (Dowling 1998)

$$\tau'_f \approx \frac{\sigma'_f}{\sqrt{3}}, \qquad \gamma'_f \approx \sqrt{3}\varepsilon'_f \tag{7.18}$$

As described in Chapter 6, sections of the contacting materials of dimension $2a \times a$ are selected for fatigue analysis, where a is the Hertzian semi-contact width. The selected material is located at $-4 \leq x/a \leq -2$ at the start of the analysis and then passes through the contact area during transient analysis, moving to $2 \leq x/a \leq 4$. The calculated damage is the damage in such a traverse of the contact area.

As in Chapter 6, the critical plane is defined by θ, which is the angle between the normal of the critical plane and the x-axis. The fatigue damage is computed for $0° \leq \theta \leq 180°$ with an interval of 1° to search for the critical plane.

The profiles of rough surfaces used for the analyses are shown in Figure 2.2 of Chapter 2 where they are denoted as profiles (A), (B) and (C).

Table 7.1: 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	ν_1, ν_2	0.3

7.4.1 Damage Distributions

Calculation of damage as described in Section 7.2 was carried out for the contact between two rough surfaces each with profile (C) (denoted as 'C-C'), and with slide roll ratio ξ = 0.5, entrainment velocity $\bar{u}$ = 25m/s and viscosity η_0 = 0.001Pa·s. As in previous chapters, the upper surface is the faster moving surface and the lower surface is the slower moving surface. Figure 7.4 shows contours of damage obtained for both the faster and the slower moving rough surface in passing through the contact area once. The length of profile considered is $2a$ and its shape, which is the same as that in Figure 6.3 of Chapter 6, is also included in the figure. The contours are chosen to vary logarithmically in value as the difference in D between the areas subject to most and least calculated damage is up to five orders of magnitude. It can be seen that the accumulated damage calculated is concentrated

near the surface of particular asperity features and the islands of high D values calculated are separated by areas that are not subject to anything like the same level of damage. This observation is in a good agreement with that presented in Chapter 6.

The damage value depends on the shape of the asperity, but also on its height relative to the local surface height envelope. Prominent asperities are subject to greater damage levels than those of similar shape that are at a lower level than their neighbours. Table 7.2 lists damage values in decreasing order for ten asperities shown in Figure 7.4 with letters A to J for both the faster and slower surfaces. These asperities selected are the same as those in Figure 6.3 in Chapter 6. It can be seen that damage values at these asperities for the two surfaces can be very different because the "collisions" experienced by each asperity differ according to the sliding speed.

Table 7.2: Damage values of ten asperities for both the faster and the slower surfaces.

Index	Faster Surface		Slower Surface	
	Damage, D	Location	Damage, D	Location
1	1.235×10^{-4}	A	1.043×10^{-4}	J
2	7.649×10^{-5}	C	8.493×10^{-5}	G
3	3.864×10^{-5}	B	4.854×10^{-5}	H
4	1.400×10^{-5}	H	3.845×10^{-5}	D
5	9.620×10^{-6}	D	3.287×10^{-5}	I
6	4.834×10^{-6}	I	3.230×10^{-5}	C
7	7.783×10^{-7}	F	2.887×10^{-5}	E
8	2.692×10^{-7}	E	1.618×10^{-5}	A
9	2.269×10^{-7}	G	9.226×10^{-6}	F
10	7.791×10^{-8}	J	6.119×10^{-6}	B

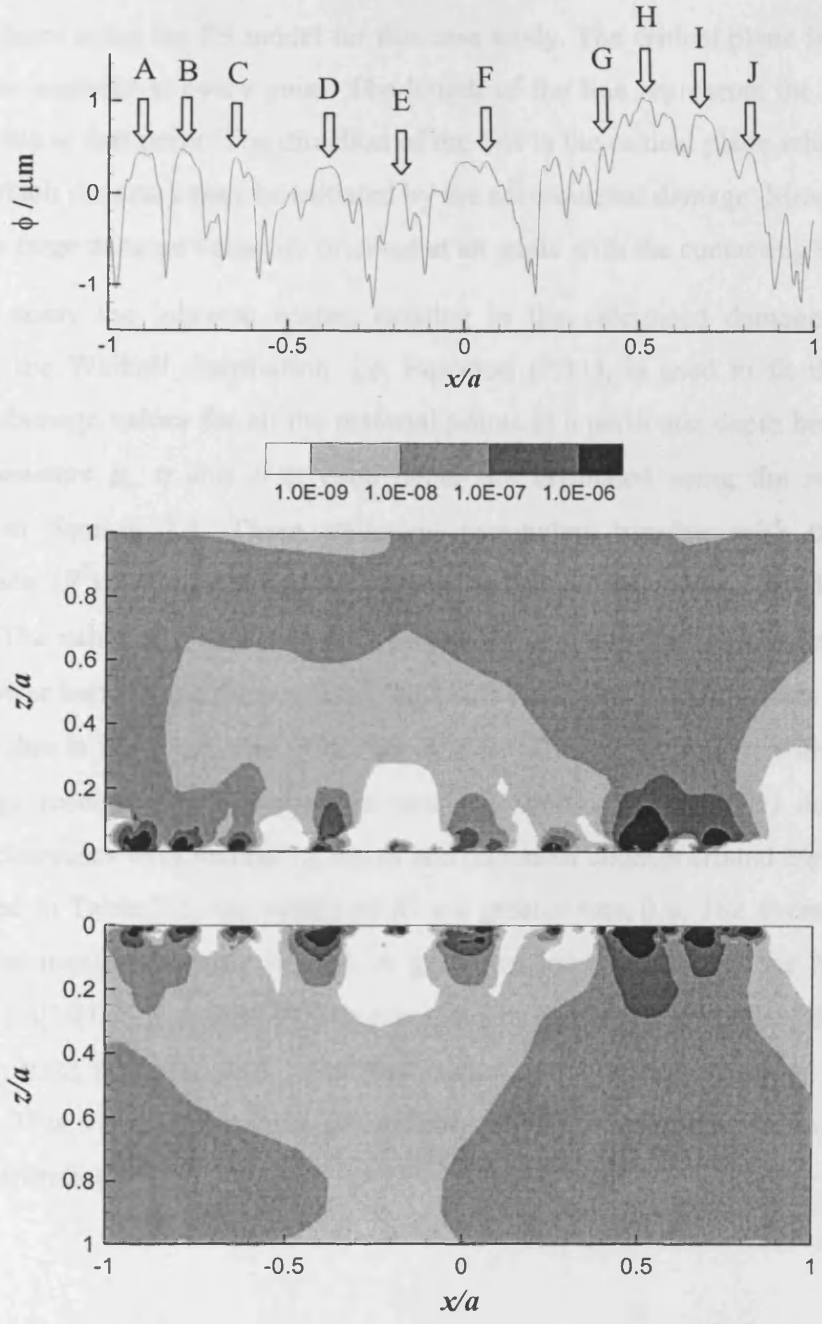


Figure 7.4: Contours of subsurface damage calculated for both faster (upper figure) and slower (lower figure) moving surfaces during transit of the contact area with ξ = 0.5 and η_0 = 0.001Pa·s. The rough profile is shown in the figure above the contour plots.

Figure 7.5 shows distributions of the critical planes determined for both the faster and slower surfaces using the FS model for this case study. The critical plane is represented by a straight line segment at every point. The length of the line represents the magnitude of the damage value at that point. The direction of the line is the critical plane which is the possible plane on which the crack may be initiated by the accumulated damage. Most critical planes at points with large damage value are oriented at an angle with the contacting surfaces.

To study the inherent scatter existing in the calculated damages due to surface roughness, the Weibull distribution, i.e. Equation (7.11), is used to fit the distribution of calculated damage values for all the material points at a particular depth beneath the surface. Three parameters μ, η and β at each depth are estimated using the regression method described in Section 7.3. These statistical parameters together with the coefficient of determination (R^2) for the case of the slower surface at the depths considered are listed in Table 7.3. The value of β increases with increasing depth indicating that damage distribution has a narrower band in the deeper layers and thus corresponding variations of damage in the x-direction due to the roughness effect are smaller. The variation of η with the depth shows that damage reaches the largest value near the surface (z/a = 0.01) due to micro-EHL pressures, decreases with increasing depth and increases slightly around z/a = 0.5. For all the depths listed in Table 7.3, the values of R^2 are greater than 0.9. The average value of R^2 is 0.96 and the maximum value is 0.99. A graphical interpretation of the linear relationship between $\ln\{\ln[1/(1-F)]\}$ and $\ln(D-\mu)$ is presented in Figure 7.6 for typical depth levels of z/a = 0.0, 0.01, 0.02, 0.05, 0.1, 0.5. The magnitudes of the corresponding R^2 are also given in the figure. This visually confirms the validity of the 3-parameter Weibull model for the damage distribution.

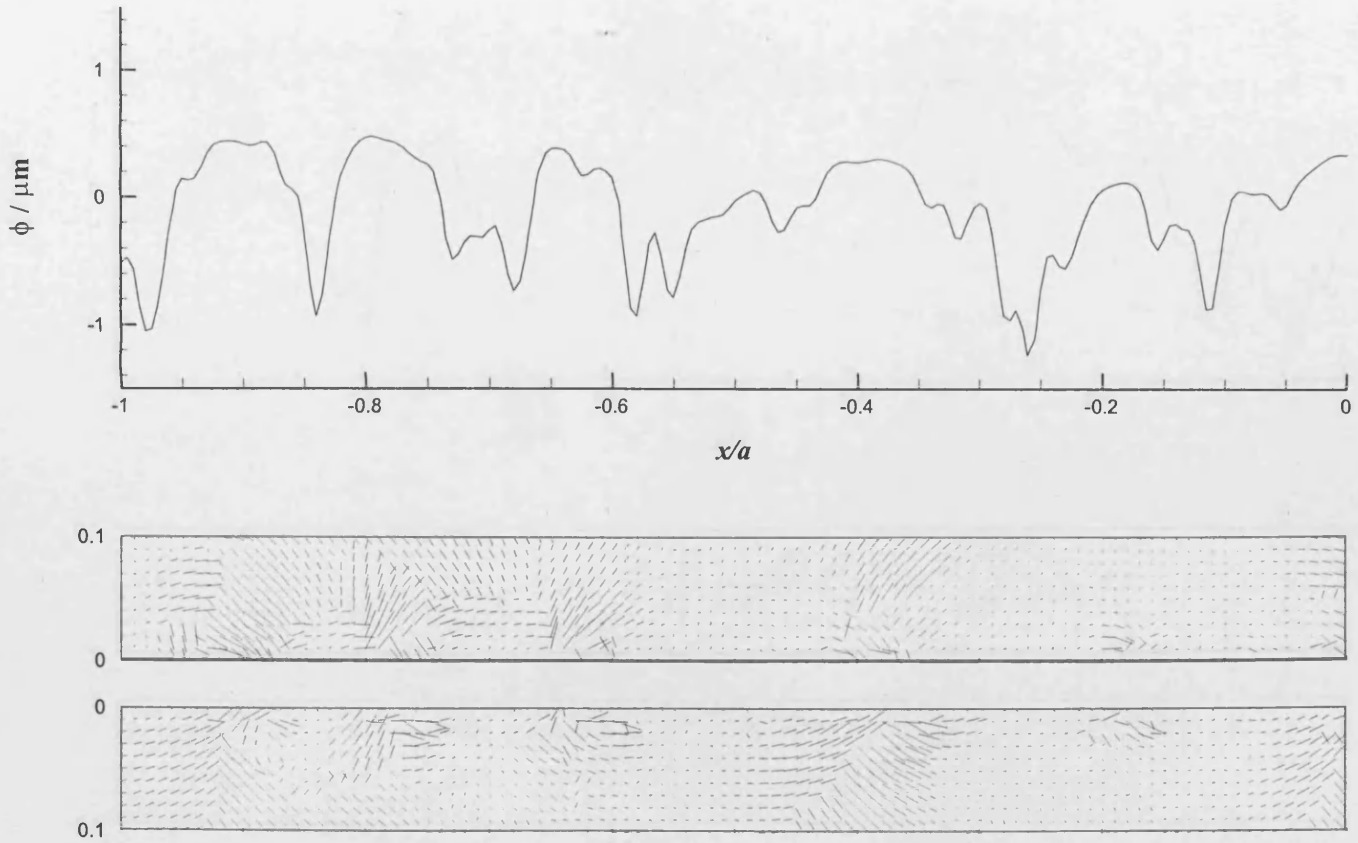


Figure 7.5(a): Distributions of critical planes determined for both faster (upper figure) and slower (lower figure) moving surfaces in the near surface layers ($0 \leq z/a \leq 0.1$) for the section of material of $-1 \leq x/a \leq 0$, whose rough profile is shown in the figure above the vector plots.

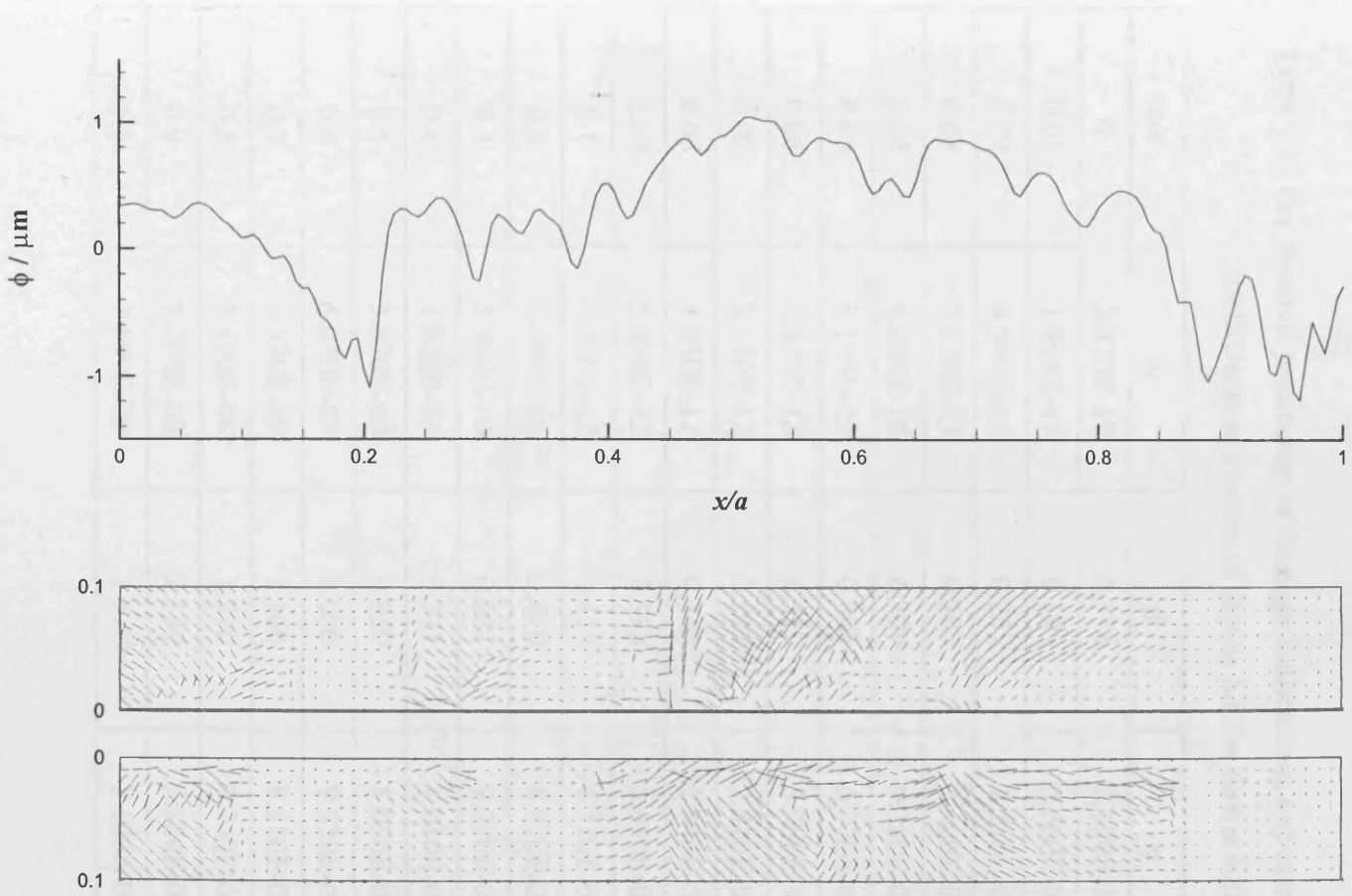
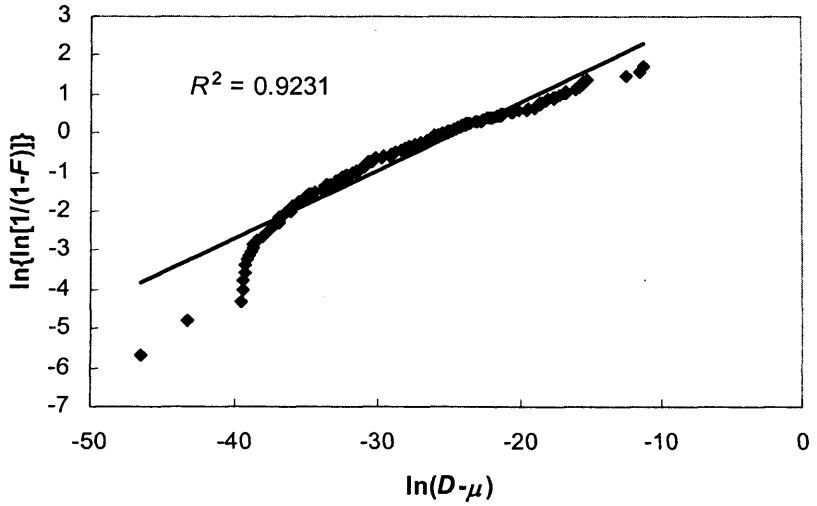


Figure 7.5(b): Distributions of critical planes determined for both faster (upper figure) and slower (lower figure) moving surfaces in the near surface layers ($0 \leq z/a \leq 0.1$) for the section of material of $0 \leq x/a \leq 1$, whose rough profile is shown in the figure above the vector plots.

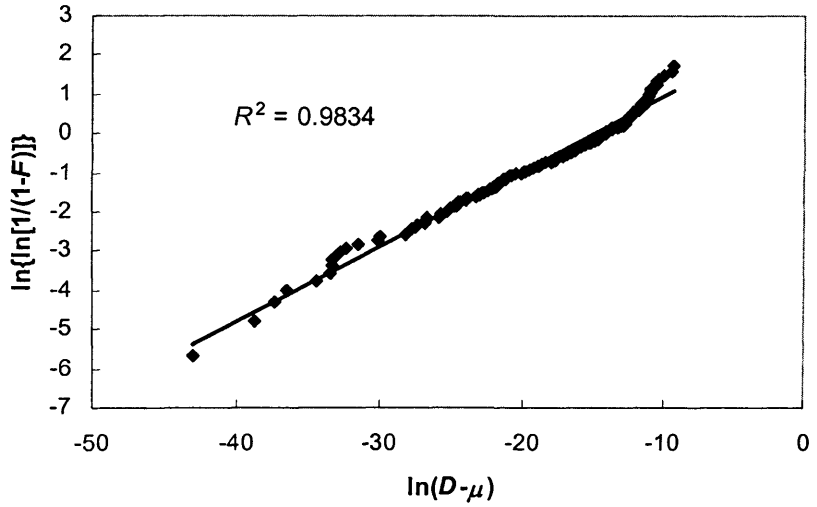
Table 7.3: The Weibull parameters of damage distributions and corresponding correlation coefficients at a series of depths for the slower surface.

z/a	μ	β	η	R^2
0	5.870E-19	0.174	2.190E-11	0.92
0.01	1.860E-17	0.191	3.360E-07	0.98
0.02	4.590E-15	0.257	2.310E-07	0.99
0.03	7.570E-13	0.295	1.400E-07	0.99
0.04	4.060E-12	0.309	9.780E-08	0.95
0.05	1.120E-11	0.303	5.770E-08	0.95
0.06	1.820E-12	0.312	3.900E-08	0.92
0.07	2.170E-12	0.319	2.810E-08	0.94
0.08	3.950E-12	0.345	2.340E-08	0.93
0.09	5.930E-12	0.386	2.050E-08	0.96
0.1	7.220E-12	0.416	1.770E-08	0.97
0.2	2.040E-10	0.619	8.970E-09	0.96
0.3	5.960E-10	0.681	9.700E-09	0.94
0.4	1.920E-09	0.771	1.010E-08	0.97
0.5	3.870E-09	0.959	1.070E-08	0.98
0.6	6.070E-09	1.000	9.920E-09	0.98
0.7	7.570E-09	1.160	8.880E-09	0.98
0.8	7.490E-09	1.770	7.740E-09	0.97
0.9	7.280E-09	2.150	5.920E-09	0.96
1.0	7.690E-09	1.470	3.180E-09	0.97

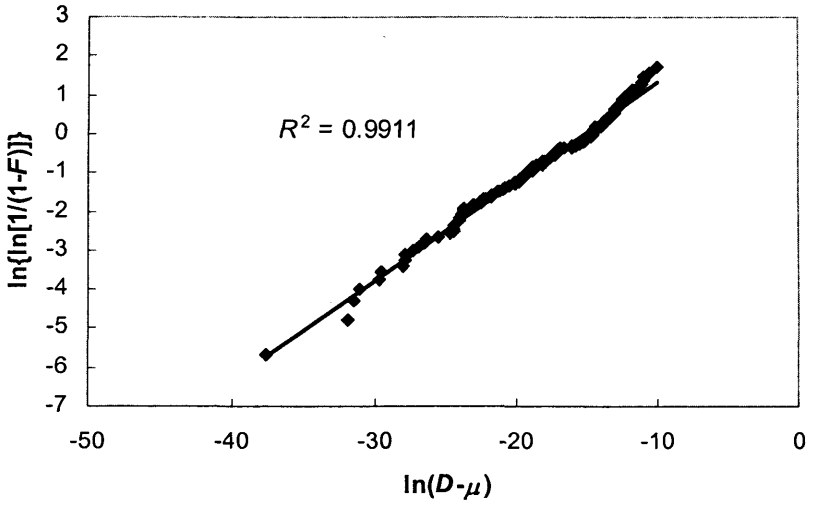
(1) $z/a = 0$


R^2 = 0.9231
ln{ln[1/(1-F)]}
ln(D-μ)
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-50
-40
-30
-20
-10
0

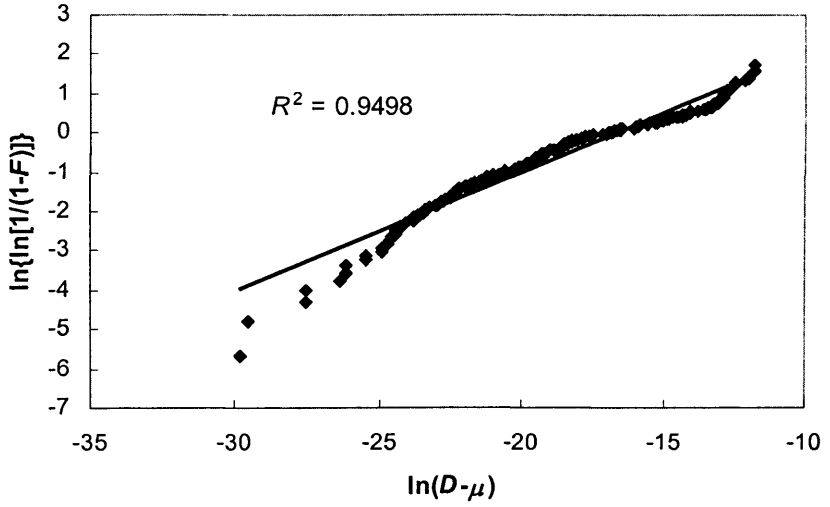
(2) $z/a = 0.01$


R^2 = 0.9834
ln{ln[1/(1-F)]}
ln(D-μ)
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-50
-40
-30
-20
-10
0

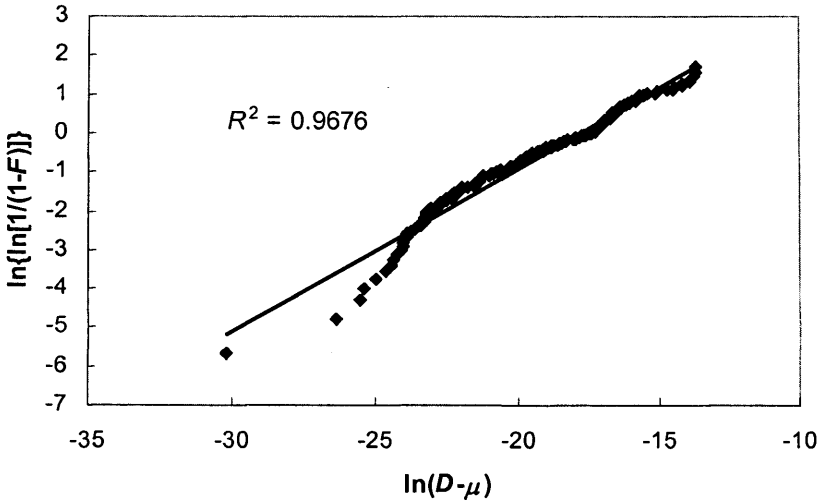
(3) z/a = 0.02



(4) z/a = 0.05



(5) z/a = 0.1



(6) z/a = 0.5

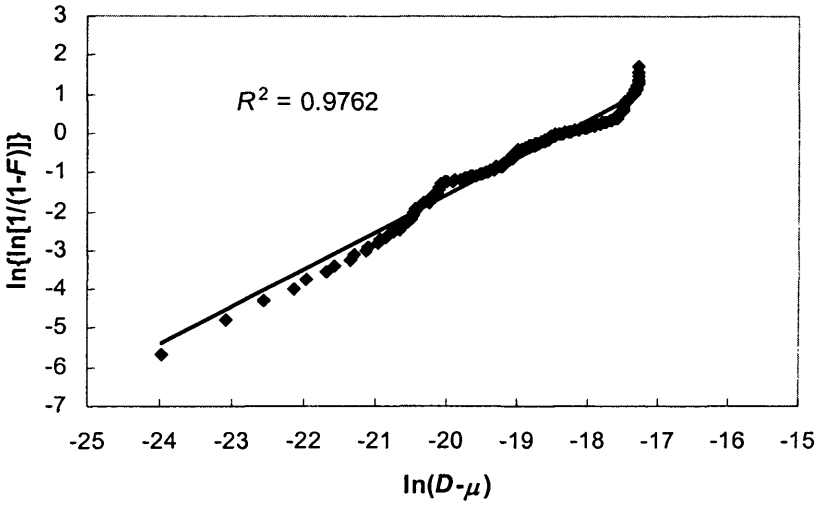


Figure 7.6: Graphical representation of the damage distribution calculated using the LSM at depth levels of z/a = 0, 0.01, 0.02, 0.05, 0.1 and 0.5.

Figure 7.7 shows the cumulative Weibull distributions of damage in terms of 1-F for the slower surface at a series of depth levels. Three extra depth levels, i.e. z/a =0.001, 0.002, and 0.004, are given in this figure to examine the variations of damage in the very near surface layers. The curves are plotted so as to give the probability that the damage level is greater than the abscissa value. Thus for the z/a = 0.01 case, for example, about 20% of material points at that depth experience calculated damage greater than $D = 3\times10^{-6}$ during a single pass through the contact area. The curves show that calculated damage is greatest at the z/a = 0.01 level as might be expected from examination of Figure 7.4. At the greater depth of z/a = 0.5 the level of damage has reduced by three or four orders of magnitude in comparison.

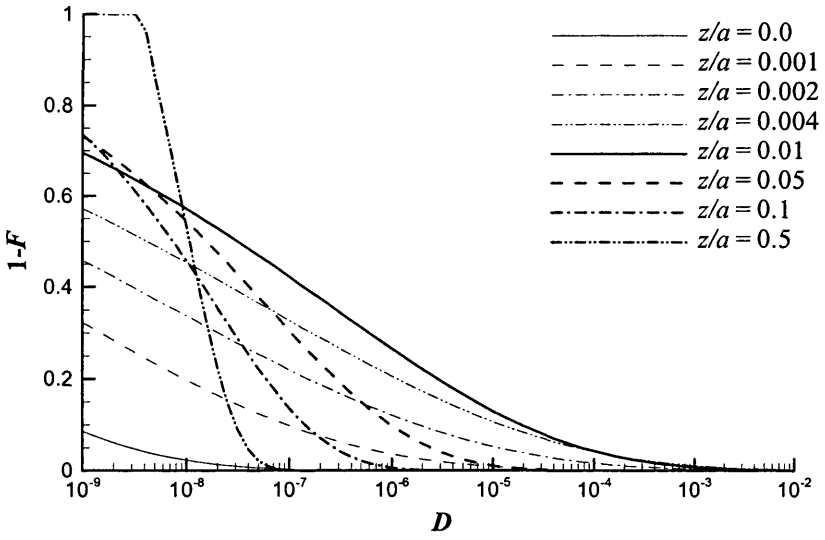


Figure 7.7: Cumulative damage distributions for slower moving surface of the case with ξ = 0.5 and η_0 = 0.001Pa·s at a series of depths beneath the surface.

In two parameter Weibull life data analysis, the parameter η is approximately equal to the mean-time-to-failure (Abernethy 1996). Similarly, it is helpful to define a characteristic damage at a given depth level as

$$D_c = \eta + \mu \tag{7.19}$$

which gives a probability of $1 - F = 1/e = 0.368$ for all damage distributions, regardless of the value of β. Figure 7.8 shows the variations of D_c versus the depth for both faster and slower surfaces. The calculated damage variation for the case with smooth surfaces is also included in the figure. It can be seen that for the rough surfaces D_c increases drastically starting from the surface and reaches the highest value in the near surface layer due to micro-EHL pressures. Then D_c drops rapidly as depth increases, and increases again slightly in the maximum shear stress zone corresponding to the macro-EHL pressure. Comparing the curves for the faster and slower surfaces, it is clear that the damage for the slower surface is greater than that for the faster surface in the near surface layer. This is because asperities of the slower moving surface are subject to higher numbers of stress cycles in traversing the contact than are those of the faster moving surface, as described in Chapter 4. This conclusion is in agreement with the unexplained observation seen in micropitting experiments with gears (Shaw 2003). That is, micropitting takes place preferentially in the root of the pinion tooth (slower surface) without occurring on the meshing tip of the wheel tooth (faster surface). Therefore, the slower moving surface is examined in the following analyses. At the greater depth where the roughness effect is no longer active, the two surfaces have the same damage levels which become close to the damage level for the case with smooth surfaces.

To investigate the sensitivity of damage distributions with respect to sections of materials selected, three different sections of materials in the slower surface are selected to perform fatigue analyses. Profiles of these sections of materials are shown in Figure 7.9 and they are denoted (1), (2) and (3). After the calculated damage values are obtained and fitted to Weibull distributions, variations of D_c along the depth are plotted in Figure 7.10. Comparing the curves in Figure 7.10, the difference between them is minor and the damage level is limited to the same order of magnitude. From a statistical view point, distributional robustness is satisfied.

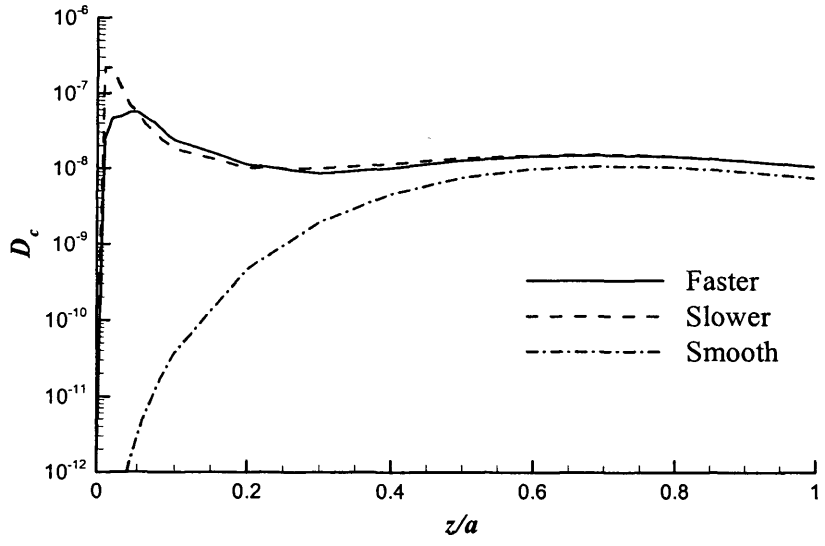


Figure 7.8: Variations of the characteristic damage D_c versus the depth for the faster and slower moving surfaces for the case with ξ = 0.5 and η_0 = 0.001Pa·s. The variation of the calculated damage for the smooth surfaces is also included for comparison.

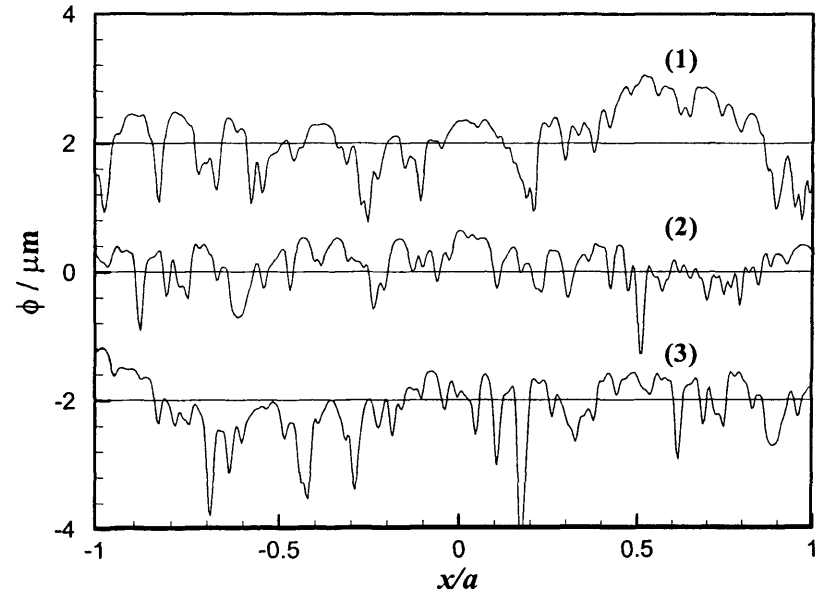


Figure 7.9: Profiles of three sections of materials selected in the slower moving surface for the case with ξ = 0.5 and η_0 = 0.001Pa·s to perform sensitivity analysis.

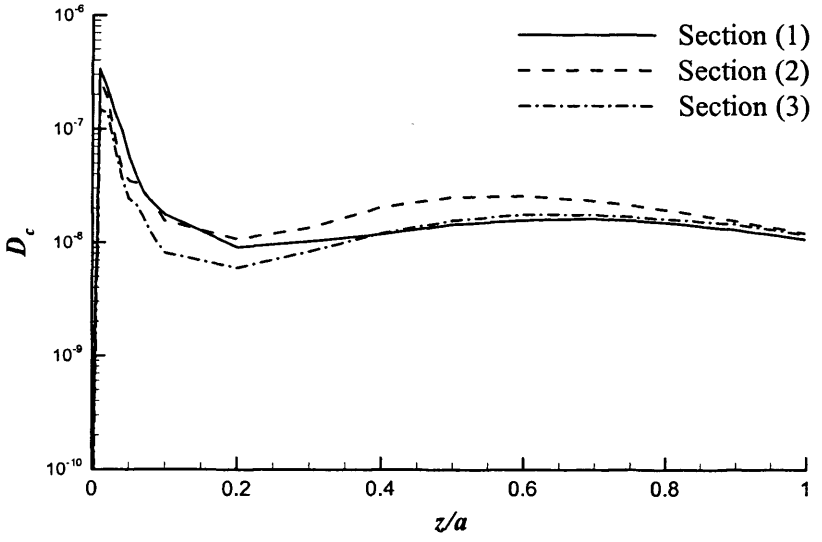


Figure 7.10: Variations of D_c versus the depth for three sections of materials shown in Figure 7.9.

7.4.2 Effect of Slide Roll Ratio

The effect of slide roll ratio on fatigue damage is studied in this section for the contact of C-C with a range of ξ values between 0 and 1.5 and $\xi = 0$ means pure rolling. The viscosity is kept constant at η_0 = 0.001Pa·s and the entrainment velocity is $\bar{u}$ = 25m/s. Figure 7.11 compares the damage contours obtained with these values of ξ for the same section of rough profile illustrated in Figure 7.4. In general increased ξ values lead to higher calculated values of D and the location of the damage hotspots is unchanged. The volume subject to a damage level of 10^{-6} surrounding the peak damage locations is seen to increase. This observation is in agreement of the results presented in Chapter 6.

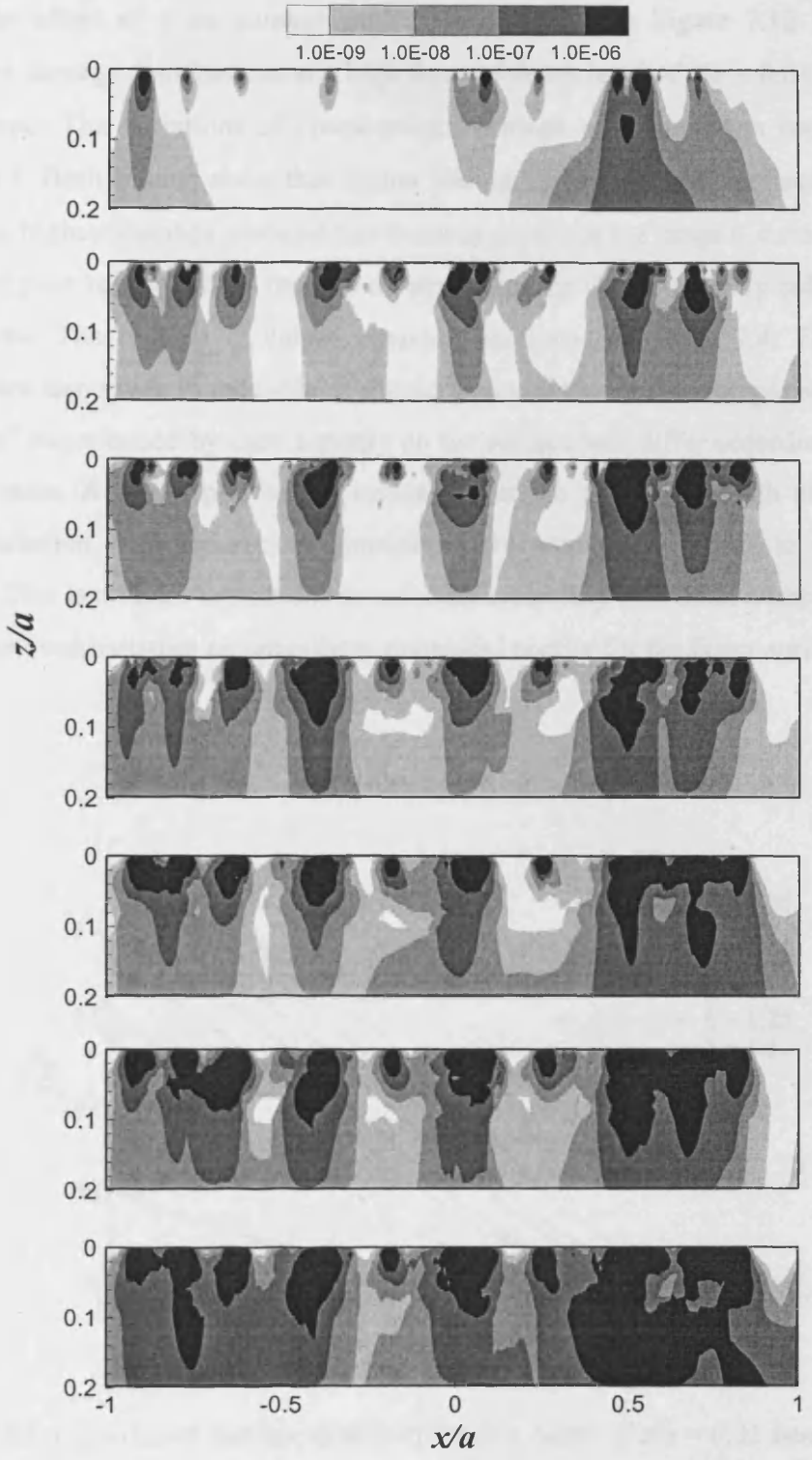


Figure 7.11: Contours of subsurface damage calculated for the section of the slower moving surface illustrated in Figure 7.4 during its transit of the contact area with η_0 = 0.001Pa·s and ξ = 0 (upper figure), 0.25, 0.5, 0.75, 1.0, 1.25, 1.5 (lower figure).

The effect of ξ on damage probabilities is seen in Figure 7.12 which shows the cumulative damage distributions at a high damage depth level of $z/a = 0.01$ for each value of ξ considered. The variations of characteristic damage with the depth levels are shown in Figure 7.13. Both figures show that higher sliding leads to greater damage probabilities. In general the highest damage probabilities occur at depths in the range $0 < z/a \leq 0.05$ except for the case of pure rolling where there is no stress cycling due to asperity collisions within the contact area. The highest D values obtained are listed in Table 7.4. The corresponding locations are also given in the table. It should be noted that in this comparison, the number of "collisions" experienced by each asperity on the surface will differ according to the specified slide roll ratio. Also the part of the opposing surface passing through the contact area in direct interaction with the section considered will change according to the sliding speed specified. This last factor introduces an inherent variability into these comparisons that could only be removed by using an unrealistic sinusoidal profile for the faster surface.

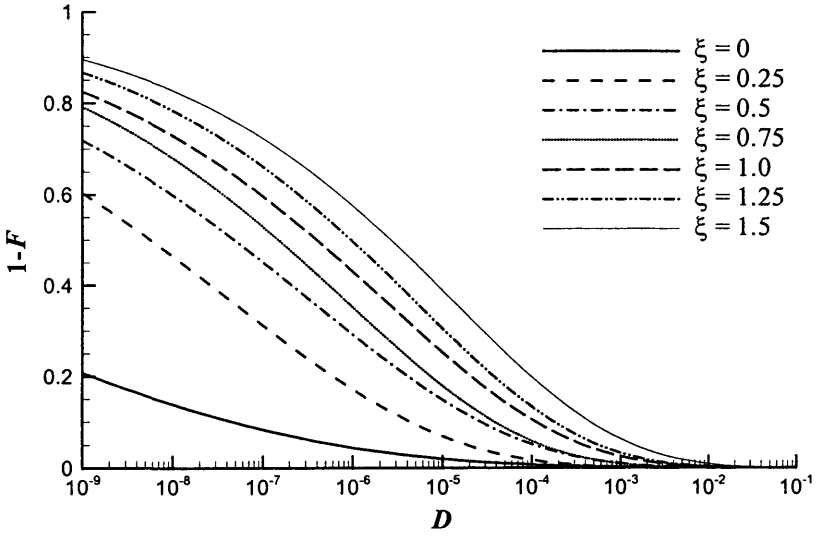


Figure 7.12: Cumulative damage distributions at a depth of $z/a = 0.01$ beneath the surface for a series of ξ values.

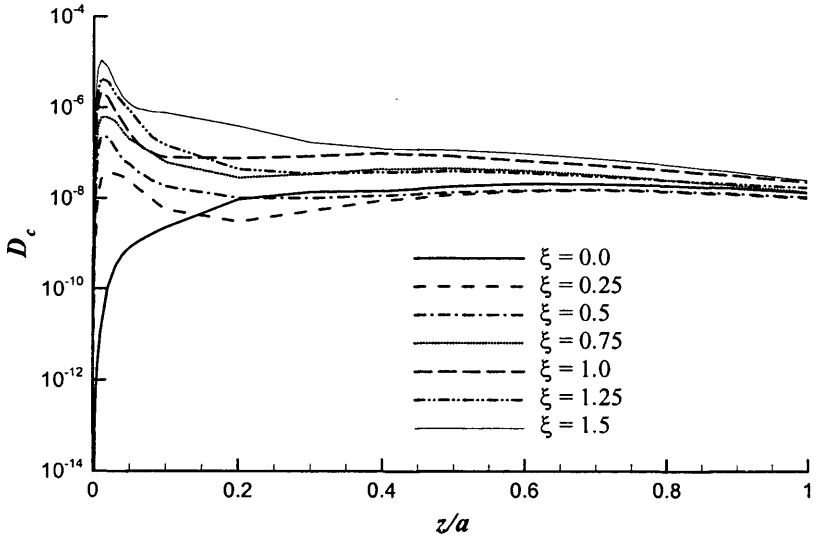


Figure 7.13: Variations of D_c versus the depth for a series of ξ values.

Table 7.4: Maximum calculated damage for a range of ξ.

Slide roll ratio, ξ	Maximum damage, D_{max}	Location
0	7.612×10^{-5}	H
0.25	5.474×10^{-5}	D
0.5	1.043×10^{-4}	J
0.75	1.178×10^{-4}	D
1.0	1.053×10^{-3}	A
1.25	2.840×10^{-4}	D
1.5	9.347×10^{-3}	A

7.4.3 Effect of Viscosity

The effect of viscosity on fatigue damage is considered for the contact of C-C with ξ = 0.5, $\bar{u}$ = 25m/s and a series of η_0 values, i.e. η_0 = 0.001, 0.002, 0.004, 0.008, 0.016 Pa·s. The corresponding dry contact analysis (denoted as η_0 = 0) is also included for comparison. Figure 7.14 compares the damage contours obtained with these values of η_0. The cumulative damage probabilities are shown in Figure 7.15 at a depth of z/a = 0.01 for each value of η_0 considered. The corresponding distributions of characteristic damage D_c versus the depth are shown in Figure 7.16. It can be seen that the fatigue damage decreases with increased η_0. With lower viscosities, the highest damage occurs near the surface. With greater viscosities, for example η_0 = 0.008 and 0.016 Pa·s, the damage levels near the surface are reduced drastically and the highest damage occurs in the deeper layer. This observation is slightly different to the results presented in Chapter 6, where the Ioannides-Harris model gives the highest failure probability near the surface for the case of η_0 = 0.008 Pa·s. The difference can be explained by the fact that D_c can be regarded as a mean value of damage, whereas the Ioannides-Harris model is very sensitive to the highest stress value. From the contour for the case of η_0 = 0.008 Pa·s in Figure 7.14, it can be seen that the highest damage still occurs near the surface.

The most severe situation by far is that of dry contact, in which the characteristic damage D_c is greater than that of the case of η_0 = 0.001 Pa·s by three orders of magnitude at the depth of z/a = 0.01. At the surface z/a = 0, the difference can be up to six orders of magnitude due to the drastic increase of the surface friction in dry contact (An assumed coefficient of friction of 0.1 was used in this work).

The calculated highest D values are listed in Table 7.5. The Λ ratio values for the cases with lubrication are also given in this table for comparison. The Λ ratio increases with increasing viscosity and thus the fatigue damage decreases. However, this calculation by no means implies that the greater the Λ ratio, the better the performance. Thick lubricant films have disadvantages such as high power losses or high operating temperatures and a level of understanding beyond the Λ ratio is necessary to optimise operating parameters (Cann et al. 1994).

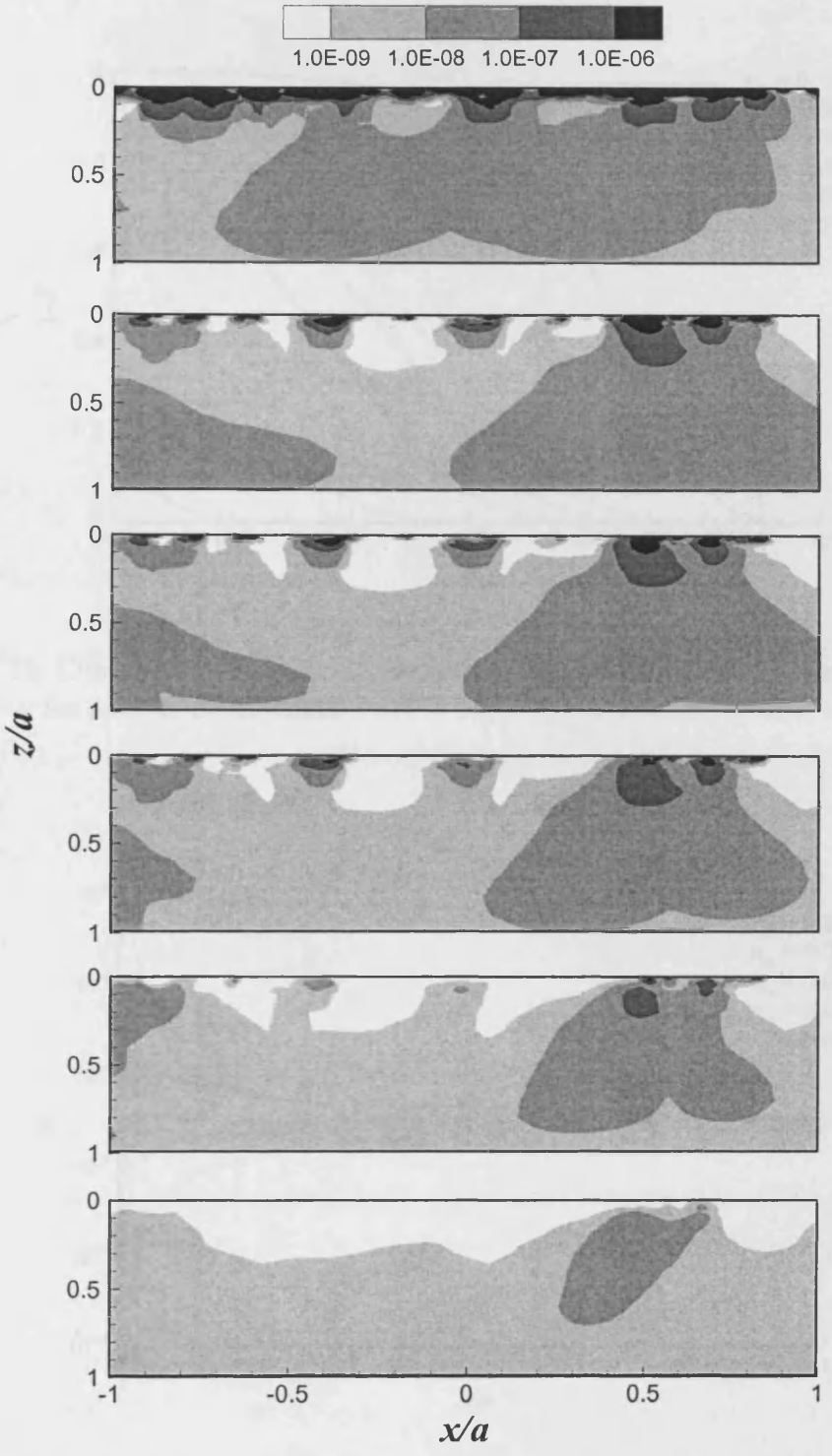


Figure 7.14: Contours of subsurface damage calculated for the section of the slower moving surface illustrated in Figure 7.4 during its transit of the contact area with ξ = 0.5 and η_0 = 0 (upper figure), 0.001, 0.002, 0.004, 0.008, 0.016 (lower figure) Pa·s.

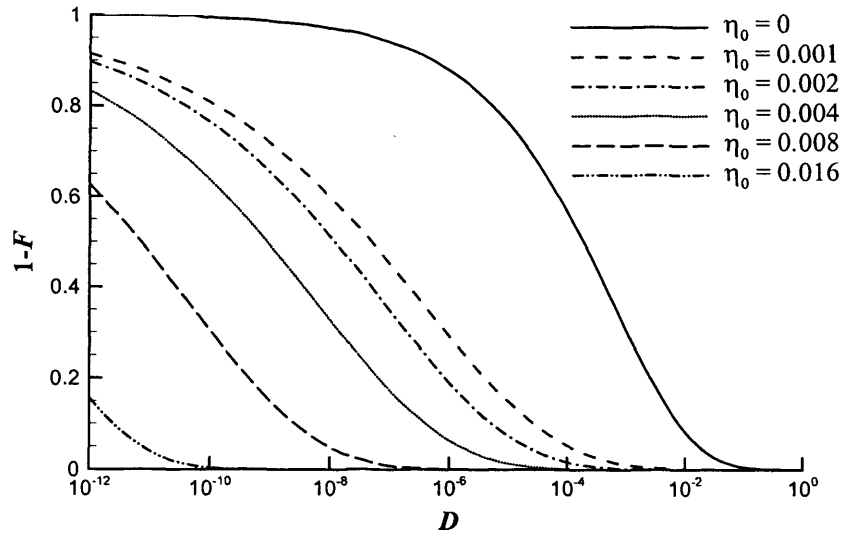


Figure 7.15: Cumulative damage distributions at a depth of z/a = 0.01 beneath the surface for a series of η_0 values.

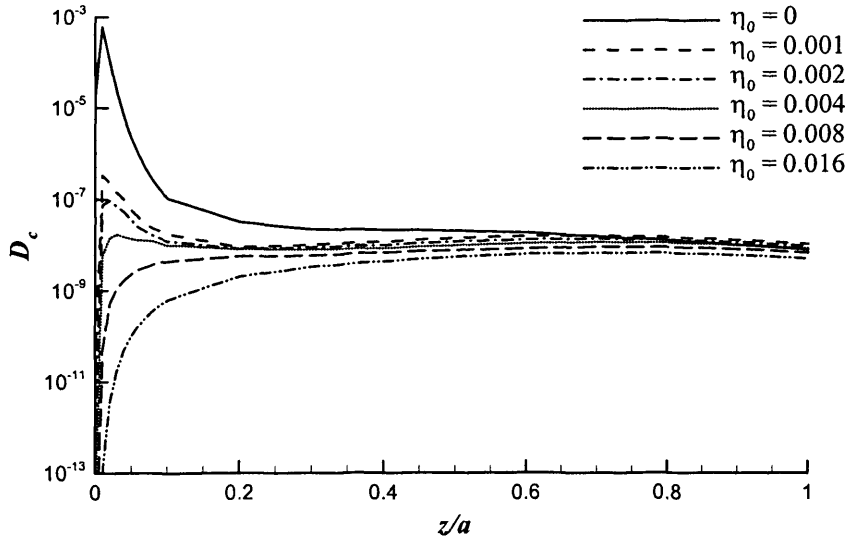


Figure 7.16: Variations of D_c versus the depth for a series of η_0 values.

Table 7.5: Maximum calculated damage for a range of η_0.

Viscosity, η_0 (Pa·s)	Λ ratio	Maximum damage, D_{max}	Location
0	–	5.265×10^{-3}	C
0.001	0.384	1.043×10^{-4}	J
0.002	0.611	6.579×10^{-5}	J
0.004	0.973	2.095×10^{-5}	J
0.008	1.548	1.768×10^{-6}	H
0.016	2.448	3.490×10^{-8}	H

7.4.4 Effect of Roughness

The influence of roughness on the fatigue damage can be further investigated by the use of different rough profiles. Three groups of contact conditions are considered by combinations of different rough profiles. These contact conditions are as follows:

(1) A-A, B-A and C-A. That is, the profile (A) is considered as the slower surface and the opposing surfaces are, respectively, profiles (A), (B) and (C).

(2) A-B, B-B and C-B. That is, the profile (B) is considered as the slower surface and the opposing surfaces are, respectively, profiles (A), (B) and (C).

(3) A-C, B-C and C-C. That is, the profile (C) is considered as the slower surface and the opposing surfaces are, respectively, profiles (A), (B) and (C).

Contours of damage obtained for a section of material of $2a\times a$ within the slower surface are shown in Figures 7.17 to 7.19 for the three groups of contact situations with $\xi =$ 0.5, $\bar{u}$ = 25m/s and η_0 = 0.001 Pa·s. The upper part of each figure is the corresponding shape of the rough profile considered as the slower surface. In each group of these contact situations, profile (C) as a faster moving surface generally causes more localised damage near the surface asperities than other two profiles and thus the volume surrounding the peak damage levels are larger.

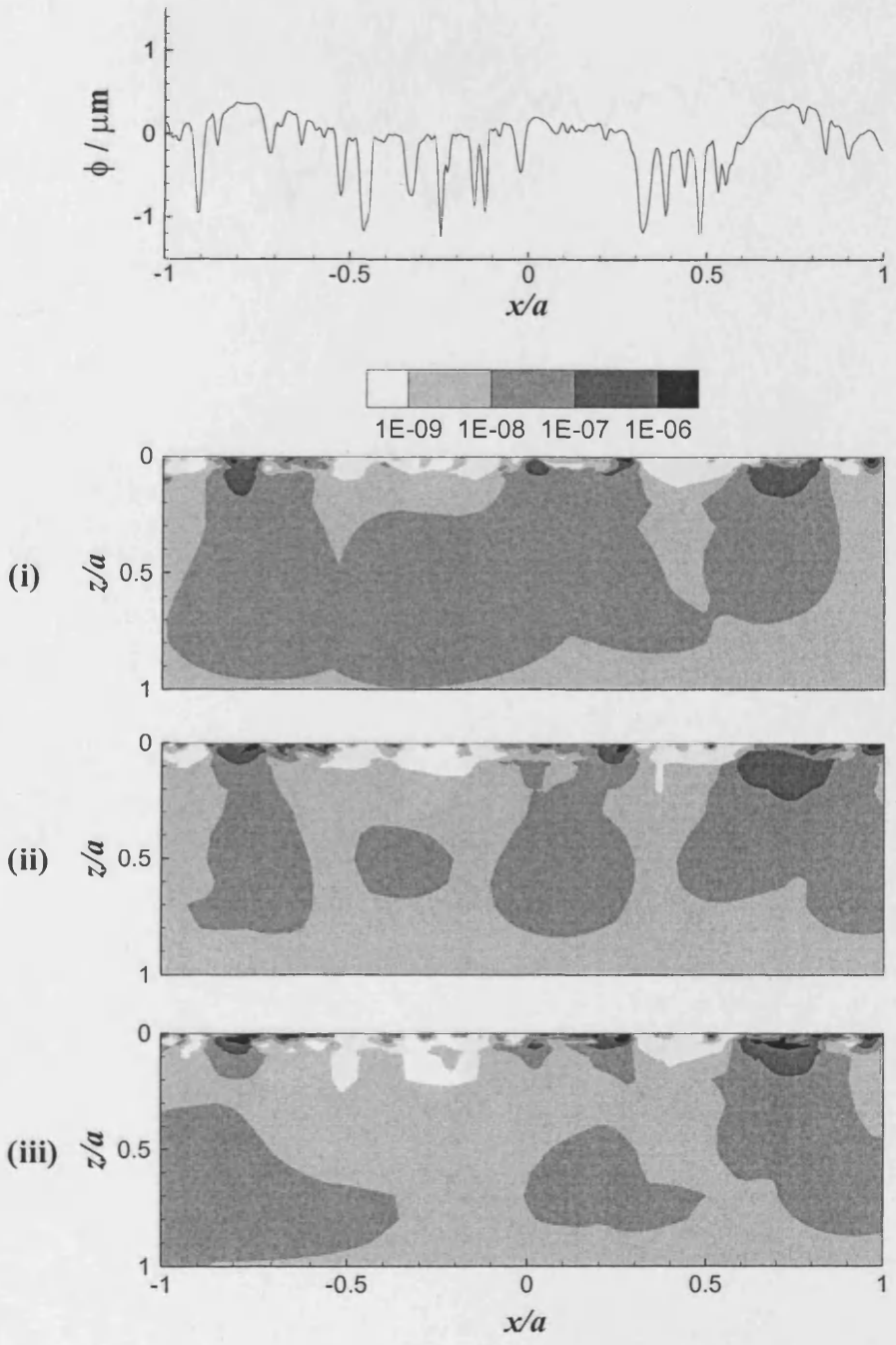


Figure 7.17: Contours of calculated damage of the slower moving surface (A) with ξ = 0.5, $\bar{u}$ = 25m/s and η_0 = 0.001Pa·s for (i) A-A, (ii) B-A and (iii) C-A. The rough profile is shown in the upper figure.

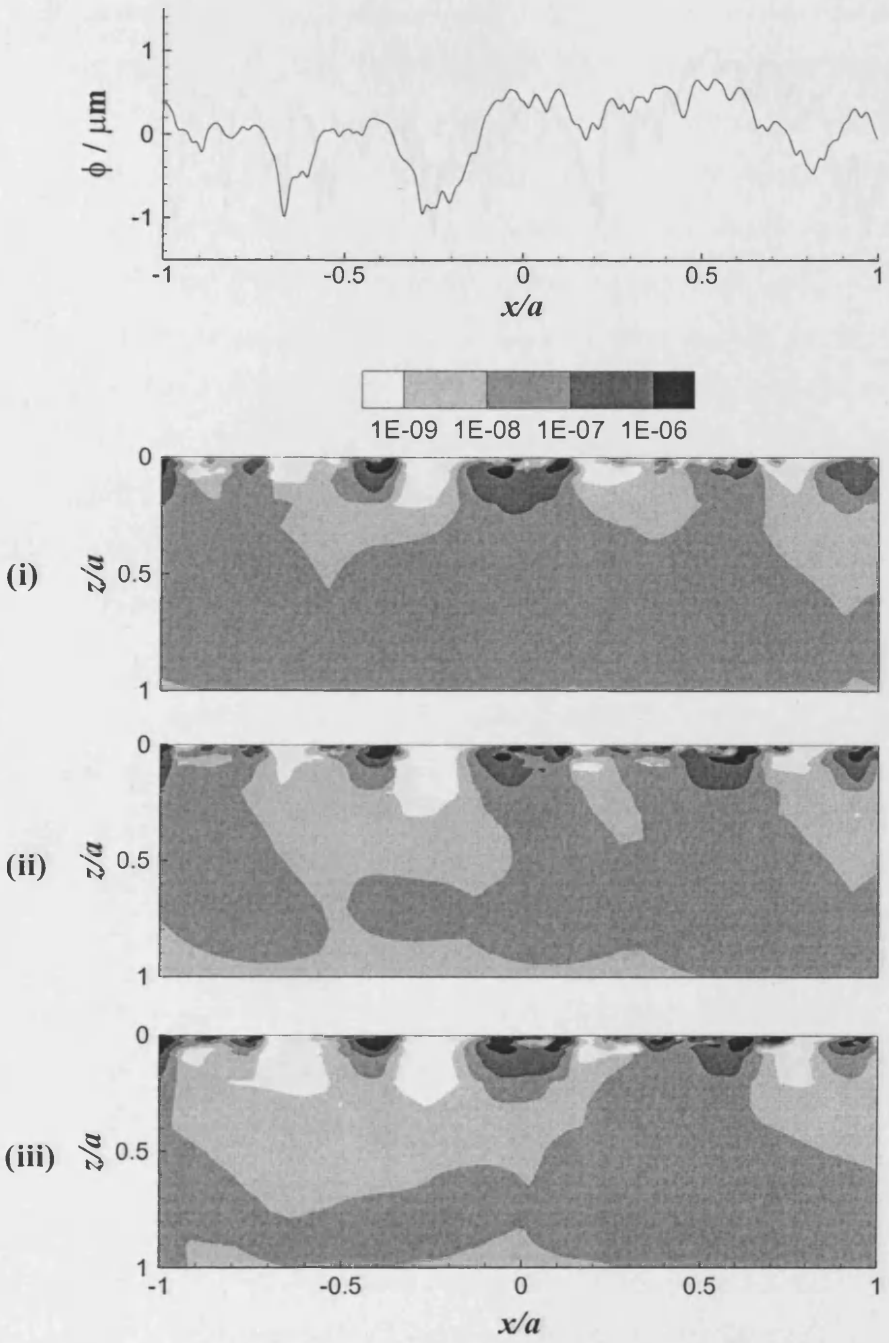


Figure 7.18: Contours of calculated damage of the slower moving surface (B) with ξ = 0.5, $\overline{u}$ = 25m/s and η_0 = 0.001Pa·s for (i) A-B, (ii) B-B and (iii) C-B. The rough profile is shown in the upper figure.

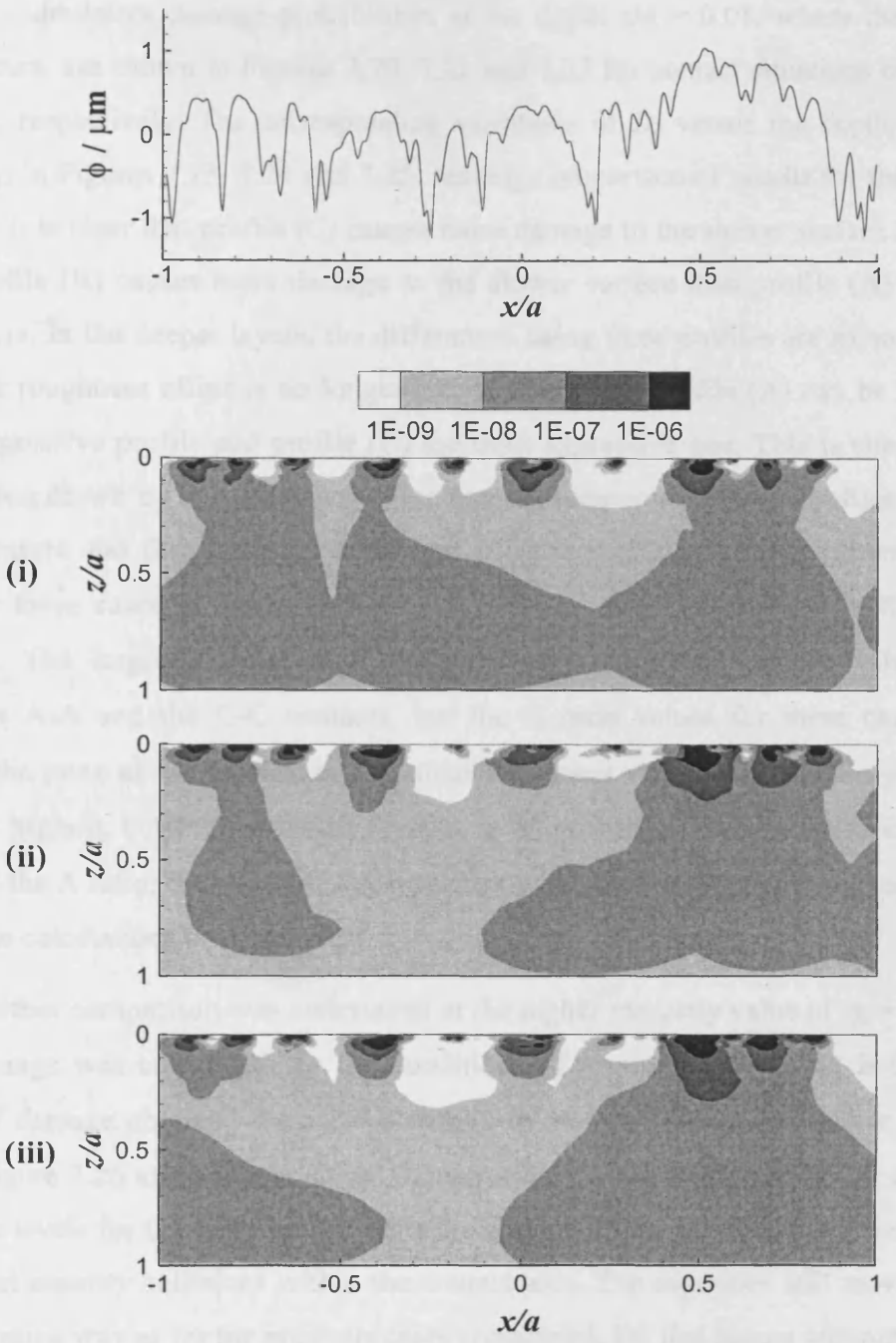


Figure 7.19: Contours of calculated damage of the slower moving surface (C) with ξ = 0.5, $\bar{u}$ = 25m/s and η_0 = 0.001Pa·s for (i) A-C, (ii) B-C and (iii) C-C. The rough profile is shown in the upper figure.

The cumulative damage probabilities at the depth z/a = 0.01, where the maximum damage occurs, are shown in Figures 7.20, 7.21 and 7.22 for contact situations of group (1), (2) and (3), respectively. The corresponding variations of D_c versus the depth are shown, respectively, in Figures 7.23, 7.24 and 7.25. Making comparison of results for the profiles in each group, it is clear that profile (C) causes more damage to the slower surface than profile (B) and profile (B) causes more damage to the slower surface than profile (A) in the near surface layers. In the deeper layers, the differences using three profiles are minor due to the fact that the roughness effect is no longer active. Therefore, profile (A) can be regarded as the least aggressive profile and profile (C) the most aggressive one. This is consistent with the conclusion drawn by considering the line contact behaviour of these profiles in terms of extreme pressure and film thickness behaviour (Tao et al. 2002). The maximum calculated damage for these cases is listed in Table 7.6. The Λ ratio is also listed in the table for comparison. The largest difference of the calculated maximum damage values is seen between the A-A and the C-C contacts, but the Λ ratio values for these two cases are essentially the same as the profiles have similar roughness values. The Λ ratio value for the B-B case is highest, but the calculated damage is intermediate. These observations indicate that besides the Λ ratio, the shape of the asperities is an important property of the surface as far as fatigue calculations are concerned.

A further comparison was undertaken at the higher viscosity value of η_0 = 0.008 Pa·s. Contact damage was calculated for the conditions of group (3), i.e. A-C, B-C and C-C. Contours of damage obtained for a $2a \times a$ section of material within the slower surface are shown in Figure 7.26 and corresponding D_c curves are shown in Figure 7.27. It can be seen that damage levels for the three profile pairs are essentially the same due to the low number of significant asperity collisions within the contact area. The asperities still move past each other in the same way as for the previous cases considered, but this occurs without generating micro-EHL pressures of the same magnitude between the asperities. As a result the effects of roughness are reduced with higher viscosity.

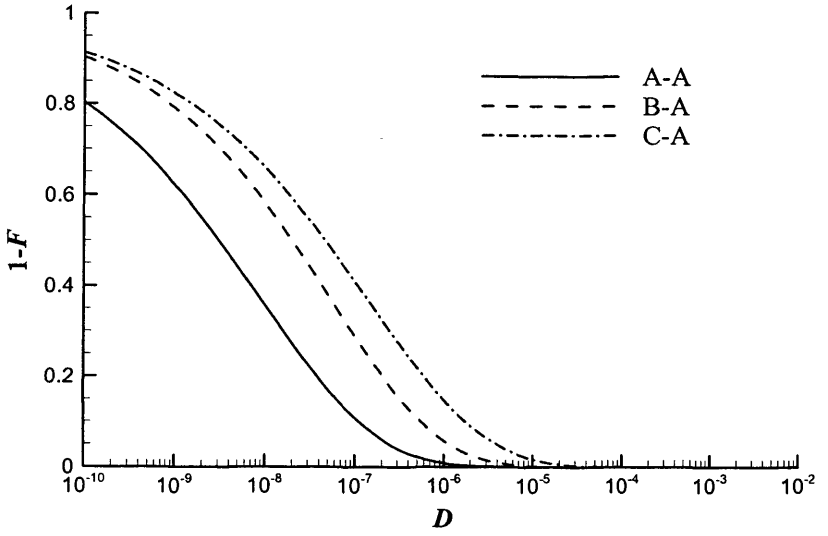


Figure 7.20: Cumulative damage distributions at a depth of z/a = 0.01 beneath the surface for A-A, B-A and C-A.

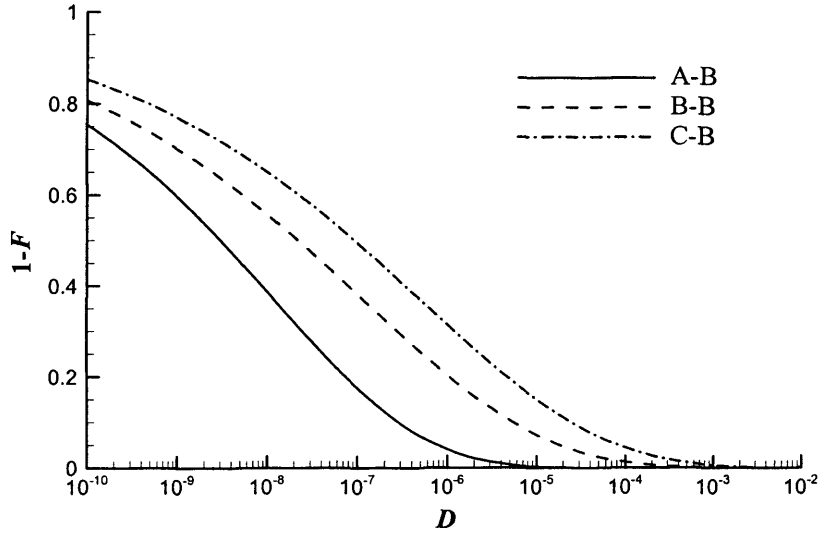


Figure 7.21: Cumulative damage distributions at a depth of z/a = 0.01 beneath the surface for A-B, B-B and C-B.

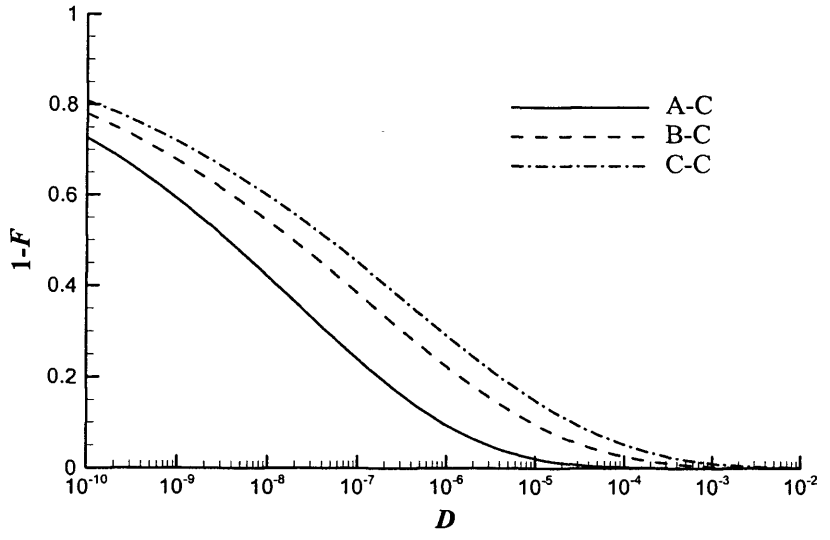


Figure 7.22: Cumulative damage distributions at a depth of z/a = 0.01 beneath the surface for A-C, B-C and C-C.

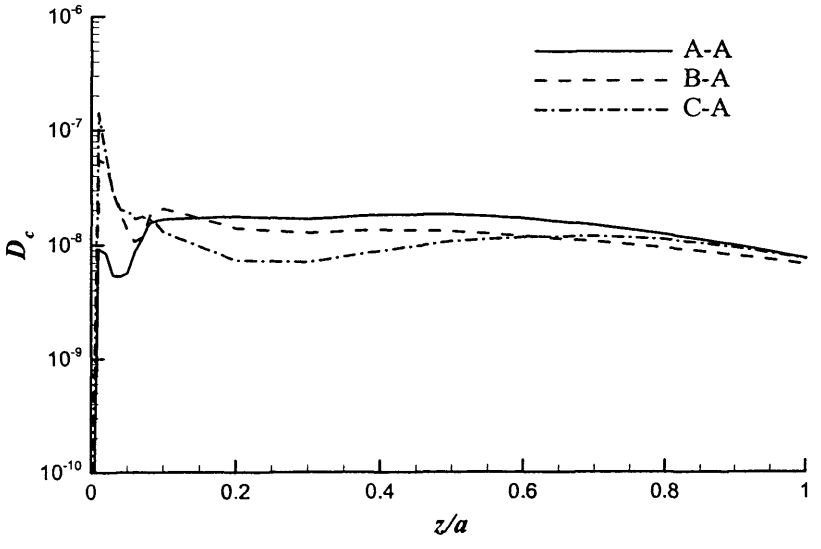


Figure 7.23: Variations of D_c versus the depth for A-A, B-A and C-A.

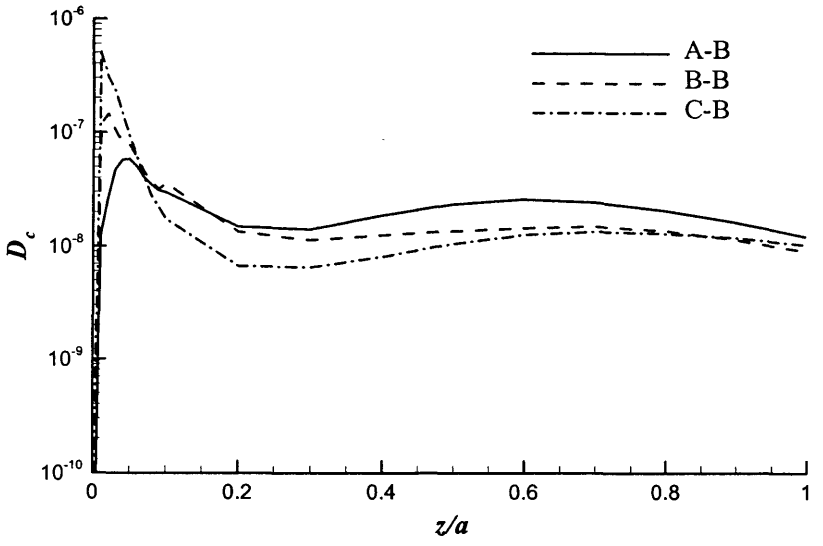


Figure 7.24: Variations of D_c versus the depth for A-B, B-B and C-B.

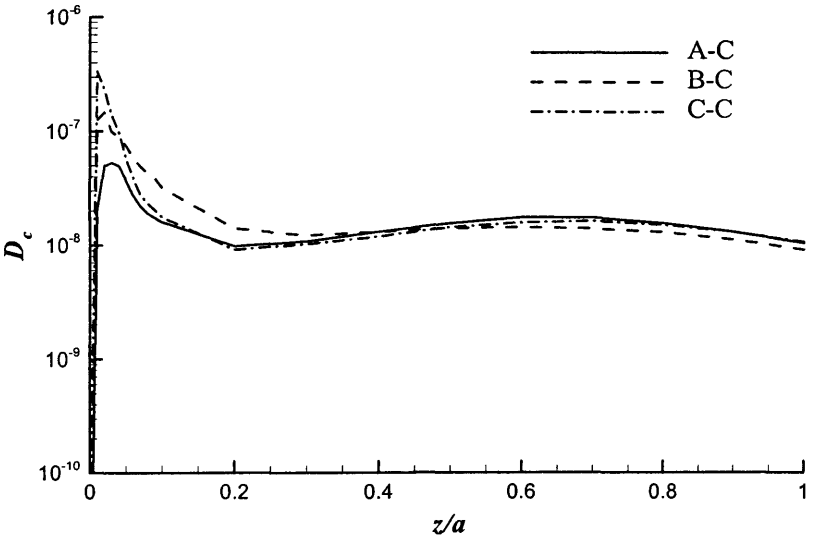


Figure 7.25: Variations of D_c versus the depth for A-C, B-C and C-C.

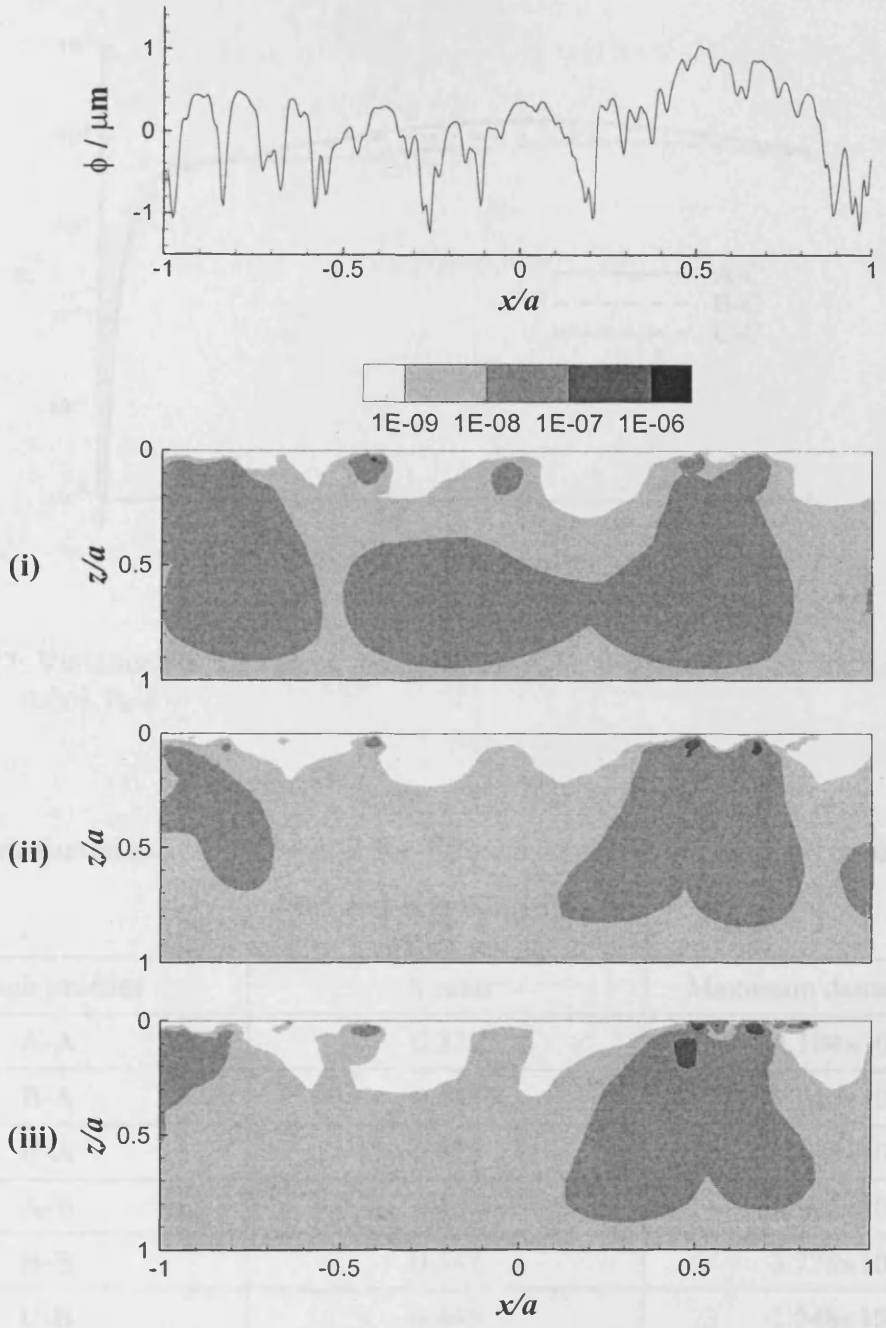


Figure 7.26: Contours of calculated damage of the slower surface illustrated in Figure 7.19 with the case of η_0 = 0.008 Pa·s for (i) A-C, (ii) B-C and (iii) C-C.

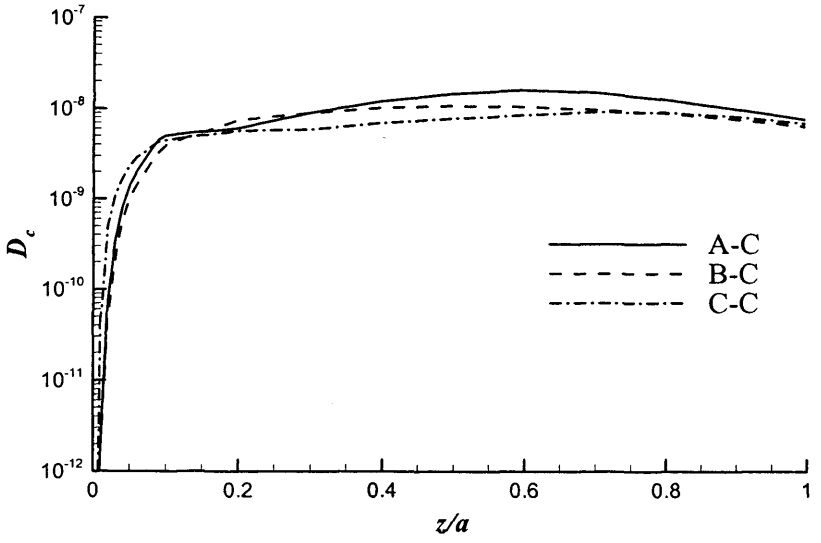


Figure 7.27: Variations of D_c versus the depth for A-C, B-C and C-C for the case of η_0 = 0.008 Pa·s.

Table 7.6: Maximum calculated damage for different combinations of rough profiles with ξ = 0.5 and η_0 = 0.001 Pa·s.

Rough profiles	Λ ratio	Maximum damage, D_{max}
A-A	0.372	4.104×10^{-6}
B-A	0.434	7.040×10^{-6}
C-A	0.378	1.490×10^{-5}
A-B	0.434	2.992×10^{-5}
B-B	0.542	3.778×10^{-5}
C-B	0.443	1.048×10^{-4}
A-C	0.378	2.866×10^{-5}
B-C	0.443	3.761×10^{-5}
C-C	0.384	1.043×10^{-4}

The relationship between the maximum calculated damage, D_{max}, and the Λ ratio for Tables 7.5 and 7.6 is visually shown in Figures 7.28 and 7.29, respectively. It can be seen that D_{max} changes with Λ monotonically for the case with viscosity variations, but this is not true for the case with rough profile variations.

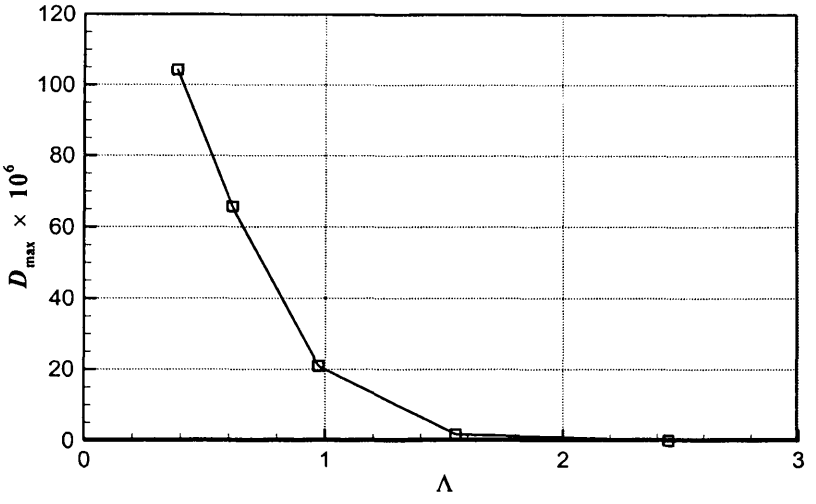


Figure 7.28: Variations of D_c versus the Λ ratio for the case of viscosity variations.

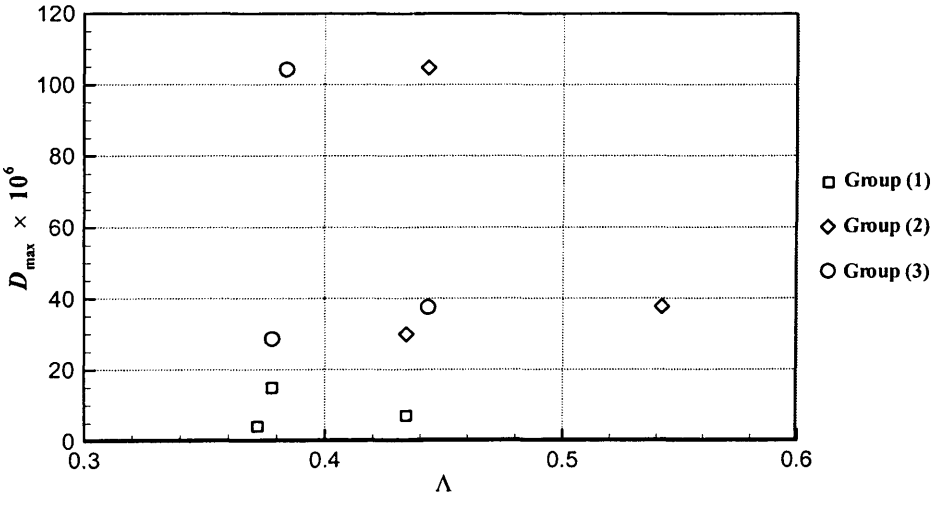


Figure 7.29: Variations of D_c versus the Λ ratio for the case of profile variations.

7.5 Summary

A numerical procedure to estimate the cumulative fatigue damage in a rough EHL contact is developed in this chapter. The procedure is based on the critical plane approach and the Fatemi-Socie shear strain model. The plane of maximum damage is treated as the critical plane. The rainflow cycle counting method is applied to quantify fatigue cycles from the shear strain history. The Miner-Palmgren linear damage rule is used to compute the cumulative damage. The effects of slide roll ratio, viscosity and surface roughness on fatigue damage are demonstrated in numerical analyses and can be summarized as follows:

- Higher sliding generally leads to greater calculated fatigue damage.
- Increasing viscosity generally leads to decreasing calculated fatigue damage.
- Surfaces with smaller roughness may provide longer life; however, the shape of the asperities also plays a significant role.

The numerical results have shown that particular asperities can sustain calculated damage levels that are significantly higher than those experienced by the surrounding material. The location and size of the area within the material for which high damage is calculated is consistent with the type of surface damage seen in micropitting.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

A numerical procedure for predicting contact fatigue damage in elastohydrodynamic lubricated line contacts between two rough surfaces is presented in this thesis. Various aspects of this procedure are described in detail in the previous chapters.

The coupled solution method incorporating a differential deflection technique is used to determine the pressure and film thickness distributions in rough surface EHL problem. The Eyring rheological model is adopted to account for the non-Newtonian effects. Real (or measured) roughness is included in a transient line contact analysis. Interpolation is necessary for real roughness data whose measuring spacing is different from the computational mesh size. The use of Fourier series is demonstrated and comparison with the cubic spline approach is made. Numerical results have shown that the two approaches give essentially the same answer.

Thermal effects are considered by solving the energy equations. Numerical results have shown that the temperature rise is dominated by shear heating. With increased slide roll ratio, speed, viscosity and pressure coefficient, the temperature rise is found to increase. Generally, temperature rise has little effect on the pressure distributions, although it plays an important role in scuffing failures. Therefore, thermal effects are not considered in the subsequent contact stress and fatigue calculations.

Elastic stress analysis is performed to evaluate stress fields and stress histories experienced by the contacting bodies. The DC-FFT method is adopted to determine the

convolution integrals for the three stress components in a plane strain problem. The accuracy and efficiency of the approach is demonstrated by comparison with other approaches. Numerical results have shown that stress deviations and cycling are most apparent close to the surface of the material due to the presence of surface roughness.

Fatigue damage is predicted by applying the Ioannides and Harris (1985) damage model together with multiaxial fatigue criteria (such as the Findley criterion, the Matake criterion, and the Dang Van criterion), and the Fatemi and Socie (1988) shear strain model together with the rainflow counting method and the cumulative damage theory. Various factors affecting fatigue damage are investigated numerically. The results obtained have shown that:

(1) Prominent asperities can sustain calculated damage levels that are significantly higher than those experienced by the surrounding material. The highest damage levels occur in the near surface layer due to the presence of surface roughness.

(2) Both the height and shape of surface asperities play a significant role in fatigue calculations. Λ ratio alone is not enough to describe the severity of contact between two rough surfaces regarding fatigue calculations.

(3) The damage for the slower surface in a contacting pair is generally greater than that for the faster surface because asperities of the slower moving surface are subject to higher numbers of stress cycles in traversing the contact than are those of the faster moving surface.

(4) Increasing sliding generally leads to an increase in the calculated fatigue damage.

(5) Increasing viscosity generally leads to decreasing calculated fatigue damage.

8.2 Future Work

Contact fatigue in gears is a complex problem and has attracted many researchers in both theoretical and experimental studies. As stated above, this thesis introduces a numerical

procedure to predict the contact fatigue damage in rough surface EHL. Due to time constraints, there are still many important issues not considered in this study. The following gives a list of tasks that may be considered in the future for further investigation:

(1) Some of the conclusions drawn in Section 8.1 are already verified by experiments. To further study the validity and applicability of the procedure for prediction of fatigue damage in rough surface EHL, more well designed experiments are required to correlate numerical results with experimental results.

(2) The work reported in this thesis assumes that the material is homogeneous. That is, the material properties are independent of position. Real gears are often treated by surface hardening, surface coatings, etc. to increase the resistance to fatigue damage. Therefore, the variation of material properties should be considered for such cases.

(3) The work reported in this thesis assumes that the material is perfect. That is, there are no voids, inclusions or initial cracks. If these initial impurities are considered, an approach based on crack propagation, such as that proposed by Kudish (2004), may be required.

(4) Residual stresses, which are produced in machining process and heat treatment or created due to plastic deformation especially during the early contact stages, are widely recognized to have an effect on fatigue life. Therefore, measuring and modeling residual stresses are very important for predicting the fatigue damage correctly.

References

Abernethy, R.B. The New Weibull Handbook, Second Edition. ISBN0-9653062-0-8, published by R.B. Abernethy, 1996.

Adams, G.G. and Nosonovsky, M. Contact modeling – forces. Tribology International, 2000; 33: 431-442.

Ai, X. Effect of three-dimensional random surface roughness on fatigue life of a lubricated contact. Transactions of the ASME, Journal of Tribology, 1998; 120: 159-164.

Ai, X. and Cheng, H.S. Transient EHL analysis for line contacts with measured surface roughness using multigrid technique. Transactions of the ASME, Journal of Tribology, 1994; 116(3): 549-558.

Ai, X. and Sawamiphakdi, K. Solving elastic contact between rough surfaces as an unconstrained strain energy minimization by using CGM and FFT techniques. Transactions of the ASME, Journal of Tribology, 1999; 121: 639-647.

Allwood, J. Survey and performance assessment of solution methods for elastic rough contact problems. Transactions of the ASME, Journal of Tribology, 2005; 127: 10-23.

Amzallag, C., Gerey, J.P., Robert, J.L. and Bahuaud, J. Standardization of the rainflow counting method for fatigue analysis. International Journal of Fatigue, 1994; 16: 287-293.

Anthes, R. J. Modified rainflow counting keeping the load sequence. International Journal of Fatigue, 1997; 19: 529-535.

August, R. and Zaretsky, E.V. Incorporating finite element analysis into component life and reliability. Transactions of the ASME, Journal of Mechanical Design, 1993; 115(4): 706-710.

Ayyub, B.M. and McCuen, R.H. Numerical Methods for Engineers. Prentice Hall, New Jersey, 1996.

Bailey, D.M. and Sayles, R.S. Effect of roughness and sliding friction on contact stresses. Transactions of the ASME, Journal of Tribology, 1991; 113: 729-738.

Bannantine, J.A. and Socie, D.F. A multiaxial fatigue life estimation technique. In: Advances in Fatigue Lifetime Predictive Techniques, ASTM STP 1122, M.R. Mitchell and R.W. Landgraf, Eds,. American Society for Testing and Materials, Philadelphia, 1992, pp249-275.

Bernasconi, A. Efficient algorithms for calculation of shear stress amplitude and amplitude of the second invariant of the stress deviator in fatigue criteria applications. International Journal of Fatigue 2002; 24: 649-657.

Bhushan, B. Contact mechanics of rough surfaces in tribology: multiple asperity contact. Tribology Letter, 1998; 4: 1-35.

Boardman, B. Fatigue resistance of steels. In: Metals Handbook, Tenth Edition, Volume 1 Properties and Selection: Irons, Steels, and High-Performance Alloys. Prepared under the direction of the ASM International Handbook Committee, 1990, pp673-688.

Brigham, E.O. The Fast Fourier Transform and Its Applications. Prentice Hall Inc., Englewood, New Jersey, 1988.

Cann, P., Ioannides, E., Jacobson, B. and Lubrecht, A.A. The lambda ratio – a critical re-examination. Wear, 1994; 175: 177-188.

Carpinteri, A. and Spagnoli, A. Multiaxial high-cycle fatigue criterion for hard metals. International Journal of Fatigue, 2001; 23: 135-145.

Carslaw, H.S. and Jaeger, J.C. Conduction of Heat in Solids, Second Edition. Oxford University Press, Oxford, 1959.

Chang, L. Traction in thermal elastohydrodynamic lubrication of rough surfaces. Transactions of the ASME, Journal of Tribology, 1992; 114(1):186-191.

Chang, L. Deterministic modeling and numerical simulation of lubrication between rough surfaces – a review of recent developments. Wear, 1995a; 184: 155-160.

Chang, L. A deterministic model for line contact partial elastohydrodynamic lubrication. Tribology International, 1995b; 28(2): 75-84.

Chang, L. and Farnum, C. A thermal model for elastohydrodynamic lubrication of rough surfaces. Tribology Transactions, 1992; 35(2): 281-286.

Chang, L. and Webster, M.N. A study of elastohydrodynamic lubrication of rough surfaces. Transactions of the ASME, Journal of Tribology, 1991; 113: 110-115.

Chang, L. and Zhao, W. Fundamental differences between Newtonian and non-Newtonian micro-EHL results. Transactions of the ASME, Journal of Tribology, 1995: 117(1): 29-35.

Chang, L. and Zhao, Y. On the sensitivity of the asperity pressures and temperatures to the fluid pressure distribution in mixed-film lubrication. Transactions of ASME, Journal of Tribology, 2000; 122(1): 77-85.

Chiu, Y.P. The mechanism of bearing surface fatigue – experiments and theories. Tribology Transactions, 1997; 40(4): 658-666.

Chu, C.C. Fatigue damage calculation using the critical plane approach. Transactions of the ASME, Journal of Engineering Materials and Technology, 1995; 117: 41-49.

Ciavarella, M. and Maitournam, H. 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, 2004; 27: 523-528.

Colangelo, V. J. and Heiser, F. A. Analysis of Metallurgical Failures, Second Edition. John Wiley & Sons, 1987.

Cole, S.J. and Sayles, R.S. A numerical model for the contact of layered elastic bodies with real rough surfaces. Transactions of the ASME, Journal of Tribology, 1992; 114: 334-340.

Colin, F. and Lubrecht, A.A. Comparison of FFT-MLMI for elastic deformation calculations. Transactions of the ASME, Journal of Tribology, 2001; 123: 884-887.

Conry, T.F., Wang, S. and Cusano, C. A Reynolds-Eyring equation for elastohydrodynamic lubrication in line contacts. Transactions of the ASME, Journal of Tribology, 1987; 109: 648-654.

Dagnall, H. Exploring Surface Texture. Rank Taylor Hobson, Leicester, 1986.

Dang Van, K., Cailletaud, G., Flavenot, J.F., Le Douaron, A. and Lieurade, H.P. Criterion for high cycle fatigue failure under multiaxial loading. In: Biaxial and Multiaxial Fatigue, EGF 3 (Edited by M.W. Brown and K.J. Miller), Mechanical Engineering Publications, London, 1989, pp459-478.

Datsyshyn, O.P. and Panasyuk, V.V. Pitting of rolling bodies contact surface. Wear, 2001; 251:1347-1355.

Davies, C. Effects of Non-Newtonian Rheology on the Line Contact Elastohydrodynamic Lubrication Problem. PhD Thesis, Cardiff University, 2005.

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

Downing, S.D. Simple rainflow counting algorithms. International Journal of Fatigue, 1982; 4: 31-40.

Downson, D and Higginson, G.R. Elasto-hydrodynamic Lubrication. Pergamon Press, Oxford, 1966.

Ekberg, A. Rolling contact fatigue of railway wheels – a parametric study. Wear, 1997; 211: 280-288.

Elcoate, C.D., Evans, H.P., Hughes, T.G. and Snidle, R.W. 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, 2001; 215: 319-337.

Elsharkawy, A.A. and Hamrock, B.J. Subsurface stresses in micro-EHL line contacts. Transactions of the ASME, Journal of Tribology; 1991; 113: 645-655.

Epstein, D., Yu, T., Wang, Q.J., Keer, L.M., Cheng, H.S., Liu, S. Harris, S.J. and Gangopadhyay, A. An efficient method of analyzing the effect of roughness on fatigue life in mixed-EHL contact. Tribology Transactions, 2003a; 46(2): 273-281.

Epstein, D., Keer, L.M., Wang, Q.J., Cheng, H.S. and Zhu, D. Effect of surface topography on contact fatigue in mixed lubrication. Tribology Transactions, 2003b; 46(4): 506-513.

Evans, H.P. and Hughes, T.G. 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, 2000; 214: 563-584.

Fatemi, A. and Socie, D.F. A critical plane approach to multiaxial fatigue damage including out-of-phase loading. Fatigue & Fracture of Engineering Materials & Structures, 1988; 11(3):149-165.

Fatemi, A. and Yang, L. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. International Journal of Fatigue, 1998; 20(1): 9-34.

Fernandes, P.J.L. and McDuling, C. Surface contact fatigue failures in gears. Engineering Failure Analysis, 1997; 4(2): 99-107.

Fernandez Rico, J.E., Hernandez Battez, A. and Garcia Cuervo, A. Rolling contact fatigue in lubricated contacts. Tribology International, 2003; 36: 35-40.

Flasker, J., Fajdiga, G., Glodez, S. and Hellen, T.K. Numerical simulation of surface pitting due to contact loading. International Journal of Fatigue, 2001; 23: 599-605.

Gao, J., Lee, S.C., Ai, X. and Nixon, H. An FFT-based transient flash temperature model for general three-dimensional rough surface contacts. Transactions of the ASME, Journal of Tribology, 2000; 122: 519-523.

Gao, C., Qi, X., Evans, H.P. and Snidle, R.W. Effect of film thickness ratio on gearing contact fatigue life. Proceedings of IUTAM Symposium on Elastohydrodynamics and Micro-Elastohydrodynamics, 2005, to be published.

Girish, D.V., Mayuram, M.M. and Krishnamurthy, S. Influence of shot peening on the surface durability of thermomechanically treated En24 steel spur gears. Tribology International 1997; 30(12): 865-870.

Glinka, G. and Kam, J.C.P. Rainflow counting algorithm for very long stress histories. International Journal of Fatigue, 1987; 9: 223-228.

Glodez, S., Ren, Z. and Flasker, J. Simulation of surface pitting due to contact loading. International Journal for Numerical Methods in Engineering, 1998; 43: 33-50.

Glovnea, R.P., Forrest, A.K., Olver, A.V. and Spikes, H.A. Measurement of sub-nanometer lubricant films using ultra-thin Film interferometry. Tribology Letters, 2003; 15(3): 217-230.

Goglia, P., Cusano, C. and Conry, T.F. The effects of surface irregularities on elastohydrodynamic lubrication of sliding line contacts – part I: single irregularities. Transactions of the ASME, Journal. of Tribology, 1984a; 106(1): 104-112.

Goglia, P., Cusano, C. and Conry, T.F. The effects of surface irregularities on elastohydrodynamic lubrication of sliding line contacts – part II: wavy surfaces. Transactions of the ASME, Journal of Tribology, 1984b; 106(1): 113-119.

Gohar, R. Elastohydrodynamics. Ellis Horwood Limited, Chichester, 1988.

Guagliano, M., Riva, E. and Guidetti, M. Contact fatigue failure analysis of shot-peened gears. Engineering Failure Analysis, 2002; 9: 147-158.

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

Hoglund, E. Influence of lubricant properties on elastohydrodynamic lubrication. Wear, 1999; 232: 176-184.

Hohn, B.R. and Michaelis, K. Influence of oil temperature on gear failures. Tribology International, 2004; 37:103-109.

Holmes, M.J.A., Evans, H.P., Hughes, T.G. and Snidle, R.W. 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, 2003a; 217: 289-303.

Holmes, M.J.A., Evans, H.P., Hughes, T.G. and Snidle, R.W. 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, 2003b; 217: 305-321.

Holmes, M.J.A., Qiao, H., Evans, H.P. and Snidle, R.W. Transient effects in EHL point and line contacts having rough transverse surface finish. In: Proceedings of the 30th Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 2004, pp201-212.

Holmes, M.J.A., Qiao, H., Evans, H.P. and Snidle, R.W. Surface contact and damage in micro-EHL. In: Proceedings of the 31st Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 2005a, to be published.

Holmes, M.J.A., Evans, H.P. and Snidle, R.W. Analysis of mixed lubrication effects in simulated gear tooth contacts. Transactions of the ASME, Journal of Tribology, 2005b; 127: 61-69.

Hooke, C.J. and Li, K.Y. An inverse approach to the validation of pressure predictions in rough elastohydrodynamic contacts. Transactions of the ASME, Journal of Tribology, 2002; 124, 103-108.

Hooke, C.J. and Li, K.Y. Use of the inverse method to check the stress predictions in rough elastohydrodynamic contacts lubricated with a non-Newtonian fluid for a range of asperity geometries. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2003; 217: 935-944.

Hsiao, H.-S. and Hamrock, B.J. Complete solution for thermal-elastohydrodynamic lubrication of line contacts using circular non-Newtonian fluid model. Transactions of the ASME, Journal of Tribology, 1992; 114(3):540-552.

Hsiao, H.-S. and Hamrock, B.J. Non-Newtonian and thermal effects on film generation and traction reduction in EHL line contact conjunctions. Transactions of the ASME, Journal of Tribology, 1994; 116(3):559-568.

Hu, Y.-Z., Barber, G.C. and Zhu, D. Numerical analysis for the elastic contact of real rough surfaces. Tribology transactions, 1999; 42(3): 443-452.

Hu, Y.-Z. and Zhu, D. Full numerical solution to the mixed lubrication in point contacts. Transactions of the ASME, Journal of Tribology, 2000; 122(1): 1-9.

Hughes, T.G., Elcoate, C.D. and Evans, H.P. 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, 2000; 214: 585-598.

Ioannides, E. and Harris, T.A. A new fatigue life model for rolling bearings. Transactions of the ASME, Journal of Tribology, 1985; 107: 367-378.

Ioannides, E., Jacobson, B. and Tripp, J.H. Prediction of rolling bearing life under practical operating conditions. In: Proceedings of 15th Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 1989, pp181-187.

Jiang, X., Hua, D.Y., Cheng, H.S., Ai, X. and Lee, S.C. Mixed elastohydrodynamic lubrication model with asperity contact. Transactions of the ASME, Journal of Tribology, 1999; 121(3): 481-491.

Jiang, Y. A fatigue criterion for general multiaxial loading. Fatigue & Fracture of Engineering Materials & Structures, 2000; 23: 19-32.

Jiang, Y. and Sehitoglu, H. A model for rolling contact failure. Wear, 1999; 224: 38-49.

Johnson, K.L. Contact Mechanics. Cambridge University Press, 1985.

Johnson, K.L. The strength of surfaces in rolling contact. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 1989; 203: 151-163.

Kadiric, A., Sayles, R.S., Zhou, X.B. and Ioannides, E. A numerical study of the contact mechanics and sub-surface stress effects experienced over a range of machined surface coatings in rough surface. Transactions of the ASME, Journal of Tribology, 2003; 125: 720-730.

Kazama, T. Effects of approximations regarding stress and temperature profiles on the solution in thermal elastohydrodynamically lubricated line contacts. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2004; 218(2): 83-93.

Kazama, T., Ehret, P. and Taylor, C.M. On the effects of temperature profile approximation in thermal Newtonian solutions of elastohydrodynamic lubrication line contacts. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2001; 215(1): 109-120.

Khonsari, M.M. and Hua, D.Y. Thermal elastohydrodynamic analysis using a generalized non-Newtonian formulation with application to Bair-Winer constitutive equation. Transactions of the ASME, Journal of Tribology, 1994; 116(1):37-45.

Kim, K.S. and Park, J.C. Shear strain based multiaxial fatigue parameters applied to variable amplitude loading. International Journal of Fatigue, 21(1999) 475-483

Kim, T.W., Cho, Y.J., An, D.M. and Lee, H.W. The fatigue crack initiation life prediction based on several high-cycle fatigue criteria under spherical rolling contact. Tribology Transactions, 2003; 46(1): 76-82.

Kim, T.H. and Olver, A.V. Fatigue and brittle fracture analysis of surface engineered material in rolling contact, elastohydrodynamics. In: Proceedings of 23rd Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 1997, pp37-48.

Kim, T.H. and Olver, A.V. Stress history in rolling-sliding contact of rough surfaces. Tribology International, 1998; 31(12): 727-736.

Kimura, Y., Sekizawa, M. and Nitanai, A. Wear and fatigue in rolling contact. Wear, 2002; 253: 9-16.

Kogut, L. and Etsion, I. A finite element based elastic-plastic model for the contact of rough surfaces. Tribology Transactions, 2003; 46(3): 383-390.

Krantz, T.L., Alanou, M.P., Evans, H.P. and Snidle, R.W. Surface fatigue lives of case-carburized gears with an improved surface finish. Transactions of the ASME, Journal of Tribology, 2001; 123: 709-716.

Krantz, T.L. and Kahraman, A. An experimental investigation of the influence of the lubricant viscosity and additives on gear wear. Tribology Transactions, 2004; 47(1): 138-148.

Kudish, I.I. and Burris, K.W. Modern state of experimentation and modeling in contact fatigue phenomenon: Part I-II. Tribology Transactions, 2000; 43(2): 187-196, 293-301.

Kudish, I.I. Lubricant-crack interaction, origin of pitting, and fatigue of drivers and followers. Tribology Transactions, 2002; 45(4): 583-594.

Kudish, I.I. and Burris, K.W. Modeling of surface and subsurface crack behavior under contact load in the presence of lubricant. International Journal of Fracture, 2004; 125: 125-147.

Kumar, P., Jain, S.C. and Ray, S. Thermal EHL of rough rolling/sliding line contacts using a mixture of two fluids at dynamic loads. Transactions of the ASME, Journal of Tribology, 2002; 124(4):709-715.

Kweh, C.C., Evans, H.P. and Snidle, R.W. Micro-elastohydrodynamic lubrication of an elliptical contact with transverse and three-dimensional sinusoidal roughness. Transactions of the ASME, Journal of Tribology, 1989; 111(4): 577-584.

Kweh, C.C., Patching, M.J., Evans, H.P. and Snidle, R.W. Simulation of elastohydrodynamic contact between rough surfaces. Transactions of the ASME, Journal of Tribology, 1992; 114(3): 412-419.

Lee, R.-T. and Hamrock, B.J. A circular non-Newtonian fluid model: Part I – used in elastohydrodynamic lubrication. Transactions of the ASME, Journal of Tribology, 1990; 112: 486-496.

Lee, R.-T. and Hsu, C.-H. A fast method for the analysis of thermal-elastohydrodynamic lubrication of rolling/sliding line contacts. Wear, 1993; 166:107-117.

Lee, R.-T., Hsu, C.-H. and Kuo, W.-F. Multilevel solution for thermal elastohydrodynamic lubrication of rolling /sliding circular contacts. Tribology International, 1995; 28(8): 541-552.

Liu, G., Wang, Q. and Lin, C. A survey of current models for simulating the contact between rough surfaces. Tribology Transactions, 1999; 42(3): 581-591.

Liu, G., Zhu, J., Yu, L. and Wang, Q. Elasto-plastic contact of rough surfaces. Tribology Transactions, 2001a; 44(3): 437-443.

Liu, Y.-C., Hu, Y.-Z., Wang, W.-Z. and Wang, H. Simulation of temperature distribution of point contacts in mixed lubrication. Science in China (Series E), 2002; 45(4): 365-372.

Liu, S., Rodgers, M.J., Wang, Q. and Keer, L.M. A fast and effective method for transient thermoelastic displacement analyses. Transactions of the ASME, Journal of Tribology, 2001b; 123: 479-485.

Liu, S. and Wang, Q. Studying contact stress fields caused by surface tractions with a discrete convolution and fast Fourier transform algorithm. Transactions of the ASME, Journal of Tribology, 2002; 124: 36-45.

Liu, S., Wang, Q., and Liu, G. A versatile method of discrete convolution and FFT (DC-FFT) for contact analysis. Wear, 2000; 243: 101-111.

Lubrecht, A.A., ten Napel, W.E. and Bosma, R. The influence of longitudinal and transverse roughness of the elastohydrodynamic lubrication of circular contacts. Transactions of the ASME, Journal of Tribology, 1988; 110(3): 421-426.

Ma, M.-T. Expedient approach to the non-Newtonian thermal EHL in heavily loaded point contacts. Wear, 1997; 206: 100-112.

Ma, M.-T. Effects of the non-Newtonian behavior of lubricants on the temperature, traction, and film thickness in an elliptical EHD contact under heavy loads. Transactions of the ASME, Journal of Tribology, 1998; 120(4): 685-694.

Mao, K., Bell, T. and Sun, Y. Effect of sliding friction on contact stresses for multi-layered elastic bodies with rough surfaces. Transactions of the ASME, Journal of Tribology, 1997; 119: 476-480.

Mayeur, C., Sainsot, P. and Flamand, L. A numerical elastoplastic model for rough contact. Transactions of the ASME, Journal of Tribology, 1995; 117: 422-429.

Meeker, W.Q. and Escobar, L.A. Statistical Methods for Reliability Data. John Wiley & Sons, New York, 1998.

Mihailidis, A., Bakolas, V. and Drivakos, N. Subsurface stress field of a dry line contact. Wear, 2001; 249: 546-556.

Mihailidis, A., Retzepis, J., Salpistis, C. and Panajiotidis, K. Calculation of friction coefficient and temperature field of line contacts lubricated with a non-Newtonian fluid. Wear, 1999; 232: 213-220.

Moore, A.J. The behavior of lubricants in elastohydrodynamic contacts. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 1997; 211: 91-106.

Moore, D.F. Principles and Applications of Tribology. Pergamon Press, Oxford, 1975.

Morales-Espejel, G.E., Kapoor, A. and Rodriguez-Sanchez, S. Shakedown in dry and lubricated surfaces. In: Proceedings of the 25th Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 1999, pp 255-265.

Nelias, D., Dumont, M.L., Champiot, F., Vincent, A., Girodin, D., Fougeres, R. and Flamand, L. Role of inclusions, surface roughness and operating conditions on rolling contact fatigue. Transactions of the ASME, Journal of Tribology, 1999; 121(2): 240-251.

Nie H. A modified rainflow counting method. International Journal of Fatigue, 1991; 13: 465-469.

Nogi, T. and Kato, T. Influence of a hard surface layer on the limit of elastic contact – part I: analysis using a real surface model. Transactions of the ASME, Journal of Tribology, 1997; 119: 493-500.

Nogi, T. and Kato, T. Influence of a hard surface layer on the limit of elastic contact – part II: analysis using a modified GW model. Transactions of the ASME, Journal of Tribology, 2002; 124: 785-793.

Olver, A.V. Gear lubrication – a review. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2002; 216: 255-267.

Pandey, R.K. and Ghosh, M.K. Thermal effects on film thickness and traction in rolling/sliding EHL line contacts - an accurate inlet zone analysis. Wear, 1996; 192:118-127.

Pandey, R.K. and Ghosh, M.K. Temperature rise due to sliding in rolling/sliding elastohydrodynamic lubricant line contacts: an efficient numerical analysis for contact zone temperatures. Tribology International, 1998; 31(12): 745-752.

Papadopoulos, I.V. Critical plane approaches in high-cycle fatigue: on the definition of the amplitude and mean value of the shear stress acting on the critical plane. Fatigue & Fracture of Engineering Materials & Structures, 1998; 21: 269-285.

Papadopoulos, I.V., Davoli, P., Gorla, C., Filippini, M. and Bernasconi, A. A comparative study of multiaxial high-cycle fatigue criteria for metals. International Journal of Fatigue, 1997; 19(3): 219-235.

Patching, M.J., Kweh, C.C., Evans, H.P. and Snidle, R.W. Conditions for scuffing failure of ground and superfinished steel disks at high sliding speeds using a gas turbine engine oil. Transactions of the ASME, Journal of Tribology, 1995; 117(3): 482-489.

Peng, W. and Bhushan, B. Three-dimensional contact analysis of layered elastic/plastic solids with rough surfaces. Wear, 2001; 249: 741-760.

Peng, W. and Bhushan, B. Sliding contact analysis of layered elastic/plastic solids with rough surfaces. Transactions of the ASME, Journal of Tribology, 2002; 124: 46-61.

Polonsky, I.A. and Keer, L.M. A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear, 1999; 231: 206-219.

Polonsky, I.A. and Keer, L.M. Fast methods for solving rough contact problems: a comparative study. Transactions of the ASME, Journal of Tribology, 2000; 122: 36-41.

Prasad, D. and Chhabra, R.P. Thermal and normal squeezing effects in lubrication of rollers by a power-law fluid. Wear, 1991; 145(1): 61-76.

Prasad, D., Singh, P. and Sinha, P. Thermal and inertia effects in hydrodynamic lubrication of rollers by a power law fluid considering cavitation. Transactions of the ASME, Journal of Tribology, 1993; 115(2):319-326.

Qiu, L. and Cheng, H.S. Temperature rise simulation of three-dimensional rough surfaces in mixed lubricated contact. Transactions of the ASME, Journal of Tribology, 1998; 120: 310-318.

Rao, J. S., Pathak, A., and Chawla, A. Blade life: a comparison by cumulative damage theories. Transactions of the ASME, Journal of Engineering for Gas Turbines and Power, 2001; 123: 886-892.

Ringsberg, J.W. Life prediction of rolling contact fatigue crack initiation. International Journal of Fatigue, 2001; 23: 575-586.

Ringsberg, J.W., Loo-Morrey, M., Josefson, B.L., Kapoor, A. and Beynon, J.H. Prediction of fatigue crack initiation for rolling contact fatigue. International Journal of Fatigue, 2000; 22: 205-215.

Rodkiewicz, C.M. and Yang, P. Non-Newtonian TEHL analysis of tilting-pad bearings subjected to inlet pressure build-up. Transactions of the ASME, Journal of Tribology, 1995; 117(3): 461-467.

Salehizadeh, H. and Saka, N. Thermal non-Newtonian elastohydrodynamic lubrication of rolling line contacts. Transactions of the ASME, Journal of Tribology, 1991; 113(3): 481-491.

Sayles, R.S. Basic principles of rough surface contact analysis using numerical methods. Tribology International, 1996; 29(8): 639-650.

Seireg, A. Thermal stress effects on the surface durability of gear teeth. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2001; 215: 973-979.

Sharif, K.J.H., Holt, C.A., Evans, H.P. and Snidle, R.W. Simplified analysis of non-Newtonian effects in a circular elastohydrodynamic contact and comparison with experiment. Tribology Transactions, 1999; 42(1): 39-45.

Sharif, K., Kong, S., Evans, H.P. and Snidle, R.W. Contact and elastohydrodynamic analysis of worm gears, part 1: theoretical formulation. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2001a; 215: 817-830.

Sharif, K., Morris, S.J., Evans, H.P. and Snidle, R.W. Comparison of non-Newtonian EHL models in high sliding applications. In: Proceedings of 27th Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 2001b, pp787-796.

Sharif, K., Evans, H.P., Snidle, R.W. and Newall, J.P. Modeling of film thickness and traction in a variable ratio traction drive rig. Transactions of the ASME, Journal of Tribology, 2004; 126: 92-104.

Shaw, B. Personal Communication. Design Unit, Newcastle University, 2003.

Society of Automotive Engineers. Fatigue Design Handbook, Second Edition. Prepared under the auspices of the Design Handbook Division of the SAE Fatigue Design and Evaluation Technical Committee; Richard C. Rice, editor, Brian N. Leis, associate editor, Drew V. Nelson, associate editor. Warrendale, Pa, 1988.

Song, B., Wang, L. and Qi, Y. Effect of two-phase liquid lubricant on fatigue life. Wear, 1996; 201: 94-98.

Sraml, M., Flasker, J. and Potrc, I. Numerical procedure for predicting the rolling contact fatigue crack initiation. International Journal of Fatigue, 2003; 25: 585-595.

Sraml, M., Flasker, J. and Potrc, I. Critical plane modeling of fatigue initiation under rolling and sliding contact. Journal of Strain Analysis for Engineering Design, 2004; 39(2): 225-236.

Stachowiak, G.W. and Batchelor, A.W. Engineering Tribology. Elsevier, Amsterdam, 1993.

Stewart, S. and Ahmed, R. Rolling contact fatigue of surface coatings – a review. Wear, 2002; 253: 1132-1144.

Sui, P.C. and Sadeghi, F. Non-Newtonian thermal elastohydrodynamic lubrication. Transactions of the ASME, Journal of Tribology, 1991; 113(2): 390-396.

Suresh, S. Fatigue of Materials, Second Edition. Cambridge University Press, 1998.

Suryanarayana, N.V. Engineering Heat Transfer. West Publishing Company, Minneapolis, 1994.

Tallian, T.E. Simplified contact fatigue life prediction model, Part I-II. Transactions of the ASME, Journal of Tribology, 1992; 114: 207-222.

Tallian, T.E. Data-fitted bearing life prediction model, Part I-III. Tribology Transactions, 1996; 39(2): 249-275.

Tallian, T.E. Data fitted bearing life prediction model for variable operating conditions. Tribology Transactions, 1999; 42: 241-249.

Tannehill, J.C., Anderson, D.A. and Pletcher, R.H. Computational Fluid Mechanics and Heat Transfer, Second Edition. Taylor & Francis, 1997.

Tao, J., Hughes, T.G., Evans, H.P. and Snidle, R.W. Elastohydrodynamic response of transverse ground gear teeth. In: Proceedings of the 28th Leeds-Lyon Symposium on Tribology, Elsevier, Amsterdam, 2002, pp447-458.

Tao, J., Hughes, T.G., Evans, H.P., Snidle, R.W., Hopkinson, N.A., Talks, M. and Starbuck, J.M. Elastohydrodynamic lubrication analysis of gear tooth surfaces from micropitting tests. Transactions of the ASME, Journal of Tribology, 2003; 125: 267-274.

Timoshenko, S. and Goodier, J.N. Theory of Elasticity, Third Edition. McGraw-Hill Book Company, 1970.

Tolstov, G.P. Fourier Series, Translated from Russian by Richard A. Silverman. Dover Publications Inc., New York, 1976.

Venner, C.H. and Lubrecht, A.A. Transient analysis of surface features in an EHL line contact in the case of sliding. Transactions of the ASME, Journal of Tribology, 1994; 116(2): 186-193.

Wang, S., Cusano, C. and Conry, T.F. Thermal analysis of elastohydrodynamic lubrication of line contacts using the Ree-Eyring fluid model. Transactions of the ASME, Journal of Tribology, 1991; 113(2): 232-244.

Wang, S., Conry, T.F. and Cusano, C. Thermal non-Newtonian elastohydrodynamic lubrication of line contacts under simple sliding conditions. Transactions of the ASME, Journal of Tribology, 1992; 114(2): 317-327.

Wang, Q., Zhu, D., Cheng, H.S., Yu, T., Jiang, X. and Liu, S. Mixed lubrication analyses by a macro-micro approach and full scale mixed EHL model. Transactions of ASME, Journal of Tribology, 2004a; 126: 81-91.

Wang, W.-Z., Liu, Y.-C., Wang, H. and Hu, Y.-Z. A computer thermal model of mixed lubrication in point contacts. Transactions of the ASME, Journal of Tribology, 2004b; 126(1): 162-170.

Wang, Y., Li, H., Tong, J. and Yang, P. Transient thermoelastohydrodynamic lubrication analysis of an involute spur gear. Tribology International, 2004c; 37: 773-782.

Webster, M.N. and Sayles, R.S. A numerical model for the elastic frictionless contact of real rough surfaces. Transactions of the ASME, Journal of Tribology, 1986; 108: 314-320.

Widmark, M. and Melander, A. Effect of material, heat treatment, grinding and shot peening on contact fatigue life of carburized steels. International Journal of Fatigue, 1999; 21: 309-327.

Wilkinson, C.M.R. and Olver, A.V. Durability of gear and disc specimens – Part I: the effect of some novel materials and surface treatments. Tribology Transactions, 1999; 42(3): 503-510.

Williams, J.A. Engineering Tribology. Oxford University Press, Oxford, 1994.

Xu, S. A Collection of C Programs for Common Algorithms (in Chinese). Tsinghua University Press, Beijing, 1994.

Xu, G. and Sadeghi, F. Thermal EHL analysis of circular contacts with measured surface roughness. Transactions of the ASME, Journal of Tribology, 1996; 118: 473-483.

Yang, P., Jin, Z.M., Liu, F. and Dowson, D. On the time-dependent, thermal and non-Newtonian elastohydrodynamic lubrication of line contacts subjected to normal and tangential vibrations. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2004; 218(2): 71-82.

Yang, P., Qu, S., Chang, Q. and Guo, F. On the theory of thermal elastohydrodynamic lubrication at high slide-roll ratios – line contact solution. Transactions of the ASME, Journal of Tribology, 2001; 123: 36-41.

Yang, P and Wen, S. A generalized Reynolds equation for non-Newtonian thermal elastohydrodynamic lubrication. Transactions of the ASME, Journal of Tribology, 1990; 112(4): 631-636.

Yang, P. and Wen, S. Behavior of non-Newtonian thermal EHL film in line contacts at dynamic loads. Transactions of the ASME, Journal of Tribology, 1992; 114(1): 81-85.

You, B.R. and Lee, S.B. A critical review on multiaxial fatigue assessments of metals. International Journal of Fatigue, 1996; 18(4): 235-244.

Yu, M.-H. and Bhushan, B. Contact analysis of three-dimensional rough surfaces under frictionless and frictional contact. Wear, 1996; 200(1-2): 265-280.

Zahavi, E. and Torbilo V. Fatigue Design, Life Expectancy of Machine Parts. CRC Press, 1996.

Zaretsky, E.V. Fatigue criterion to system design, life and reliability. Journal of Propulsion and Power, 1987; 3: 76-83.

Zaretsky, E.V. Design for life, plan for death. Machine Design, 1994; 66(15): 57-59.

Zaretsky, E.V., Poplawski, J.V. and Peters, S.M. Comparison of life theories for rolling-element bearings. Tribology Transactions, 1996; 39(2): 237-248.

Zhai, X. and Chang, L. Transient thermal model for mixed-film contact. Tribology Transactions, 2000; 43(3): 427-434.

Zhai, X. and Chang, L. Some insights into asperity temperatures in mixed-film lubrication. Tribology International, 2001; 34: 381-387.

Zhao, J., Sadeghi, F. and Hoeprich, M.H. Analysis of EHL circular contact start up: Part I - Mixed contact model with pressure and film thickness results. Transactions of the ASME, Journal of Tribology, 2001; 123(1): 67-74.

Zhu, D. Elastohydrodynamic lubrication in extended parameter ranges – part I: speed effect. Tribology Transactions, 2002a; 45(4): 540-548.

Zhu, D. Elastohydrodynamic lubrication in extended parameter ranges – part II: load effect. Tribology Transactions, 2002b; 45(4): 549-555.

Zhu, D. and Ai, X. Point contact EHL based on optically measured three-dimensional rough surfaces. Transactions of the ASME, Journal of Tribology, 1997; 119: 375-384.

Zhu, D., Wang, Q., Liu, S. and Martini, A. A transient contact stress analysis for mixed-EHL of rough surfaces, stress field animation, and a virtual texturing technology. In: Proceedings of 2003 STLE/ASME Joint International Tribology Conference, Ponte Vedra Beach, Florida USA, October 26-29, 2003. 2003-TRIB-0273, pp93-98.

Zhu, Y., Zhang, S. and Yan, M. A cycle counting method considering load sequence. International Journal of Fatigue, 1993; 15: 407-411.