U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

DEPARTAMENTO DE ENGENHARIA MECÂNICA E GESTÃO INDUSTRIAL

GEAR MICROPITTING PREDICTION USING THE DANG VAN HIGH-CYCLE FATIGUE CRITERION

JOSÉ AUGUSTO DE SOUSA FERREIRA BRANDÃO

2007

JOSÉ AUGUSTO DE SOUSA FERREIRA BRANDÃO

GEAR MICROPITTING PREDICTION USING THE DANG VAN HIGH-CYCLE FATIGUE CRITERION

A DISSERTATION SUBMITTED TO THE
FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
FOR THE DEGREE OF *MESTRE EM ENGENHARIA MECÂNICA*

Supervisor: Prof. Jorge H. O. Seabra
Co-supervisor: Prof. Manuel Jorge D. Castro

DEPARTAMENTO DE ENGENHARIA MECÂNICA E GESTÃO INDUSTRIAL
FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Abstract

The micropitting gear tooth flank damage has attracted considerable attention in recent times. This is due to the fact that, as the performance of the lubricant oils and surface coatings of the gears allows for an increase in speeds and power transmission within the gearboxes, this type of damage becomes increasingly frequent, especially in surface hardened materials with current surface finishing.

The micropitting damage consists in small craters (in the order of 10 μm in width and depth) appearing on the surface of a gear tooth. These craters are produced by the propagation of fatigue cracks that originate from the surface of the gear tooth, where a highly complex stress state is to be found because of the interaction of the roughness features of the teeth. Because the propagation of one such fatigue crack is nearly instantaneous, when compared with the initiation time, and because of the highly localized nature of the phenomenon, allied to the complex multi-axial stress history of an initiation site, the classic fatigue criteria such as the Goodman or Soderberg line are inadequate for the present use.

The aim of this work is to develop a model suitable for the analysis of tooth flank surface fatigue damage in spur gears, in particular micropitting contact fatigue initiation and mild wear. The prominent features of the model are a mixed film lubrication model and the application of the Dang Van high-cycle multi-axial fatigue criterion. By taking into account the roughness of the contacting gear teeth, the non-Newtonian behaviour of the lubricating oil, the temperature variations within the contact and the elastic and fatigue properties of the gear material, it allows for a realistic, if approximate, modelling of the physical phenomena behind the onset of micropitting.

The micropitting initiation model was used to simulate an actual micropitting test. The comparison of the test with the simulation results proved sufficiently satisfactory to warrant further work on the model.

Resumo

O dano por micropitting nos flancos das engrenagens tem atraído uma considerável atenção nos últimos tempos. Isso deve-se ao facto de que, à medida que as performances dos óleos lubrificantes e dos revestimentos superficiais das engrenagens vão permitindo aumentar as velocidades e potências transmitidas nas caixas de engrenagens, esse tipo de dano torna-se cada vez mais frequente, especialmente nas engrenagens endurecidas superficialmente e com acabamento superficial corrente.

O dano de micropitting consiste no aparecimento de pequenas crateras (na ordem de 10 μm em largura e profundidade) na superfície dos flancos de engrenagem. Essas crateras são produzidas pela propagação de fissuras de fadiga originadas na superfície do dente, onde um estado de tensão altamente complexo impera, causado pela interacção entre as rugosidades dos dentes. A propagação dessas fendas é praticamente instantânea, quando comparada com o tempo de iniciação. Em conjunto com isso, a escala minúscula do fenómeno, aliada à complexidade do historial de tensões duma zona de iniciação tornam inviável o uso de critérios de fadiga clássicos, tais como as linhas de Goodman ou Soderberg.

O objectivo deste trabalho é o desenvolvimento de um modelo adequado para a análise dos danos de fadiga nas superfícies dos flancos de engrenagens cilíndricas, com particular incidência sobre a iniciação de fadiga de contacto de micropitting e sobre o desgaste ligeiro. Os aspectos mais proeminentes do modelo são um modelo de lubrificação por filme misto e a aplicação do critério de fadiga multi-axial de elevado número de ciclos de Dang Van. Porque o modelo incorpora as rugosidades das engrenagens, a reologia não Newtoniana dos óleos de lubrificação, as variações térmicas no contacto e as propriedades elásticas e de fadiga do material de engrenagem, ele permite uma simulação realista, embora aproximada, dos fenómenos físicos que levam ao micropitting.

O modelo de iniciação de micropitting foi usado para simular um teste de micropitting. A comparação do teste com a simulação revelou-se suficientemente satisfatória para dar continuação a trabalhos adicionais no âmbito deste modelo.

Résumé

L'endommagement de micropitting sur les flancs d'engrenage a attiré dans les dernières années une attention considérable. Cela se doit au fait que, a mesure que la performance des huiles de lubrification et des revêtements de surface des flancs d'engrenage s'améliorent et permettent l'augmentation des vitesses et puissances transmises dans les boites d'engrenages, ce type de damage devient de plus en plus fréquent, spécialement dans le cas d'engrenages avec endurcissement superficiel et une finition de surface courante.

L'endommagement de micropitting consiste en l'apparition de petites cratères (de l'ordre de 10 μm en largeur et profondeur) sur les flancs des dents d'engrenage. Ces cratères son produites par la propagation de fissures de fatigue originaires de la surface du flanc, où règne un état de tension extrêmement complexe, a cause de l'interaction des rugosités des dents en contact. Parce que la propagation d'une telle fissure de fatigue est quasiment instantanée, surtout si comparée a la durée d'initiation, et parce que le phénomène est hautement local et a cause de la compexe histoire de tensions multi-axiales du site d'initiation de la fissure, les critères de fatigue classiques, tel que la ligne de Goodman ou de Soderberg, sont inadéquats a l'utilisation présente.

L'objectif de ce travail est de développer un modèle propre a l'analyse des endommagements sur la surface des flancs d'engrenage cylindrique, en particulier l'initiation de fatigue de contact de micropitting et l'usure légère. Les aspects les plus proéminents du modèle sont un modèle de contact en régime de lubrification en filme mixte et l'application du critère de fatigue de Dang Van, un critère multi-axial et de nombre élevé de cycles. Par le fait que le modèle prend en compte les rugosités des flancs des dents d'engrenage, la rhéologie non Newtonienne de l'huile de lubrification, les variations thermiques dans le contact et les propriétés élastiques et de fatigue du matériau d'engrenage, il permet une modélisation réaliste, bien qu'approximative, des phénomènes physiques qui conduisent a l'apparition du micropitting.

Ce modèle a été utilisé pour simuler un essai de micropitting. Les comparaisons entre les résultats du test et de la simulation sont suffisamment satisfaisantes pour encourager la continuation de travaux additionnels sur le modèle.

Keywords

Spur gears
Gear micropitting
Mixed film lubrication
Fatigue crack initiation
Dang Van fatigue criterion

Palavras chave

Engrenagens cilíndricas
Micropitting em engrenagens
Lubricação em filme misto
Iniciação de fendas de fadiga
Critério de fadiga de Dang Van

Mots-clés

Engrenages cylindriques
Micropitting des engrenages
Lubrication en film mixte
Initiation de fentes de fatigue
Critère de fatigue de Dang Van

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Jorge Seabra for his patient guidance and support throughout the course of this work. I would also like to express my appreciation of the guidance and careful review provided by my co-supervisor Prof. Jorge Castro.

I wish to thank my colleagues and staff at CETRIB-INEGI: Beatriz Graça, Armando Campos, Jorge Castro, Ramiro Martins and most particularly Nuno Cardoso, whose dissertation I shamelessly plundered in the elaboration of this work.

My thanks also go to FEUP and to CETRIB-INEGI for their having given me the material means to realize this work.

Last but not least, my gratitude and love go to my wife and children for the monumental forbearance they showed during the many times when this work lead me to neglect them.

To my wife, Deborah and my children, Tomás and Carolina ...

Contents

List of Figures

List of Tables

Introduction

The trend in gear design has been, for the past few decades, in an improvement of the gear materials and surface treatments. This, alongside the improvement in lubricant oils formulation has allowed for ever higher speeds and power density in gear boxes. These improvements have helped in reducing the effect of the most destructive kinds of fatigue damage. In particular, the sort of progressive, in-depth originated fatigue damages, whose most characteristic example is the spalling damage, have been greatly reduced. As a result, surface originated fatigue damage have gained importance in the determination of the life of a gear [3] [4].

These types of damage, and preponderant among them the micropitting damage, come from the propagation of cracks that initiate at the surface of the gear and progress, first inward and then outward, until a surface pit is produced [2]. A crucial characteristic of these fatigue cracks is that they have a very short propagation time [3]. Thus, in practice, the total life of a crack is equal to the initiation time. In this setting, an initiation model is very useful when dealing with such a fatigue mechanism.

Because these types of damage develop wholly within the first few tens of micrometers of the depth below the surface of a tooth, the important stress perturbations caused by the interaction of the roughness of the teeth cannot be ignored [5] [6]. In practise, this translates itself in the need for the solution of the lubrication problem in the mixed film regime, in which important surface pressure distributions occur due to the interaction of the roughness peaks of each tooth with the surface of the opposing one, be that interaction through direct metal-metal contact, be it mediated by a very thin, highly pressurized lubricating film [7].

This work concerns itself with the prediction of micropitting damage and mild wear in a spur gear. Micropitting is mainly defined by the small size of its pits, typically in the order of 10 μm in width and depth [8]. While not immediately destructive, the appearance of micropitting on the surface of a gear results in an increase in the noise level and in the geometric inaccuracy of the teeth, causing a loss in efficiency of the transmission. In extreme cases, a crack originating in a micropit may progress in depth with catastrophic consequences. Another possible scenario is that widespread micropitting so weakens the surface that a layer of material is removed at once [9].

The aim of this work is to provide a numerical model for the prediction of the initiation of micropitting fatigue cracks and mild wear. This model hinges on the solution of the contact problem between gear teeth, characterized by the mixed film regime of elastohydrodinamic lubrication, and on the application of the Dang Van

high cycle multi-axial fatigue criterion, selected for use in this work because of its suitability for the prediction of fatigue crack initiation in complex loading cases.

This dissertation is divided into six chapters, in addition to the present introductory one. The next three chapters deal with the elements of the model, the subsequent one integrates these into a cohesive numerical model, the last-but-one presents an attempt at simulating an actual micropitting test and the final chapter discusses the conclusions to draw from this work.

In Chapter 1, several aspects of the geometry of a gear tooth, the kinematics of gear meshing and the Hertzian contact mechanics are presented. Notably, the roughness is discussed regarding its meaning, measurement and several roughness parameters relevant to the present work. The equations that describe the nominal profile of a gear tooth are established, with a bias toward the discretization of the surface of a tooth. In the sequence of this, the kinematics of gear meshing are described mathematically, with particular emphasis on the notable points on the meshing line. Finally, the Hertzian contact mechanics solution of the gear meshing is presented.

In Chapter 2 an overview of lubricants and lubrication is provided. In particular, lubricant oil rheology is discussed. Subsequently, a mixed film lubrication model for rough elastohydrodinamic lubrication is presented.

In Chapter 3 an overview of the method of obtaining the stresses within a gear tooth as well as an overview of the Dang Van fatigue criterion are presented. In particular, the method by which the Dang Van criterion may be applied from the previous knowledge of the stress history of each point in the tooth is discussed. The chapter closes with a discussion of the various types of damage to be encountered when dealing with gears, with particular emphasis on the micropitting phenomenon and its relation with fatigue crack initiation.

Chapter 4 presents an overview of the numerical model—which integrates the various subjects discussed in the previous chapters—alongside with the algorithms used to solve the most difficult numerical problems: the solution of the contact at each instant, the calculation of the stresses in each point of the tooth and at each instant and, finally, the calculation of the mesoscopic stresses, a quantity necessary in order to apply the Dang Van criterion.

Chapter 5 describes a micropitting test and its simulation by the model developed here. Subsequently, a comparison of the test and simulation results is presented in order to validate the adequacy of the model for the simulation of the initiation of micropitting cracks.

Finally, Chapter 6 closes this work with a presentation of the conclusions and some proposals for future works that might extend and improve the present one.

1 Geometry, kinematics and contact mechanics of a spur gear

1.1 Introduction

The geometry of a disc or a bearing ball is so simple as to hardly need any discussion. What is more, the geometric and kinematic conditions of such mechanical elements do not change during a load cycle. Thus, the geometry of these elements need hardly be mentioned when studying them. This is far from true in the case of spur gears. Because of this an entire chapter—this one—is dedicated to the study of the geometry of the spur gear tooth flank (in which is included its roughness), the kinematics of the meshing of spur gears and its Hertzian contact mechanics.

1.2 Roughness

Every engineering surface possesses a nominal geometry, which is the mathematically ideal shape that is to be approximated to some specified degree. Because of the imprecise nature of any manufacturing process, the actual shape is always different from the ideal one. This difference is usually divided into shape imperfection—a therm that is meant here to stand both for shape deviation and waviness—and roughness. Simply put, a shape imperfection is a difference such that the macro-aspect is altered; while the roughness is a small variation around the ideal shape. The distinction between shape imperfection and roughness is largely conventional: while extreme cases are easily distinguished the boundary is blurred; so that the distinction between one and the other is dependent on the end use of the part, the overall dimensions etc. . .

This is exemplified in Figures 1.1 to 1.3. Figure 1.1 shows the roughness measurement of a gear tooth in the radial direction. It also shows the nominal shape of the tooth as a dotted line. Figure 1.2 shows the shape imperfections of the tooth and Figure 1.3 shows the roughness.

It has already been said that roughness is caused by the manufacturing processes. This is only partly true: while *initial* roughness is caused by the manufacturing processes, the roughness may vary with use because of running-in and surface damage.

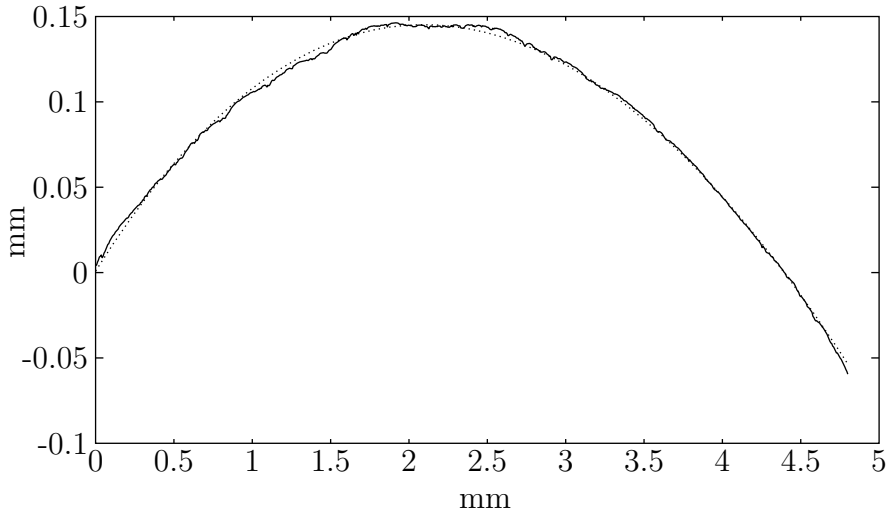


Figure 1.1: Roughness measurement of a gear tooth (full line) and its nominal shape (dotted line) in the radial direction.

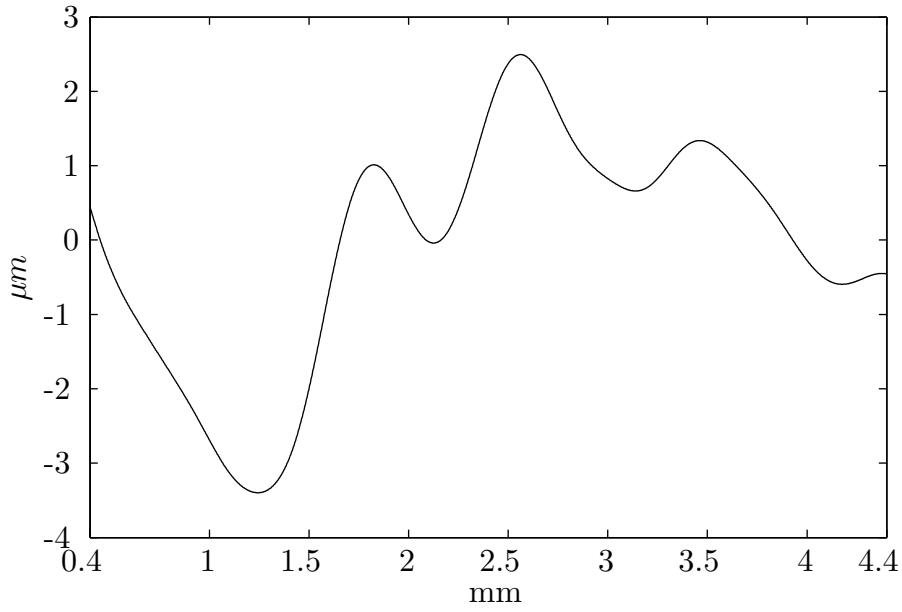


Figure 1.2: Shape imperfections of a gear tooth in the radial direction.

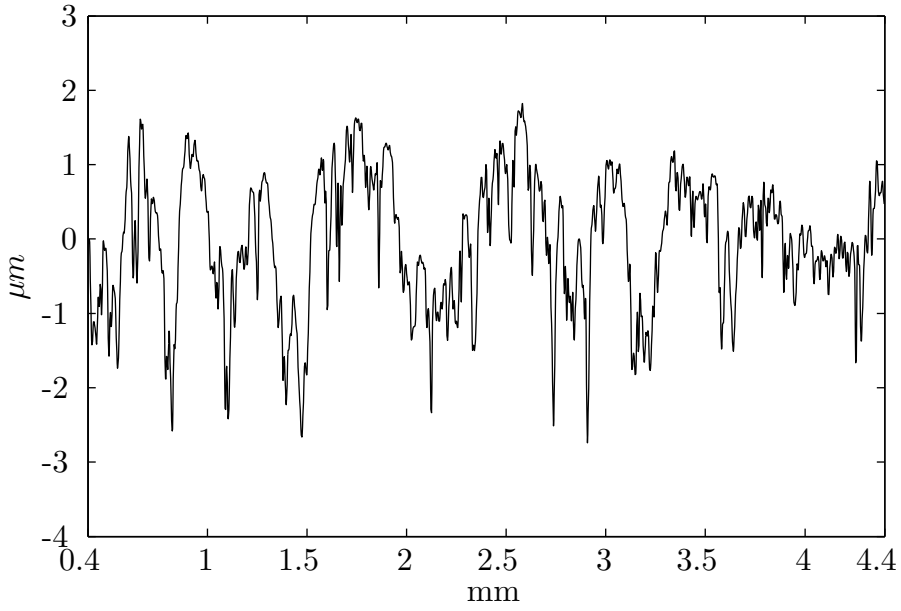


Figure 1.3: Roughness of a gear tooth in the radial direction.

1.2.1 Roughness measurement

The most usual method of roughness measurement is to slide a stylus along the surface to be measured. This stylus transmits the signal to a controller that translates these signals into coordinates and filters the data. This is the method used at the *Unidade de Tribologia, Vibrações e Manutenção Industrial* (CETRIB.) where a *Hommewerke T4000* device is used to perform the measurements. A photograph of this device is shown in Figure 1.4

1.2.2 Roughness parameters

Roughness parameters are numbers that allow the characterization of some aspect of surface roughness. All roughness parameters are specified in standards. A detailed overview of roughness measurement and roughness parameters is given in [10].

Prior to the calculation of the roughness parameters, the measured profile must be filtered in order to obtain a roughness profile like the one shown in Figure 1.3. Broadly put, the filtering process consists in decomposing the unfiltered profile in a sum of simple sine waves, whose wavelengths are compared with the cut-off length (λ_c), a parameter that differentiates between roughness on one side, and nominal shape and shape imperfections on the other. The amplitude of the waves is then reduced according to their wavelength: a wave whose wavelength is much higher than the cut-off length will be all but eliminated and one whose wavelength is much lower will be transmitted untouched.

After this, the measured length is divided into 5 segments of equal length, this length being the cut-off length.

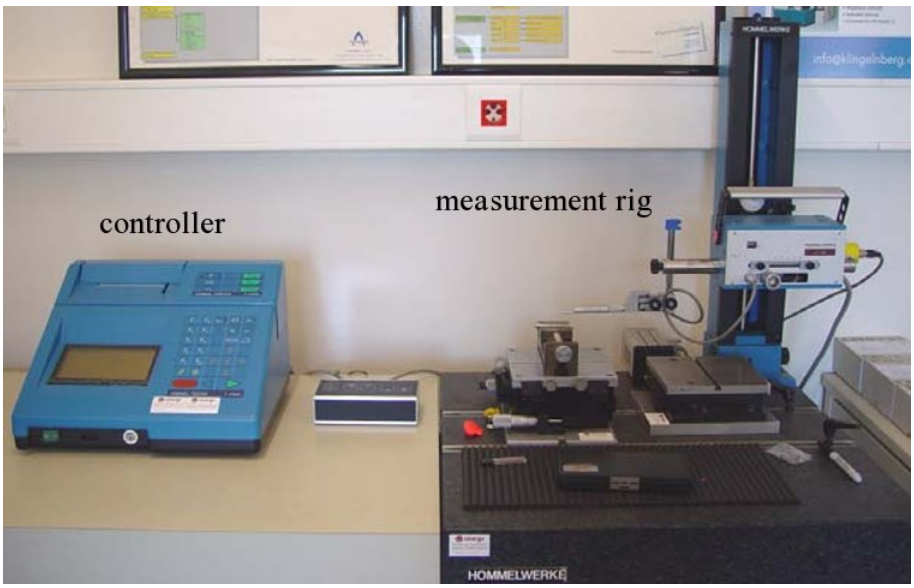


Figure 1.4: Photograph of a *Hommelwerke T4000* controller and measurement rig.

A few parameters which will be used later in this work are now described.

- **The average roughness**, R_a, is given by the equation:

$$R_a = \sqrt{\frac{1}{l_m}\int_0^{l_m} |y(x)| \; dx} \tag{1.1}$$

  where l_m is the measured length.

- **The root mean-square profile height**, R_q, is given by the equation:

$$R_q = \sqrt{\frac{1}{l_m}\int_0^{l_m} y^2(x) \; dx} \tag{1.2}$$

  where l_m is the measured length. This is none other than the standard deviation from statistics.

- **The mean peak to valley height**, $R_{Z.DIN}$, obtained with:

$$R_{Z.DIN} = \frac{1}{5}\sum_{i=1}^{5} z_i \tag{1.3}$$

  where z_i are the maximum peak to valley heights in each of the five segments of length λ_c.

  This gives a description of the amplitude of the roughness that is not too sensitive to freak events, like a single abnormally deep valley in the profile.

- **The reduced peak height**, R_{pk} and **the reduced valley depth**, R_{vk}.

  The method for obtaining these parameters, which are derived from the Abbott-Firestone curve, is rather complex and its detailed explanation is best left as supplementary reading in the work by Mummery [10].

  Their interpretation may be stated thus: the roughness may be classified in three parts. One is the core roughness, a roughness that falls within limits that encompass 40% of the material on the surface. Above this core roughness rise the highest peaks. Below the core roughness sink the lowest valleys.

  R_{pk} is the average height of the highest peaks above the core roughness and R_{vk} is the average depth of the lowest valleys below the core roughness.

  These are useful parameters because, unlike R_q and $R_{Z.Din}$, they not only represent statistical averages and maxima, but also convey some information as to the shape of the roughness profile. For instance, a roughness profile with a low R_{pk} compared to R_{vk} is in the form of a plateau.

  From the point of view of studying micropitting, they are also particularly interesting because the micropits manifest themselves as abnormally deep valleys in the roughness profile and therefore must influence the parameter R_{vk} and leave essentially unchanged the core roughness.

1.3 Profile of a gear tooth

The nominal geometry of a spur gear tooth profile can usually be described as a circle involute (see Figure 1.5). It is therefore convenient at this stage to recall the main equations that govern this geometry.

The circle involute may be likened to the trajectory of the end of a length of string unwinding around a disk. This is exemplified in Figure 1.6. Taking an arbitrary point P on the circle involute, the length $\overline{PQ}$ is the length of the unwound string and the arc length $\widehat{RQ}$ is the space previously occupied by the string on the circle. Therefore:

$$\widehat{RQ} = R_b\theta = \overline{PQ} \tag{1.4}$$

Consider a pair of versors $\vec{\imath}, \vec{k}$ respectively tangent and normal to the involute at the point P. It is found that $\vec{\imath}$ is also parallel to the radius $\overline{OQ}$, so that the position of the point P may be expressed, in relation to the centre of the base circle, as:

$$\overrightarrow{OP} = \overrightarrow{OQ} + \overrightarrow{QP} = -R_b \cdot \vec{\imath} - R_b\theta \cdot \vec{k} \tag{1.5}$$

If, as before, another arbitrary point P_0 and its associated versors $\vec{\imath}_0$ and $\vec{k}_0$ are singled out, it must of necessity obey the same law as the previous point P:

$$\overrightarrow{OP_0} = -R_b \cdot \vec{\imath}_0 - R_b\theta_0 \cdot \vec{k}_0 \tag{1.6}$$

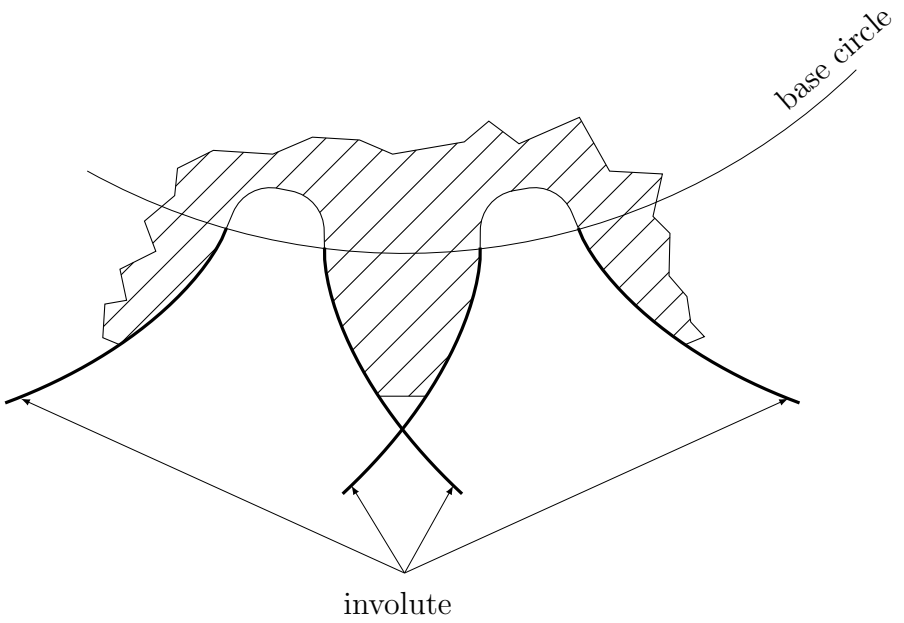


Figure 1.5: Profile of a spur gear tooth.

It would be most useful to write a parametric equation of the involute in the coordinate system defined by point P_0 and its associated versors. This is achieved by considering P_0 fixed and P moving on the curve.

It is readily observed that the pair of versors $\vec{\imath}, \vec{k}$ is nothing more than the pair $\vec{\imath}_0, \vec{k}_0$ rotated by an angle $\theta - \theta_0$:

$$\begin{aligned} \vec{\imath} &= \cos(\theta - \theta_0)\,\vec{\imath}_0 - \sin(\theta - \theta_0)\,\vec{k}_0 \\ \vec{k} &= \sin(\theta - \theta_0)\,\vec{\imath}_0 + \cos(\theta - \theta_0)\,\vec{k}_0 \end{aligned} \tag{1.7}$$

Combining Equations (1.5), (1.6) and (1.7), a parametric equation for the curve may be obtained.

$$\begin{aligned} \overrightarrow{P_0P} = R_b\,[1 - \cos(\theta - \theta_0) - \theta \sin(\theta - \theta_0)]\,\vec{\imath}_0 + \\ R_b\,[\theta + \sin(\theta - \theta_0) - \theta \cos(\theta - \theta_0)]\,\vec{k}_0 \end{aligned} \tag{1.8}$$

The coordinates x, z associated with the versors $\vec{\imath}_0, \vec{k}_0$ are derived from Equation 1.8:

$$x = R_b - R_b \cos(\theta - \theta_0) - R_b\theta \sin(\theta - \theta_0) \tag{1.9}$$

$$z = R_b\theta + R_b \sin(\theta - \theta_0) - R_b\theta \cos(\theta - \theta_0) \tag{1.10}$$

A parametric equation has now been obtained, but the parameter θ is not the most convenient one from a physical standpoint. It would be much more interesting to use the arc length $\widehat{RP}$ on the involute.

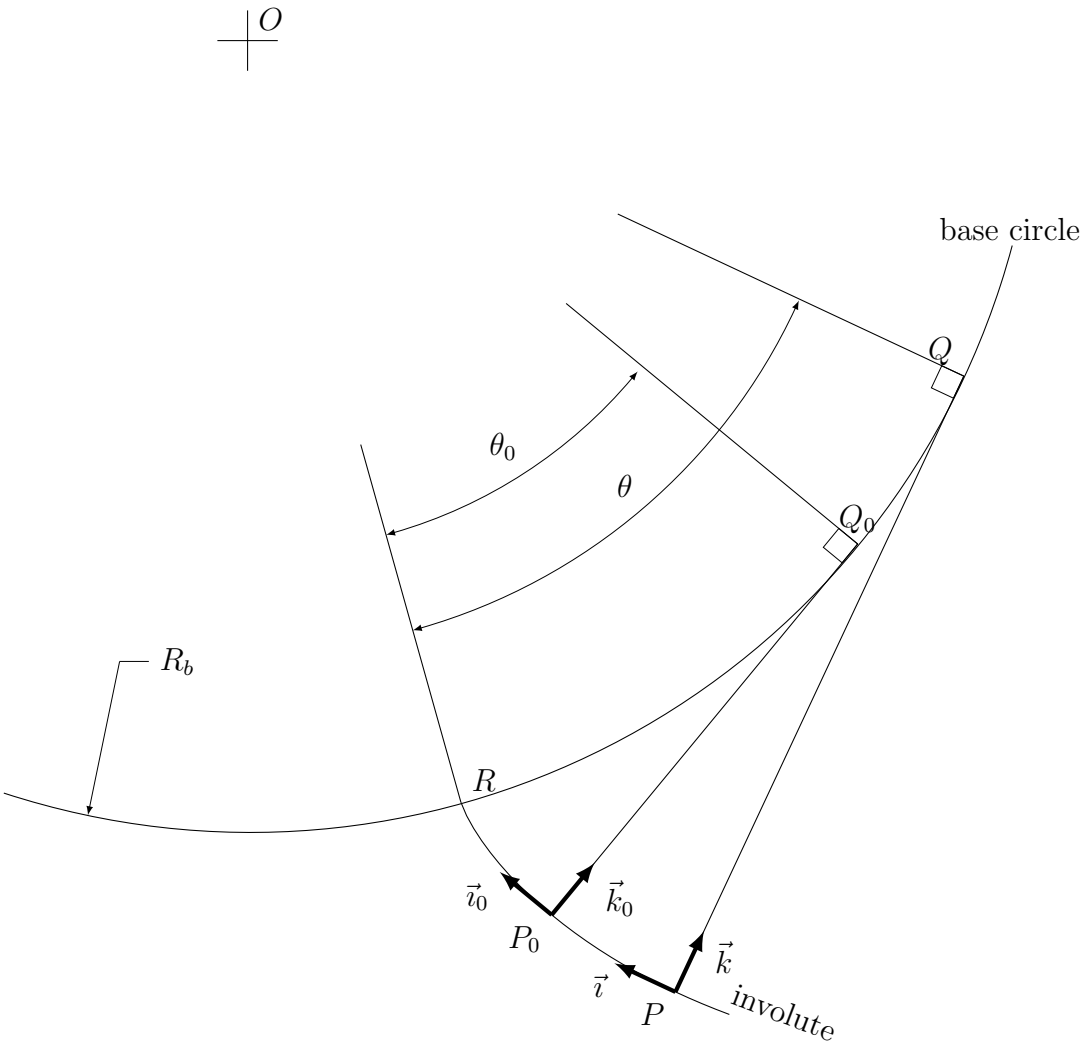


Figure 1.6: Circle involute.

As $\widehat{RP}$ may be thought of as the trajectory of a particle as it travels from point R to point P, this length may be obtained from the integration of the pseudo-velocity vector $d/d\theta\left(\overrightarrow{OP}\right)$. It is simplest to derive this vector from Equation 1.5, in which case the following is obtained:

$$\frac{d}{d\theta}\left(\overrightarrow{OP}\right) = -R_b\theta \cdot \vec{\imath} \tag{1.11}$$

In order to obtain the arc-length one needs only to integrate the norm of this pseudo-velocity vector:

$$\widehat{RP} = s\left(\theta\right) = \int_0^\theta R_b\eta \, d\eta = \frac{R_b\theta^2}{2} \tag{1.12}$$

It is now possible to rewrite the parametric equation of the x coordinate with respect to s:

$$x = R_b\left[1 - \cos\left(\sqrt{\frac{2s}{R_b}} - \sqrt{\frac{2s_0}{R_b}}\right) - \sqrt{\frac{2s}{R_b}}\sin\left(\sqrt{\frac{2s}{R_b}} - \sqrt{\frac{2s_0}{R_b}}\right)\right] \tag{1.13}$$

When trying to solve the contact problem between the meshing gear teeth, this is not a very convenient form. It is desirable to obtain s and z as functions of x.

No closed form formula exists that meets this requirement, but a Taylor expansion of Equation (1.13) at $s = s_0$ to the 3rd power yields:

$$x = s - s_0 + \frac{\left(s - s_0\right)^3}{12R_bs_0} + O\left(\left(s - s_0\right)^4\right) \tag{1.14}$$

This means that in the vicinity of point P_0, x can be taken to be $s_0 - s$ with sufficient accuracy and:

$$x = s_0 - s = R_b\frac{\theta_0^2 - \theta^2}{2} \tag{1.15}$$

$$z = R_b\left[\sin\left(\theta - \theta_0\right) + \theta\left(\cos\left(\theta - \theta_0\right)\right)\right] \tag{1.16}$$

The usefulness of this cannot be overstated for this is what allows one to use the same discretization grid on the surfaces of both meshing teeth while solving the numerical contact problem.

1.4 Mutual positioning of the gears

Figure 1.7 shows a spur gear when two pairs of teeth are meshing at once. One can identify:

a' the operating centre distance,

α' the operating pressure angle,

$\overline{T_1T_2}$ the contact line, on which the contact travels. It always remains at the same place.

$R_{b(i)}, R_{p(i)}, R_{a(i)}$ the radius of the base circle, the operating pitch circle, the addendum circle of each gear,

ω_1, ω_2 the rotational speeds of the pinion and wheel.

Not shown in the figure are Z_1 and Z_2—the number of teeth of the pinion and wheel.

Several geometric relations can be found from the examination of Figure 1.7:

$$\alpha' = \arccos \frac{R_{b1} + R_{b2}}{a'} \tag{1.17}$$

$$\overline{T_1T_2} = a' \sin \alpha' = (R_{b1} + R_{b2}) \tan \alpha' \tag{1.18}$$

$$R_{p1} = \frac{R_{b1}}{\cos \alpha'} \tag{1.19}$$

$$R_{p2} = \frac{R_{b2}}{\cos \alpha'} \tag{1.20}$$

Figure 1.8 zooms in on the meshing line, for a more detailed view of the geometry of the contacting teeth. The base pitch p_b is defined as the distance $\overline{PP'}$ between the contact points of two consecutive pairs of teeth at a given instant. From Equation 1.4 one concludes that:

$$\overline{PP'} = \widehat{R_1R_1'} = \widehat{R_2R_2'} \tag{1.21}$$

On the other hand, $\widehat{R_1R_1'}$ is the arc distance between two consecutive teeth on the pinion, so that:

$$\widehat{R_1R_1'} = R_{b1}\frac{2\pi}{Z_1} \tag{1.22}$$

Similarly, for the wheel:

$$\widehat{R_2R_2'} = R_{b2}\frac{2\pi}{Z_2} \tag{1.23}$$

Which means that the gear ratio is equal to the ratio between the base radii of the gears:

$$\frac{Z_2}{Z_1} = \frac{R_{b2}}{R_{b1}} = \frac{R_{p2}}{R_{p1}} \tag{1.24}$$

An important point that needs to be made, evident though it may be, is that to guarantee contact during the meshing, any contact point must satisfy:

$$\overline{T_1T_2} = \overline{T_1P} + \overline{T_2P} = R_{b1}\theta_1 + R_{b2}\theta_2 \tag{1.25}$$

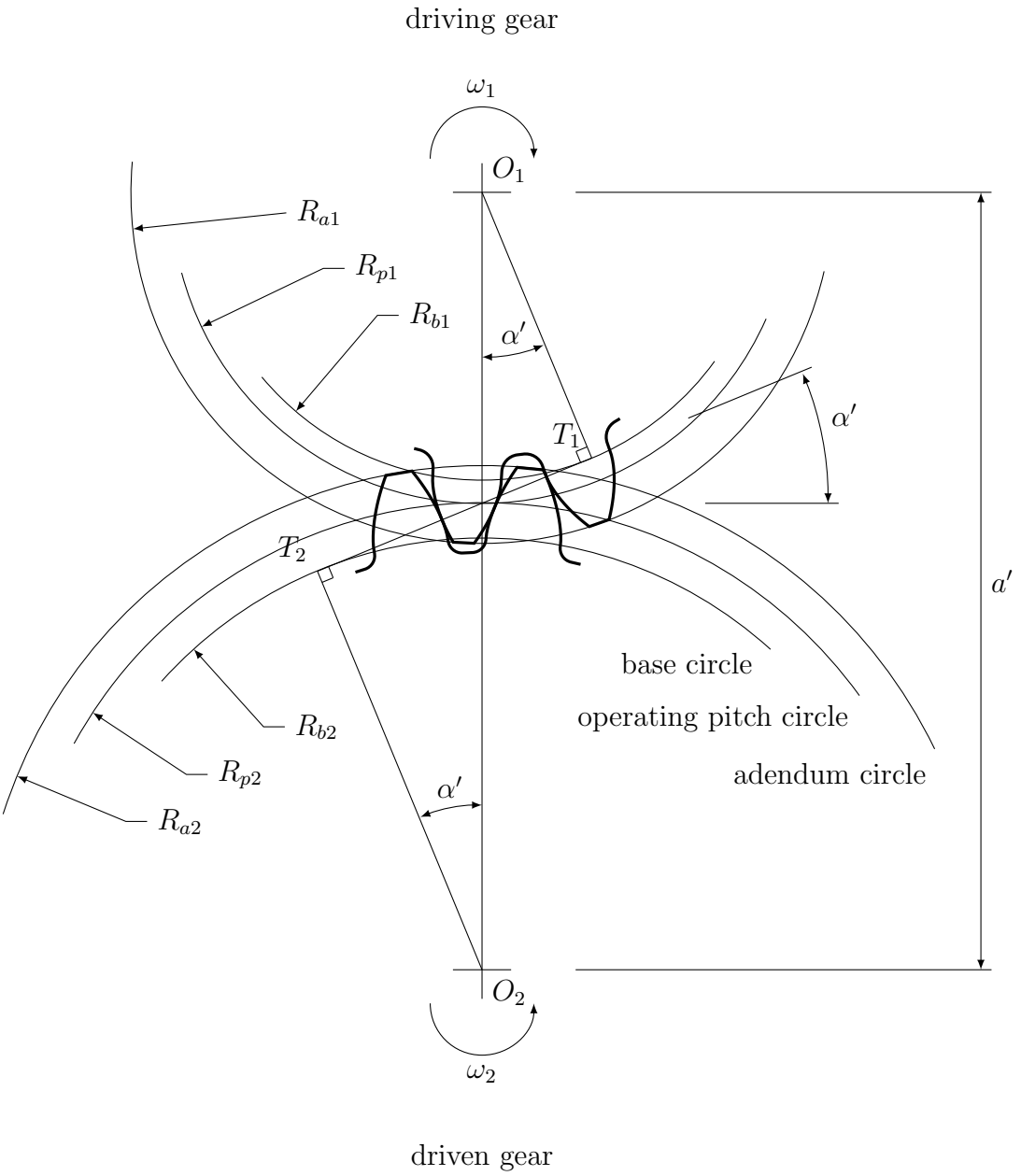


Figure 1.7: Position of the spur gears.

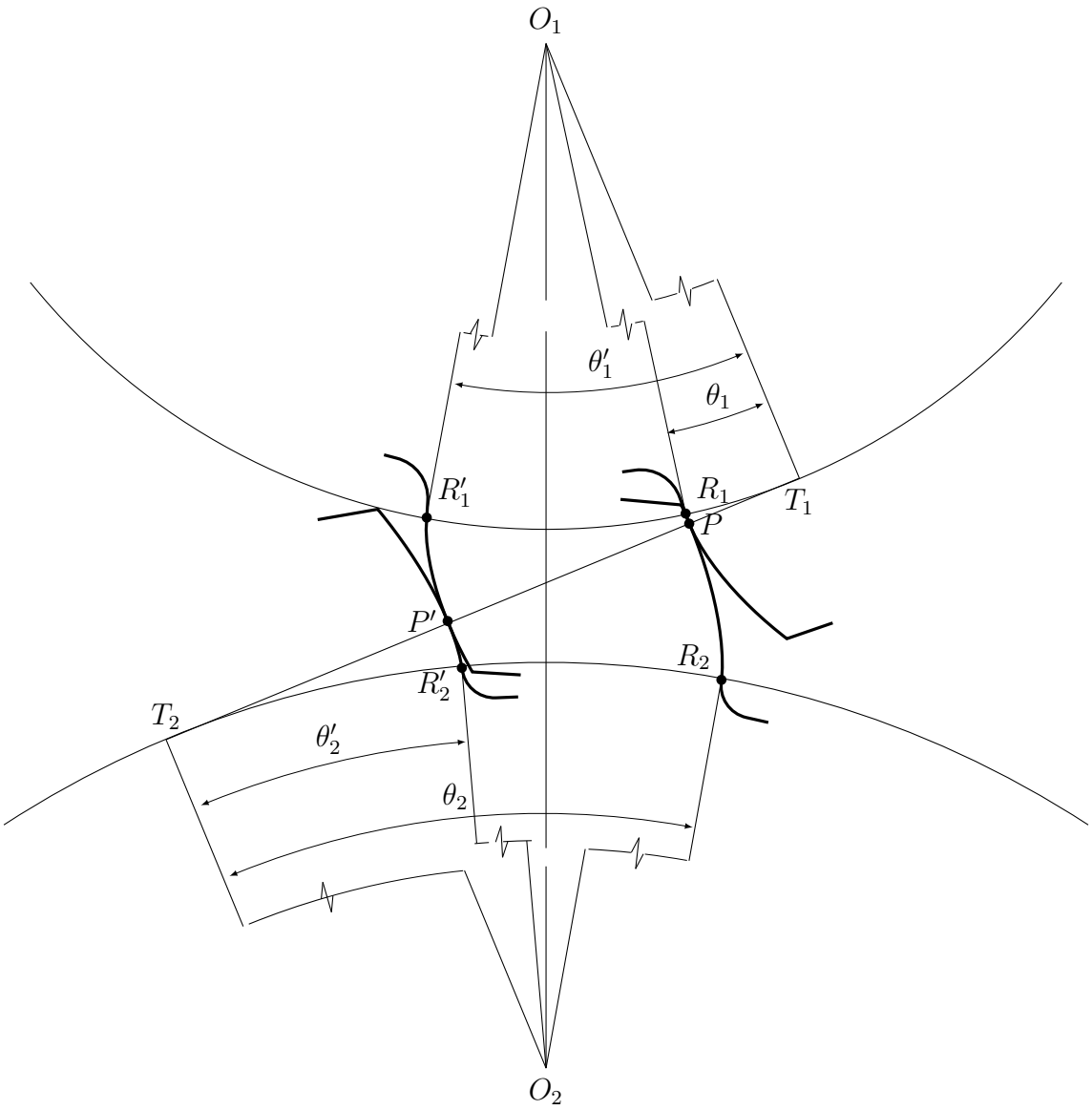


Figure 1.8: Meshing of gear teeth.

1.5 Kinematics of gear meshing

1.5.1 The contact history

During meshing, the contact points travel along the contact line but they do not actually cover all the points on $\overline{T_1T_2}$. Instead, they merely travel along a segment—$\overline{AE}$ in Figure 1.9.

To illustrate this, it is useful to observe simultaneously Figure 1.9 and 1.10 that represent the most significant instants during meshing in distinct manners: Figure 1.9 shows the superimposed positions of one pair of contacting teeth at different moments, identifying the contact points on $\overline{T_1T_2}$; while Figure 1.10 shows snapshots—in each of which the middle pair of teeth is the one represented in Figure 1.9—of the gear at each of these instants.

An explanation of the events is given in non chronological order:

Instant (a) Contact begins at A, the intersection point of the addendum circle of the wheel with the contact line.

Instant (e) Contact ends at E, the intersection point of the addendum circle of the pinion with the contact line.

Instant (b) Contact is at B. At the same time, the pair on the left ends contact, so that B is at a distance p_b from point E.

Instant (d) Contact is at D. At the same time, the pair on the right initiates contact, so that D is at a distance p_b from point A.

Instant (c) Contact is at C (pitch point), the intersection of both pitch circles with the contact line.

This means that:

$$\overline{T_2A} = \sqrt{R_{a2}^2 - R_{b2}^2} \tag{1.26}$$

$$\overline{T_1E} = \sqrt{R_{a1}^2 - R_{b1}^2} \tag{1.27}$$

$$\overline{T_1B} = \overline{T_1E} - p_b \tag{1.28}$$

$$\overline{T_2D} = \overline{T_2A} - p_b \tag{1.29}$$

$$\overline{T_1C} = R_{p1}\sin\alpha' = \frac{Z_1}{Z_1+Z_2}\overline{T_1T_2} \tag{1.30}$$

$$\overline{T_2C} = R_{p2}\sin\alpha' = \frac{Z_2}{Z_1+Z_2}\overline{T_1T_2} \tag{1.31}$$

Because the number of contacting teeth is variable in time, the load on a pair of teeth varies as well. The determination of this load is a fairly complex dynamics problem, but a fair approximation—that is used throughout this work—is shown in Figure 1.9, where F_N is the total contact load, parallel to the contact line, that is borne by the gear.

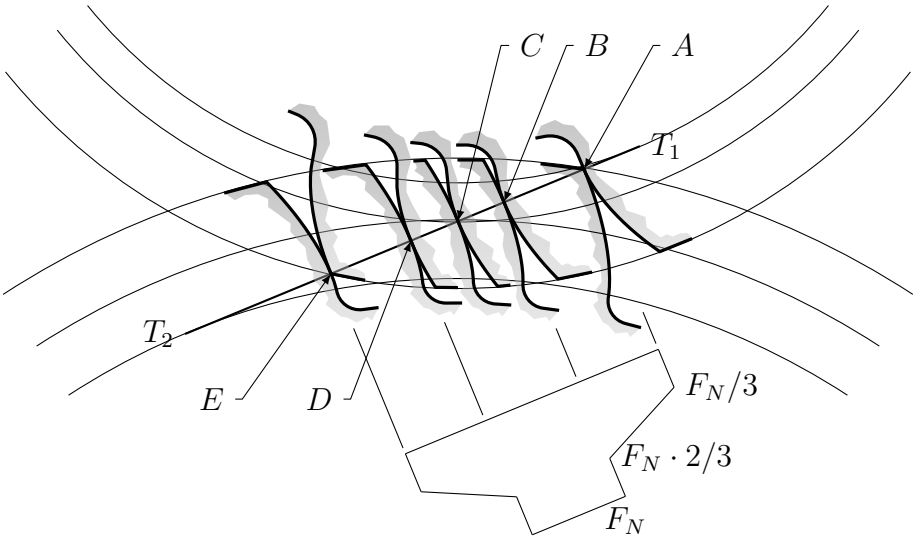
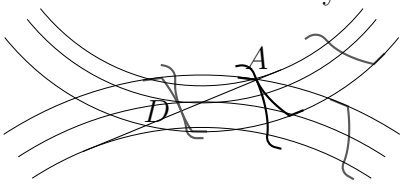
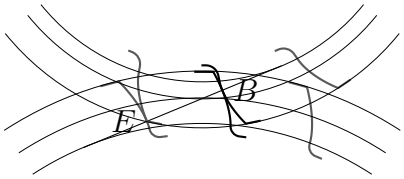


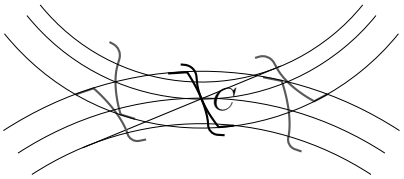
Figure 1.9: Notable moments of the meshing: the consecutive positions of a pair of contacting teeth are shown superimposed. Also shown: the share of the normal load borne by this pair of teeth.



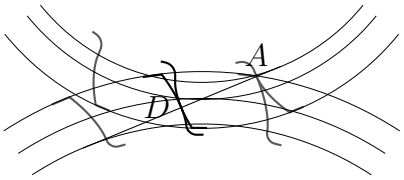
(a) the middle pair initiates contact at A



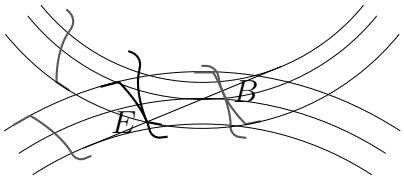
(b) the left pair ends contact at E



(c) the middle pair contacts on the pitch point C



(d) the right pair initiates contact at A



(e) the middle pair ends contact at E

Figure 1.10: Notable moments of the meshing: the middle pair of teeth is followed from the inception of contact to its end in snapshot fashion. Both the pairs immediately to the right and to the left are shown as well.

This analysis has shown what happens in a gear that, at most, admits contact between two pairs of teeth. Other situations may arise in which more, or less, teeth contact at the same time. This is determined by a parameter of the gear, the contact ratio:

$$\epsilon = \frac{\text{length of path of contact}}{\text{base pitch}} \tag{1.32}$$

In this particular case:

$$\epsilon = \frac{\overline{AE}}{p_b} = \frac{\overline{T_2A} + \overline{T_1E} - \overline{T_1T_2}}{p_b} \tag{1.33}$$

1.5.2 Velocities

To complete this kinematic analysis, the velocities of the teeth need to be discussed.

Figure 1.11 illustrates the contact of a pair of teeth at an arbitrary instant. The instantaneous velocity of the surface point of the pinion that is superimposed on the contact point is given by:

$$\begin{aligned} \vec{V_1} &= -\omega_1 \cdot \vec{\jmath} \times \overrightarrow{O_1P} \\ &= -\omega_1 \cdot \vec{\jmath} \times \left(-R_{b1} \cdot \vec{\imath} - \overline{T_1P} \cdot \vec{k}\right) \\ &= -\omega_1 \cdot R_{b1} \cdot \vec{k} + \omega_1 \cdot \overline{T_1P} \cdot \vec{\imath} \end{aligned} \tag{1.34}$$

Similarly, the instantaneous velocity of the surface point of the wheel that is superimposed on the contact point is given by:

$$\vec{V_2} = -\omega_2 \cdot R_{b2} \cdot \vec{k} + \omega_2 \cdot \overline{T_2P} \cdot \vec{\imath} \tag{1.35}$$

For the contact to be continuous, the normal velocities must be equal, so that:

$$\omega_1 R_{b1} = \omega_2 R_{b2} \tag{1.36}$$

Or:

$$\frac{\omega_1}{\omega_2} = \frac{R_{b2}}{R_{b1}} = \frac{Z_2}{Z_1} \tag{1.37}$$

From Equations (1.34) and (1.35) one can conclude that the instantaneous tangential velocities of the teeth are those of a fictitious pair of disks contacting at P, as shown in Figure 1.11, such that: one is centred on T_1, has a radius of T_1P and has an angular velocity of ω_1; the other is centred on T_2, has a radius of T_2P and has an angular velocity of ω_2. This is what allows results of tests on twin-disk machines to be extrapolated to the case of gears.

Let U_1 and U_2 be defined as the tangential velocities of the pinion and the gear. Then the sliding velocity $U_2 - U_1$—which may be interpreted as the relative velocity

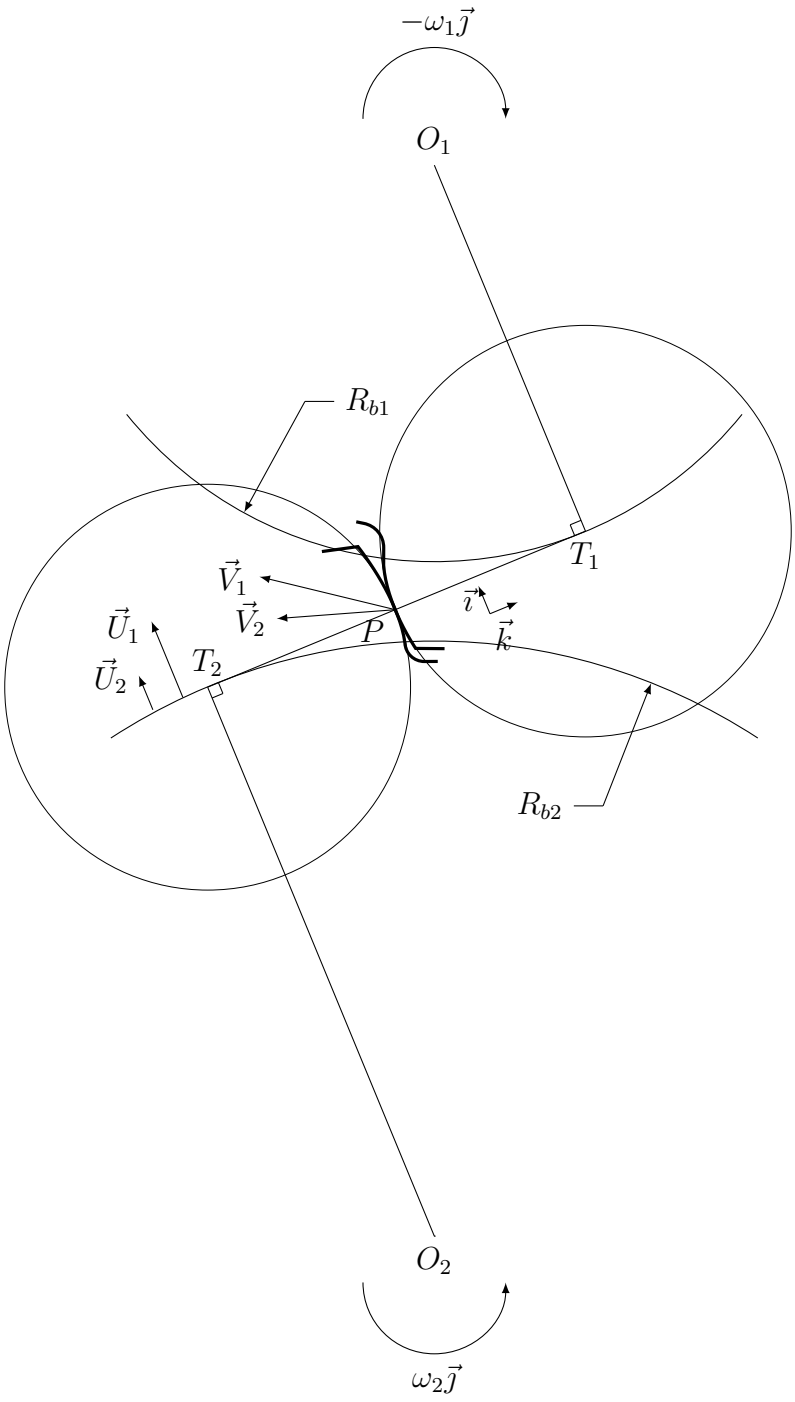


Figure 1.11: Velocities of driving and driven gear.

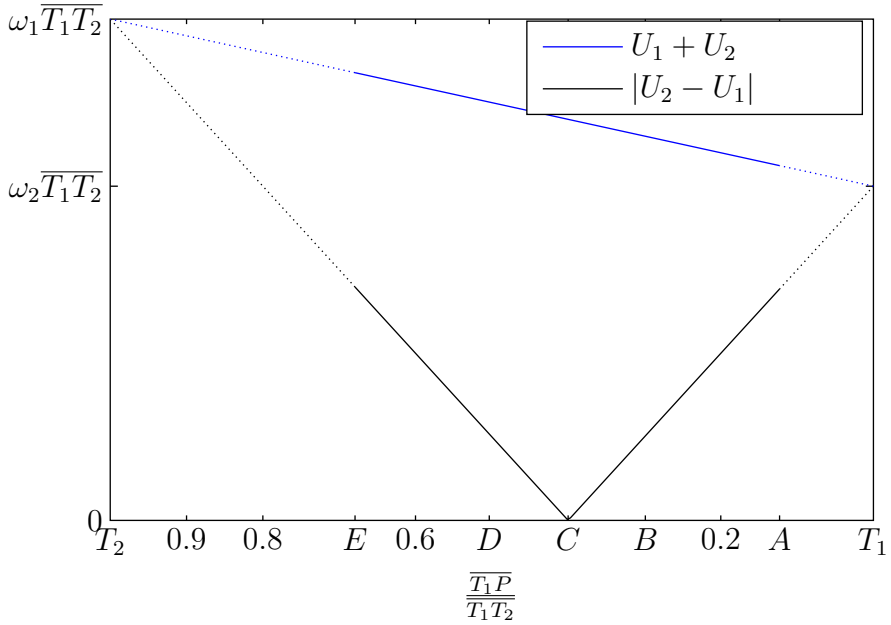


Figure 1.12: Rolling and sliding velocities along the meshing line: right of C, $U_2 > U_1$; left of C, $U_2 < U_1$.

of the movement of the wheel on the pinion—is given by:

$$\begin{aligned} U_2 - U_1 &= \omega_2 \cdot \overline{T_2P} - \omega_1 \cdot \overline{T_1P} \\ &= (\omega_1 + \omega_2)\left(\frac{\omega_2}{\omega_1 + \omega_2}\overline{T_2P} - \frac{\omega_1}{\omega_1 + \omega_2}\overline{T_1P}\right) \\ &= (\omega_1 + \omega_2)\left(\frac{\omega_2}{\omega_1 + \omega_2}\left(\overline{T_1T_2} - \overline{T_1P}\right) - \frac{\omega_1}{\omega_1 + \omega_2}\overline{T_1P}\right) \\ &= (\omega_1 + \omega_2)\left(\frac{\omega_2}{\omega_1 + \omega_2}\overline{T_1T_2} - \overline{T_1P}\right) \\ &= (\omega_1 + \omega_2)\left(\frac{Z_1}{Z_1 + Z_2}\overline{T_1T_2} - \overline{T_1P}\right) \\ &= (\omega_1 + \omega_2)\left(\overline{T_1C} - \overline{T_1P}\right) \end{aligned} \tag{1.38}$$

It is very important—for reasons that are explained in the chapter on fatigue—to determine the direction of the sliding velocity. From Equation (1.38) follows immediately that there is no sliding when the contact point is on C, that $U_2 - U_1 > 0$ when it is on the right of C and that $U_2 - U_1 < 0$ when it is on the left of C.

Figure 1.12 shows the evolution of the sliding and rolling velocities (the rolling velocity is the sum of the tangential velocities of the individual teeth $U_1 + U_2$) of a contacting pair of teeth along the meshing line.

Figure 1.13 shows the directions of the sliding velocities of each tooth on the other. Focusing on a pinion tooth, it may be said that when the contact is below the pitch circle, sliding is positive (directed towards the root of the tooth); and when the contact is above the pitch circle, the sliding is negative (pointing towards

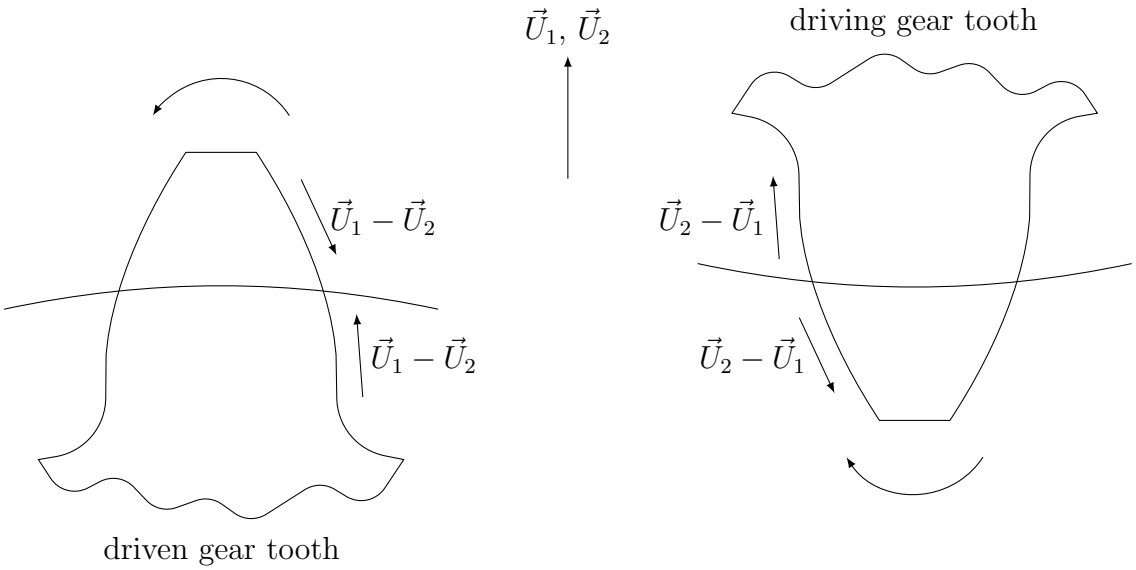


Figure 1.13: Direction of sliding (and friction force) on a tooth surface.

the tip of the tooth). As will be seen later, this fact seems to determine the direction of growth of fatigue cracks.

1.5.3 A reworking of the equations

The previous sections have described the operation of a gear in a somewhat unusual way, in that the technological aspect of gear design was disregarded. Such aspects as addendum modification, for instance, remained unmentioned. For a more mainstream and complete overview, the reader may refer to [11].

In any case, the equations hitherto obtained are here reworked to refer to more usual technological parameters: the gear ration, the numbers of teeth, the operating centre distance, the operating pressure angle, the addendum radii and the angular velocities of the gears.

Mutual positioning of the gears (see Figures 1.7 and 1.8)

gear ratio:
$$i = \frac{Z_2}{Z_1} \tag{1.39}$$

primitive radii:
$$R_{p1} = \frac{a'}{1+i} \qquad R_{p2} = \frac{ia'}{1+i} \tag{1.40}$$

base radii:
$$R_{b1} = \frac{a'}{1+i}\cos\alpha' \qquad R_{b2} = \frac{ia'}{1+i}\cos\alpha' \tag{1.41}$$

operating gear module: $$m' = \frac{2R_{p1}}{Z_1} = \frac{2R_{p2}}{Z_2} = \frac{2a'}{Z_1 + Z_2} \tag{1.42}$$

base pitch: $$p_b = \pi m' \cos\alpha' = \frac{2\pi a' \cos\alpha'}{Z_1 + Z_2} \tag{1.43}$$

meshing line: $$\overline{T_1T_2} = a' \sin\alpha' \tag{1.44}$$

The contact history (see Figure 1.9 and 1.10)

$$\overline{T_2A} = \sqrt{R_{a2}^2 - \left(\frac{ia' \cos\alpha'}{1+i}\right)^2} \tag{1.45}$$

$$\overline{T_1E} = \sqrt{R_{a1}^2 - \left(\frac{a' \cos\alpha'}{1+i}\right)^2} \tag{1.46}$$

$$\overline{T_1B} = \sqrt{R_{a1}^2 - \left(\frac{a' \cos\alpha'}{1+i}\right)^2} - \frac{2\pi a' \cos\alpha'}{Z_1 + Z_2} \tag{1.47}$$

$$\overline{T_2D} = \sqrt{R_{a2}^2 - \left(\frac{ia' \cos\alpha'}{1+i}\right)^2} - \frac{2\pi a' \cos\alpha'}{Z_1 + Z_2} \tag{1.48}$$

$$\overline{T_1C} = \frac{a' \sin\alpha'}{1+i} = \frac{Z_1 \tan\alpha'}{2\pi} p_b \tag{1.49}$$

$$\overline{T_2C} = \frac{ia' \sin\alpha'}{1+i} = \frac{Z_2 \tan\alpha'}{2\pi} p_b \tag{1.50}$$

$$\epsilon_1 = \frac{\overline{AC}}{p_b} = \frac{Z_2}{2\pi}\left(\sqrt{\left(\frac{R_{a2}}{R_{b2}}\right)^2 - 1} - \tan\alpha'\right) \tag{1.51}$$

$$\epsilon_2 = \frac{\overline{CE}}{p_b} = \frac{Z_1}{2\pi}\left(\sqrt{\left(\frac{R_{a1}}{R_{b1}}\right)^2 - 1} - \tan\alpha'\right) \tag{1.52}$$

$$\epsilon = \epsilon_1 + \epsilon_2 \tag{1.53}$$

Velocities (see Figures 1.11 and 1.13)

$$U_1 = (\omega_1 + \omega_2)\frac{i}{1+i}\overline{T_1P} \tag{1.54}$$

$$U_2 = (\omega_1 + \omega_2)\left(\frac{a' \sin\alpha'}{1+i} - \frac{\overline{T_1P}}{1+i}\right) \tag{1.55}$$

$$U_1 + U_2 = (\omega_1 + \omega_2)\left(\frac{a' \sin\alpha'}{1+i} + \overline{T_1P}\frac{i-1}{1+i}\right) \tag{1.56}$$

$$U_2 - U_1 = (\omega_1 + \omega_2)\left(\frac{a' \sin\alpha'}{1+i} - \overline{T_1P}\right) \tag{1.57}$$

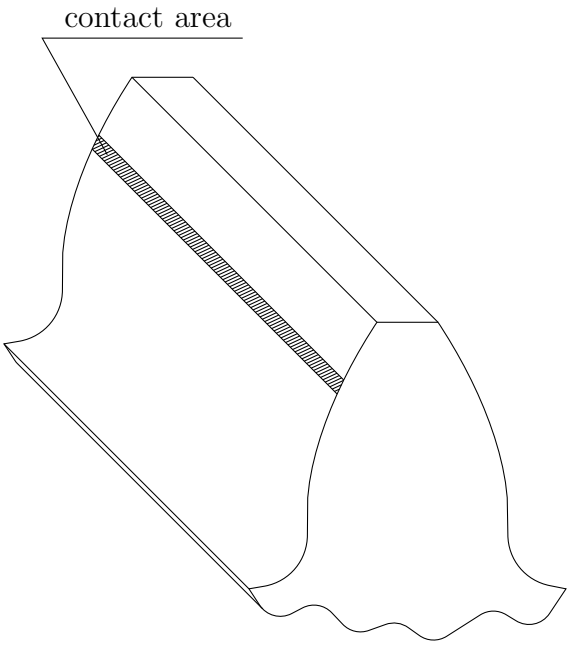


Figure 1.14: Approximate shape of the contact area.

1.6 Contact mechanics of spur gear flanks

In the previous sections, the meshing of the teeth has been studied as if they were rigid bodies: the deformation of the teeth—due either to bending or contact stresses—has been held to be null. From a kinematic viewpoint, the error thus introduced is negligible.

However, this is an unacceptable premise when dealing with the contact stresses between the teeth. While the stresses due to bending are not within the scope of this work— their effect is only felt very far from the surface—the deformation caused by contact stresses must be taken into account.

During meshing, the contacting area of the teeth is a very narrow band that spans the full width of the teeth. It is a good approximation to consider that this band is rectangular (see Figure 1.14); and because it is so narrow, it's deviation from an osculating plane is very slight, when compared to the radii of curvature of the teeth.

Under such conditions, the theory of the elastic half-plane for a plane strain state—laid out in such works as [12] and [13, Section 138–141]—may be safely used. Indeed, the use of this theory requires that the surface slopes be very small. With gear teeth, this is generally the case even when their roughness is taken into account.

This section introduces the Hertzian solution to the contact between teeth. This is by no means a detailed treatment of the subject. The interested reader may complement this overview with the reading of [12].

One of the postulates of the Hertzian theory is that the surfaces in contact are

devoid of roughness. This approximation is valid when studying the effects of the contact in the sub-surface, but must be abandoned when the conditions on the surface must be ascertained. This is precisely the case of the study of micropitting. Nevertheless, this theory is a good starting point to understand some of the factors that influence micropitting. The solution obtained with this theory—while not valid in this case—may also be seen as a standard against which more complex models must be compared.

1.6.1 Equivalent elastic cylinder/rigid plane problem

Looking at Figure 1.11, it is seen that the radius of curvature of the driving gear is $\overline{T_1P}$ and that of the driven gear is $\overline{T_2P} = \overline{T_1T_2} - \overline{T_1P}$.

Mathematically, and within the framework of the Hertzian contact, this is the same problem as that of an elastic cylinder pressed against a rigid surface. This equivalent cylinder has a radius of curvature equal to:

$$\begin{aligned} r_{eq} &= 2\left(r_1^{-1} + r_2^{-1}\right)^{-1} \\ &= 2\xi\left(1-\xi\right)\cdot\overline{T_1T_2} \end{aligned} \tag{1.58}$$

where:

r_1 is the radius of curvature of the driving gear

r_2 is the radius of curvature of the driven gear

r_{eq} is the radius of curvature of the equivalent cylinder

ξ is: $\frac{\overline{T_1P}}{\overline{T_1T_2}}$

The equivalent radius of curvature of a gear can be seen plotted against the position of the contact on the meshing line in Figure 1.15. It shows that although the equivalent radius of curvature given in Equation (1.58) is a parabola and symmetric in relation to the midpoint of the line segment T_1T_2, this symmetry is broken by the eccentricity of the position of the actual meshing line.

The equivalent Young's modulus of the equivalent cylinder is given by:

$$E_{eq} = \left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)^{-1} \tag{1.59}$$

where:

E_{eq} is the equivalent Young's modulus

E_1 is the Young's modulus of the driving gear

ν_1 is the Poisson's ratio of the driving gear

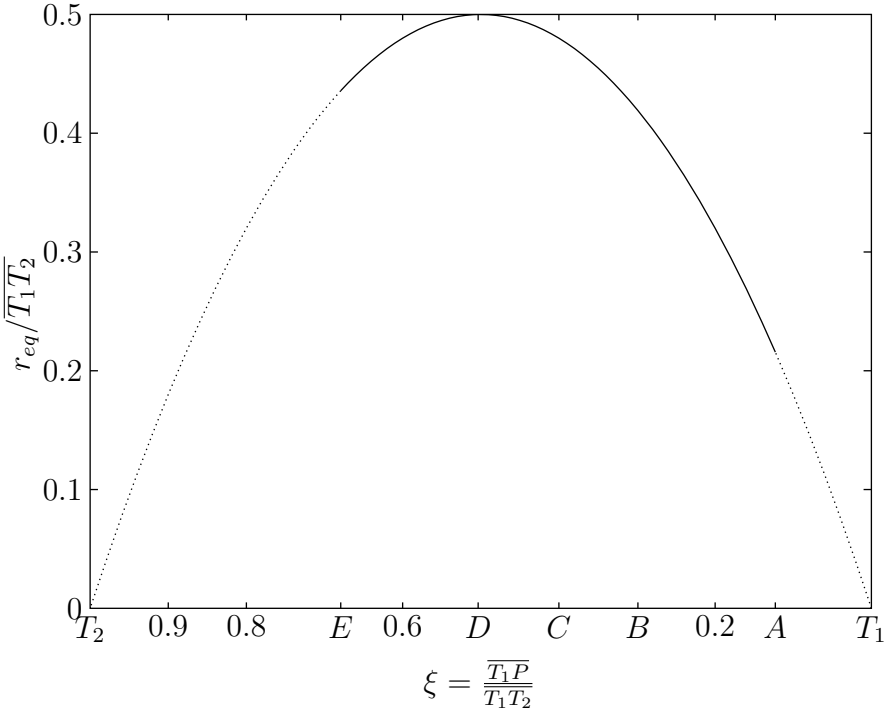


Figure 1.15: Dimensionless plot of the equivalent radius of curvature vs. the position on the meshing line.

E_2 is the Young's modulus of the driven gear

ν_2 is the Poisson's ratio of the driven gear

It is important to note that these equivalent properties can only be used to determine the contact deformation and stresses. The bulk stresses must be calculated with the actual elastic properties.

1.6.2 Hertzian stress and half-width of contact

According to the Hertzian theory the solution of the contact problem is as shown in Figure 1.16. The pressure distribution follows the equation:

$$p = \begin{cases} p_0\sqrt{1-\left(\frac{x}{a}\right)^2} & ,|x| < a \\ 0 & ,|x| > a \end{cases} \tag{1.60}$$

Thus, the variables of the problem are:

a the Hertzian contact half-width

p_0 the maximum Hertzian pressure

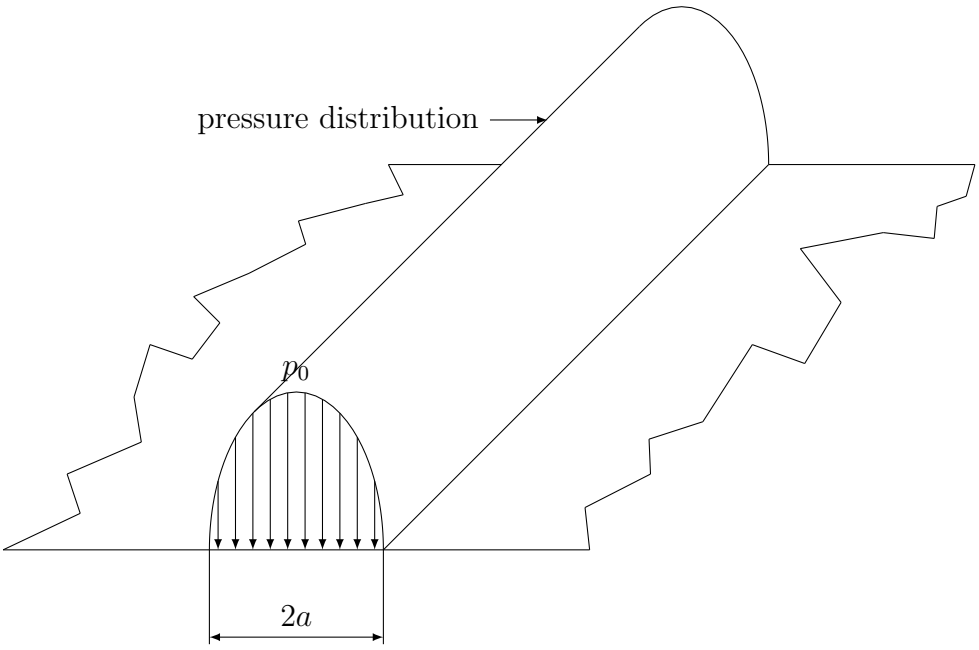


Figure 1.16: Hertzian solution to the contact problem: a is the contact half-width; p_0 is the maximum contact pressure.

These variables are given by the equations:

$$a = \sqrt{\frac{2}{\pi} f_N \frac{r_{eq}}{E_{eq}}} \tag{1.61}$$

$$p_0 = \sqrt{\frac{2}{\pi} f_N \frac{E_{eq}}{r_{eq}}} \tag{1.62}$$

where f_N is the actual contact load per unit of length between one single pair of teeth (see Figure 1.9). This is calculated by:

$$f_N = \begin{cases} 0 & \overline{T_1P} > \overline{T_1E} \text{ or } \overline{T_1A} > \overline{T_1P} \\ \dfrac{F_N}{b}\dfrac{1}{3}\left(1 + \dfrac{\overline{AP}}{\overline{AB}}\right) & \overline{T_1B} > \overline{T_1P} > \overline{T_1A} \\ \dfrac{F_N}{b} & \overline{T_1D} > \overline{T_1P} > \overline{T_1B} \\ \dfrac{F_N}{b}\dfrac{1}{3}\left(1 + \dfrac{\overline{EP}}{\overline{DE}}\right) & \overline{T_1E} > \overline{T_1P} > \overline{T_1D} \end{cases} \tag{1.63}$$

Figures 1.17 and 1.18 show the evolution of these variables along the meshing line. The dotted lines represent the values of these variables if f_N were kept constant at the key values of $F_N/(3b)$, $(2/3)(F_N/b)$, F_N/b. As was the case with the graph of the equivalent radius of curvature, these lines are perfectly symmetric in the interval T_1T_2. It's the placement of the actual meshing line that introduces the asymmetry in the values of a and p_0.

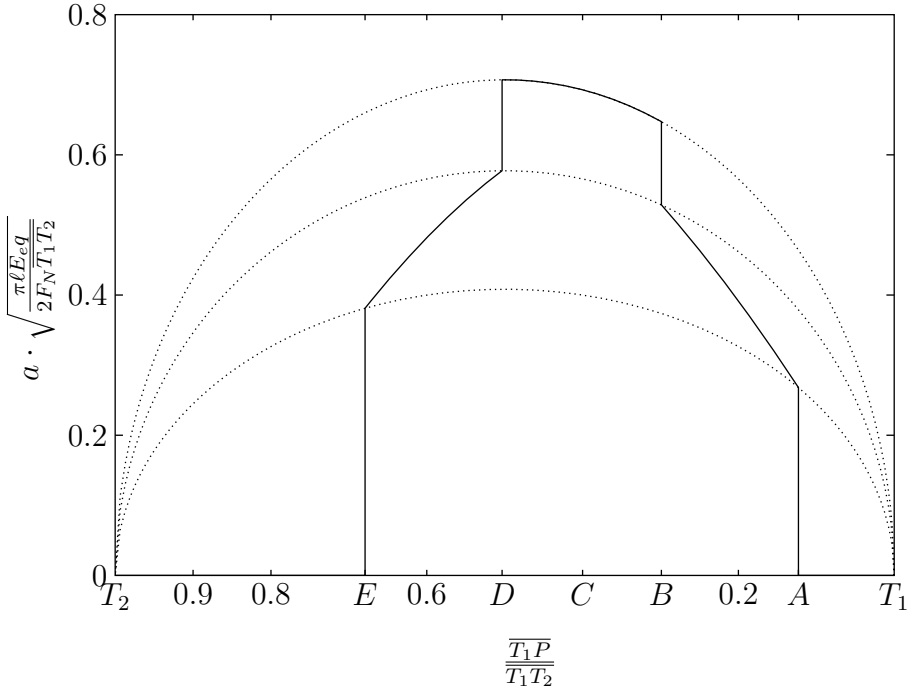


Figure 1.17: Dimensionless plot of the contact half-width vs. the position on the meshing line.

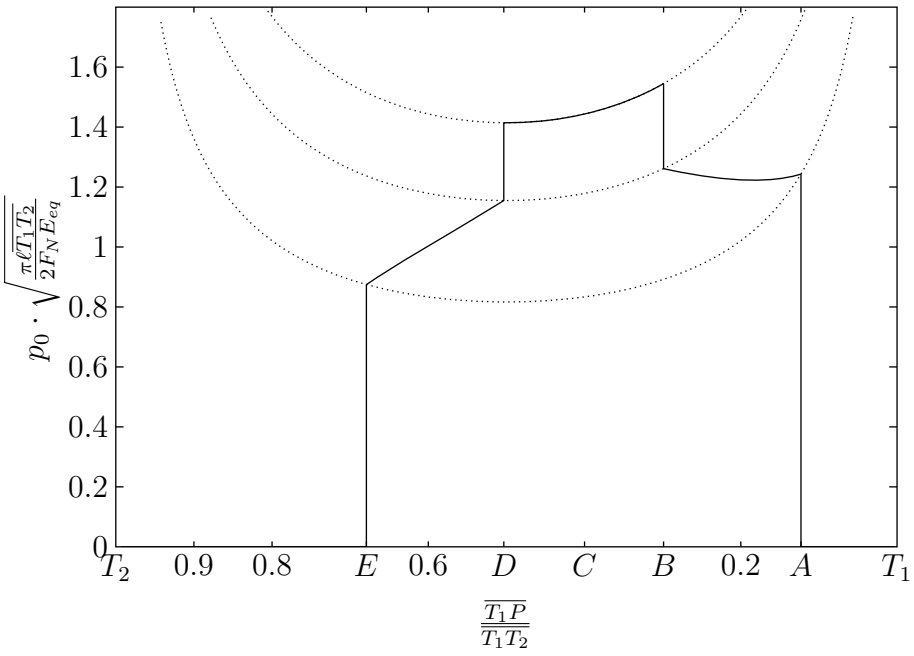


Figure 1.18: Dimensionless plot of the Hertzian stress vs. the position on the meshing line.

2 Lubricant oils and lubrication

2.1 Introduction

Lubricants play a crucial part in the contact of solids. They have mainly three functions: to separate the contacting surfaces, to reduce friction and to evacuate heat from the contact area. Unfortunately, these are often contradictory requirements so that a choice of lubricant is always a careful exercise in compromise.

Lubricants may roughly be classified according to their physical state in normal operating conditions: lubricant oils, lubricant greases, solid lubricants and gas lubricants. Of these, the first type is by far the one most used with gears. It is also the only one with which this work concerns itself.

In what follows, most aspects of the study of a lubricant oil but the ones that directly concern the simulation of contact will be glossed over. The interested reader may complement this overview by reading the course notes by Seabra et al. [14].

2.2 Lubricant oil classification according to origin

A lubricant oil may be classified according to its origin. Thus these may be distinguished:

vegetal and animal oils Historically, these were the very first oils used by Man for lubrication. They have mainly fallen into disuse, since their performance is not comparable to those of other types of oils. As an example, one can mention sperm-whale oil and colza oil.

mineral oils These are obtained from the distillation of petroleum. They can roughly be separated into parafinic based oils, naftenic based oils and aromatic based oils, according to their base molecular structure. Because of their low cost and reasonable performance, their use is very widespread.

synthetic oils The most recent addition to the panoply of available oils, these are synthesized (by opposition to the distillation process used for mineral oils) from the constituting molecules of either hydrocarbons or vegetal oils. This is a fine-tuned process that produces oils of surpassingly good performance but of correspondingly high costs. They may be classified in three major groups: the PAO, polyglicols and ester oils. Of these the ester oils merit

special mention as they have been the subject of intense research: the mounting environmental concerns of the public have made one of their properties, biodegradability, a very desirable characteristic.

In broad lines, those are the available classes of lubricant oils. But those are only the base lubricants, the reader is not to suppose that the story ends here. Indeed, from the 1920's on, lubricant vendors have been mixing additives with the base oils. In fact a base lubricant's behaviour can be significantly altered by the admixture of additives. There are additives to improve viscosity, to protect surfaces from wear, to prevent oxidation etc... So much so that a lubricant oil can almost be customized. Unfortunately there is no way to predict accurately the influence of mixing several additives, since they mutually affect themselves as well as the base lubricant. The desirable properties must be obtained by trial and error.

2.3 Lubricant oil rheology

In order to model numerically the behaviour of a lubricant oil, its constitutive relations—the rheology—must be known. To achieve this, relevant properties must be identified and introduced into a set of equations that relate shear stresses to rates of shear deformation.

2.3.1 Viscosity

The viscosity is the property of a fluid to oppose internal sliding at low velocities. The proviso of low velocities is an important one, as will be later shown.

Consider Figure 2.1, that represents the velocity field v of a fluid film between two plates—whose ideally infinite dimensions in the zz axis ensures a plane strain state—when the top plate is moving at the velocity V and the bottom plate is fixed.

When V is very small—nearly negligible—so that the flow is laminar, the shear stress σ_{xy} and the shear strain rate $\dot{\gamma}_{xy}$—the derivative of the shear strain in time are linearly related:

$$\sigma_{xy} = \eta\dot{\gamma}_{xy} = \eta\frac{dv}{dy} \tag{2.1}$$

and η is called the dynamic viscosity of the fluid.

Throughout this work, only one shear strain component and one shear stress component need to be considered for the lubricant oil film, and thus, for simplicity's sake, their indices will be dropped and they will be called γ and τ from now on. Thus, the Equation (2.2) becomes:

$$\tau = \eta\dot{\gamma} = \eta\frac{dv}{dy} \tag{2.2}$$

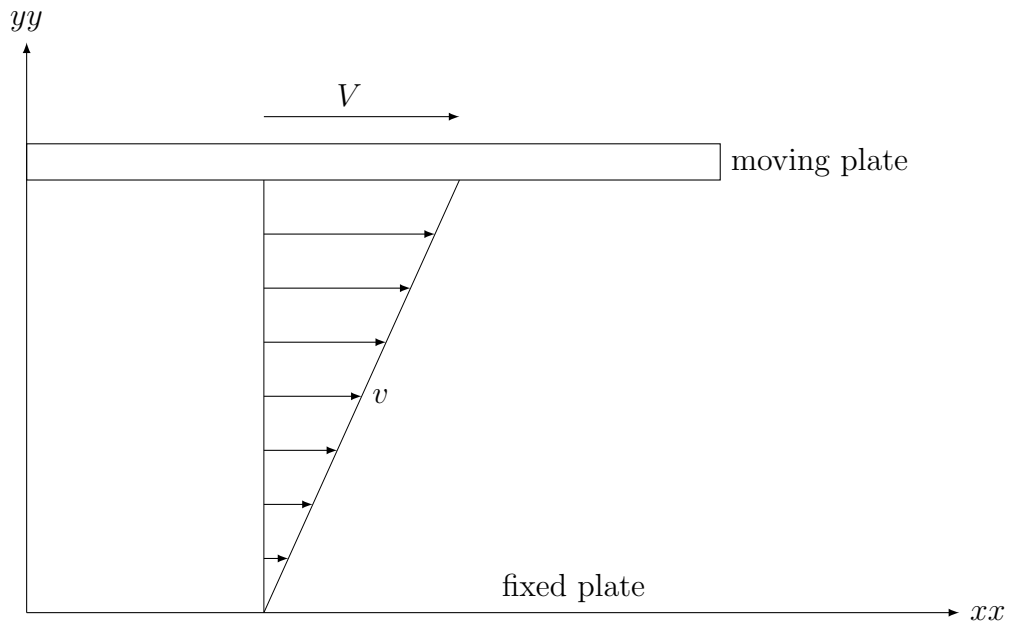


Figure 2.1: Laminar flow of a fluid between two plates.

Equation (2.2) describes the Newtonian law of fluids. It is always true for very small velocities but there exist fluids—called Newtonian fluids—for which the limitation of small velocity does not apply. An example of such fluids is water.

Naturally the viscosity is only a constant as long as pressure and temperature conditions remain unaltered. Under a variation of either or both, viscosity changes dramatically.

Thermoviscosity

Thermoviscosity is the change of the viscosity under varying temperatures. It is important to note that an increase in temperature always has a thinning effect: the viscosity diminishes with temperature.

Several laws that model this dependence have been developed. As an aside, one must bear in mind that these are valid only at constant pressure.

The simplest one, proposed by Cameron [14], is an exponential law:

$$\eta = \eta_0 \exp\big(-\beta\,(\theta - \theta_0)\big) \tag{2.3}$$

with:

θ temperature of the fluid;

θ_0 reference temperature of the fluid;

β thermoviscosity index of the fluid;

η_0 dynamic viscosity of the fluid at reference temperature θ_0;

η dynamic viscosity of the fluid at temperature θ.

This law is accurate only for a short range of temperatures around θ_0. In fact, it can be viewed as a Taylor expansion to the first degree of the logarithm of the true viscosity function.

Another—more accurate—law is that proposed by Vogel [14]:

$$\eta = K \exp \frac{b}{\theta + c} \tag{2.4}$$

with:

η dynamic viscosity of the fluid at temperature θ;

θ temperature of the fluid in [°C];

K lubricant constant, in the same units as the viscosity η;

b, c lubricant constants, in [°C].

One must be careful to note that this equation is not independent of the temperature unit used, so that all temperatures must be converted to [°C] before introducing them into the equation.

Another expression, and perhaps the most important one due to its widespread use in industry, is that specified in the ASTM D341 standard:

$$\ln \ln (\nu + a) = n - m \ln T \tag{2.5}$$

where:

ν is the kinematic viscosity of the lubricant defined as:

$$\nu = \frac{\text{dynamic viscosity}}{\text{density}} \tag{2.6}$$

T temperature of the lubricant oil in [K];

m, n, a lubricant constants.

Piezoviscosity

Piezoviscosity is the change in viscosity with varying pressure. Unlike the effect of temperature, an increase in pressure is always followed by an increase in viscosity.

The simplest law that describes this, that proposed by Barus [14] is stated thus:

$$\eta = \eta_0 \exp (\alpha p) \tag{2.7}$$

with:

p pressure of the fluid,

α piezoviscosity coefficient of the fluid,

η_0 dynamic viscosity of the fluid at atmospheric pressure,

η dynamic viscosity of the fluid at pressure p.

Thermo-piezo-viscosity

Naturally, in real applications, both temperature and pressure change at the same time. This leads to the necessity of obtaining equations that integrate both thermal and pressure effects.

This can be effected by combining Equations (2.3) and (2.7):

$$\eta = \eta_0 \exp\big(\alpha p - \beta\,(\theta - \theta_0)\big) \tag{2.8}$$

This equation has the virtue of simplicity, with its usual attendant defect, namely a low accuracy in its prediction for realistic ranges of variation.

A more interesting law is that proposed by Roelands [14]:

$$\begin{aligned} \eta &= \eta_0 \exp\left(\alpha^* p\right) \\ \alpha^* p &= (\ln \eta_0 + 9.67)\left\{\left(\frac{\theta - 138}{\theta_0 - 138}\right)^{-S_0} \cdot \left(1 + 5.1 \times 10^{-9} p\right)^Z - 1\right\} \end{aligned} \tag{2.9}$$

with:

θ temperature of the fluid in [K],

θ_0 reference temperature of the fluid in [K],

p pressure of the fluid in [Pa],

η_0 dynamic viscosity of the fluid at reference temperature θ_0 and atmospheric pressure,

η dynamic viscosity of the fluid at temperature θ in [Pa·s].

As an entertaining side note, the Roelands equation can be reworked into a more aesthetically pleasing shape, with the following equation as an end result:

$$\eta = \eta_R \cdot \exp\left(R \cdot \left(\frac{p}{p_R} + 1\right)^Z \cdot \left(\frac{\theta}{\theta_R} - 1\right)^{-S_0}\right) \tag{2.10}$$

where $\eta_R = 6.31 \times 10^{-5}$ Pa · s, $\theta_R = 138$ K, $p_R = 196$ GPa are constants of the Roelands viscosity law (independent of the fluid) and R is a non-dimensional constant of the fluid, to be determined either from the constants of Equation (2.9) or from viscosity measurements. All other parameters retain the same meaning. Note that in this new form, the equation is dimensionally consistent and no reference conditions are needed.

2.3.2 Non-Newtonian rheology

The majority of lubricant oils behave in a non-Newtonian fashion under normal operating conditions. This is easily understood if one thinks that a typical lubricant, at the rates of deformation and extremely high pressures present during the meshing of gears, and even when taking into account the thinning effect of the temperature rise caused by the dissipation of the shear forces, would need to sustain a friction coefficient in the order of 5 in order to maintain a Newtonian behaviour. This is clearly at odds with the observed evidence that points to a friction coefficient of, at most, 0.1 in lubricated contacts. This leads to alternate descriptions of the rheology of lubricant oils: non-Newtonian rheologies.

Non-linear viscous behaviour

The non-Newtonian rheology can result from an increasing non-linearity of the relation between shear stress and shear strain rate. This model, presented by Ree-Eyring [15] is expressed mathematically as:

$$\dot{\gamma} = \frac{\tau_r}{\eta} \sinh \frac{\tau}{\tau_r} \tag{2.11}$$

where τ_r, the Ree-Eyring shear stress, is the shear stress that marks the limit between a linear—i.e. Newtonian—relation between shear stress and rate of deformation.

In this model, the shear stress increases without bounds with the shear strain rate, albeit at a much slower pace than is the case with a Newtonian rheology.

Elastic behaviour

In order for a lubricant oil to manifest its viscosity, loads must be applied to it in a sufficiently long and constant manner. In the event of a shock solicitation, the loads rise so sharply and so fast that the molecular reorganization necessary for the viscous response does not have time to take place, so that the only response available to the fluid is to behave as if it were an elastic solid. In mathematical terms:

$$\tau = G\gamma \tag{2.12}$$

or, taking the derivatives in time:

$$\dot{\tau} = G\dot{\gamma} \tag{2.13}$$

where G is the elastic shear modulus of the lubricant oil.

The higher the viscosity of the fluid, the more time is needed for the viscous response to enfold, and the more it resembles a solid. In fact, in conditions of extremely high pressure, the lubricant oil may even become an amorphous solid (like glass).

This means that within high pressure contacts—like meshing gears—some measure of elastic behaviour must exist alongside the viscous one.

Plastic behaviour

This reasoning can be pursued further if one thinks that solids cannot sustain an indefinitely growing elastic deformation: yielding occurs at some point. In the same manner, any lubricating oil must have a limiting shear stress (τ_L) beyond which shear deformation occurs without any increase in shear stress.

$$\tau = \min\left(\eta\dot{\gamma}, \tau_L\right) \tag{2.14}$$

Composite behaviour

Naturally, the behaviours previously listed may occur all at once with varying degrees of severity, so that an equation must be found that integrates these various behaviours. Unfortunately, at present no such equation is to be found. There has been no success in unifying the non-linear viscosity of Ree-Eyring with the elastic and plastic behaviour.

Bair and Winer have presented a model that integrates elasto-visco-plasticity:

$$\dot{\gamma} = \frac{\dot{\tau}}{G} - \frac{\tau_L}{\eta}\ln\left(1 - \frac{\tau}{\tau_L}\right) \tag{2.15}$$

This was selected for use in this work because it seems intuitively more physically correct to assume that the shear stresses are bounded by an upper limit. A simplification is introduced in this equation, as used in this work: in the absence of sliding, the elastic part of Equation (2.15) is important and cannot be neglected but, when sliding is important, as is the case with gear meshing, the visco-plastic component becomes so preponderant that the elastic part can be neglected entirely. This yields the following equation:

$$\dot{\gamma} = -\frac{\tau_L}{\eta}\ln\left(1 - \frac{\tau}{\tau_L}\right) \tag{2.16}$$

An additional reason for this elimination is that while the visco-plastic portion of the shear stress is bounded by the limiting shear stress, the elastic part is not, which negates the principle of a limiting shear stress.

This introduces an additional rheological parameter, the limiting shear stress (τ_L), that is dependent on pressure and temperature. This dependence may be stated, according to Houpert [15], as:

$$\tau_L = \tau_{L_0}\exp\left[\alpha_{\tau_L}p + \beta_{\tau_L}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{2.17}$$

where α_{τ_L} and β_{τ_L} are constants of the lubricant oil that are independent of pressure and temperature.

2.4 Elastohydrodynamic lubrication of spur gears

Elastodydrodynamic lubrication (EHL) is characteristic of the contact between two elastically deformable bodies more or less separated by a lubricant film. The proviso of "more or less" is added because, frequently, the film thickness is insufficient to ensure that a contact between opposing roughness features does not happen.

The behaviour of such a contact is modelled by a differential equation, the Reynolds equation which, in its one-dimensional form, states:

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{\eta}\frac{\partial p}{\partial x}\right) = 12\frac{U_1 + U_2}{2}\frac{\partial\left(\rho h\right)}{\partial x} \tag{2.18}$$

given here, for simplicity's sake, for a Newtonian fluid in steady state, and an integral equation, the equation of the elastic deformations of the surfaces:

$$h\left(x\right) = h_0\left(x\right) - \frac{2}{\pi E_{eq}}\int_{-\infty}^{+\infty} \ln\left|\frac{x - x'}{L}\right| p\left(x'\right)\ dx' \tag{2.19}$$

where $h\left(x\right)$ is the separation of the surfaces, $h_0\left(x\right)$ is the undeformed separation of the surfaces and L is a constant of integration.

To Equation (2.19) must be added the limiting conditions to which the pressure and separation must obey, the equality to the normal load:

$$f_N = \int_{-\infty}^{+\infty} p\left(x\right)\ dx \tag{2.20}$$

(where f_N, introduce in the previous chapter, is the normal contact load per unit face-width between one pair of teeth at a given instant) and the non penetration rule:

$$\forall x: \quad h\left(x\right) \geq 0 \tag{2.21}$$

A thorough treatment of the Reynolds equation is given by Hamrock in [16]. A detailed study of EHL is presented by Seabra et al. in [14].

To these must be added a thermal equation that accounts for the conservation of energy [14]:

$$\frac{\partial^2 T_f}{\partial z^2} + \frac{1}{K_f}\eta\left(\frac{\partial U}{\partial z}\right)^2 = \frac{\rho_f C_{pf}}{K_f} U \frac{\partial T_f}{\partial x} \tag{2.22}$$

where, T_f is the temperature field in the lubricant, ρ_f its density, C_{pf} its specific heat capacity, K_f its thermal conductivity and U its tangent velocity field.

Historically, much of the efforts of tribologists have been directed towards the solution of the EHL problem of two smooth surfaces—surfaces with no roughness—and only in recent years has the problem of EHL for rough surfaces—in which case the lubricant film may not be sufficiently thick to prevent direct contact between the surfaces—been more seriously studied.

The reader is referred to the works by Campos [17] and Sottomayor [18] for the solution of the smooth EHL contact problem considering a non Newtonian rheology and the thermal effects on the contact.

The EHL contact is generally somewhere between two extreme cases: the ideally smooth EHL contact, in which the roughness of the surfaces is negligible; and the boundary lubrication regime, in which the roughness is such that there is no longer a film that sustains any portion of the load between the surfaces. Somewhere in the middle, a real contact partakes of the characteristics of both.

2.4.1 Ideally smooth elastohydrodynamic lubrication

In this work, no attempt is made to solve the equations that govern the EHL contact between smooth surfaces. Instead, the solution of Grubin [19, chapter 6] is used.

Figure 2.2 shows the pressure distribution of an ideal full-film EHL contact comparing it with that of a Hertzian contact in the same geometric and kinematic conditions.

The figure shows the different zones that can be distinguished: the inlet zone, where the pressures are negligible, the contact—or high pressure—zone, where the lion share of the load is borne and pressures and temperatures can attain values in the order of 1 GPa and 200 °C, and the outlet zone, where the pressure and temperature drop again to the values of the inlet zone.

It shows clearly that the differences in the pressure distribution of the two cases are trifling and that the Hertzian pressure distribution is a very good approximation to that of an EHL contact within the core contact zone. There is nevertheless one striking difference between the two pressure distributions: immediately before the start of the outlet zone, a pressure spike occurs in the EHL case that corresponds to a constriction of the film thickness in the same place.

This is the basis of the Grubin solution, where it is advocated that the pressure distribution is that of the Hertzian case and the film thickness is constant within the contact zone, once more as in the Hertzian case.

Based on these hypotheses, the film thickness may be determined by:

$$h_{00} = 0.975 \left(\alpha \eta_0 \left(U_1 + U_2\right)\right)^{\frac{8}{11}} \left(r_{eq}\right)^{\frac{4}{11}} \left(\frac{E_{eq}}{f_N}\right)^{\frac{1}{11}} \tag{2.23}$$

where:

h_{00} is the film thickness at the centre of the contact,

η_0 is the viscosity of the lubricant oil in the inlet zone of the contact (inlet temperature and atmospheric pressure),

α is the piezo-viscosity coeffictient of the lubricant oil in the inlet zone given by:

$$\alpha = \frac{\partial}{\partial p} \left(\ln \frac{\eta\left(p, T\right)}{\eta_0} \right) \Bigg|_{p=p_{\mathrm{atm}}, T=T_{inlet}} \tag{2.24}$$

This is the same as Dawson and Higginson's formula [19].

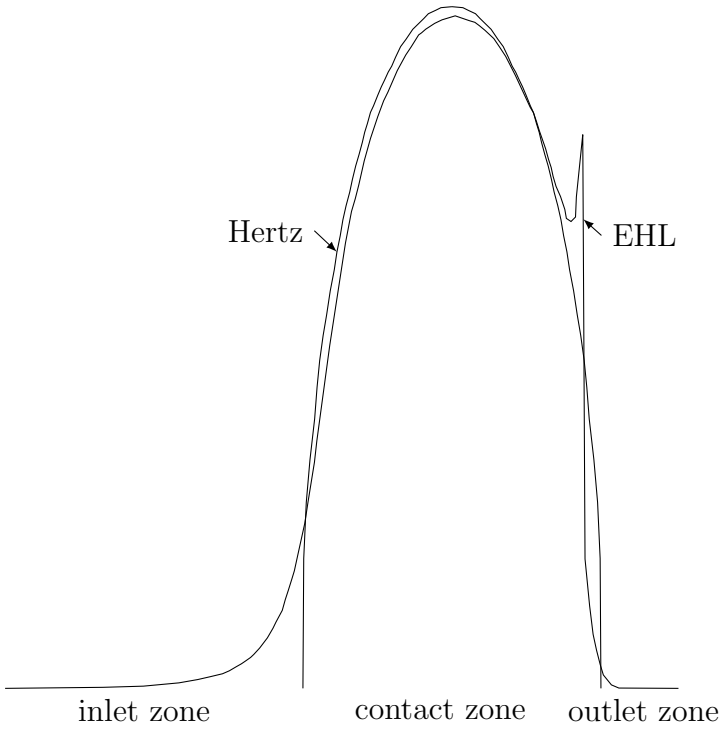


Figure 2.2: Comparison of the EHL pressure distribution with that of the Hertzian solution.

Sottomayor [18, Section 2.2.3] advocates the correction of this formula to account for the rise in temperature within the inlet zone due to the friction forces, by the coefficient:

$$\phi_T = \left[1 + 0.1\left(\frac{\beta\eta_0\left(U_1+U_2\right)^2}{K_f}\right)^{0.64}\left(1 + 14.8\left|\frac{U_1-U_2}{U_1+U_2}\right|^{0.83}\right)\right]^{-1} \tag{2.25}$$

where the thermoviscosity coefficient β is given by:

$$\beta = \frac{\partial}{\partial T}\left(\ln\frac{\eta\left(p,T\right)}{\eta_0}\right)\bigg|_{p=p_{\mathrm{atm}},T=T_{inlet}} \tag{2.26}$$

and K_f is the thermal conductivity of the lubricant oil and U_1 and U_2 are the velocities of the surface of the driving and driven gear tooth respectively.

Thus, the corrected central film thickness is:

$$h_{00,T} = h_{00}\phi_T \tag{2.27}$$

To determine the friction forces within the contact, the thermal equations must be considered. The thermal solution of the EHL contact presented is the one developed by Tevaarvek in 1980 and described in [14].

The results relevant to the present work are:

$$\Delta T_s^{\max} = \frac{1}{\sqrt{2\pi}} \frac{\mu f_N}{\mathrm{PT}\sqrt{a}} |U_1 - U_2| \left(\frac{1}{\sqrt{U_1}} + \frac{1}{\sqrt{U_2}} \right) \tag{2.28}$$

$$\Delta T_f^{\max} = \frac{1}{\beta} \ln \left(\frac{\beta (U_1 + U_2)^2 \eta}{8 K_f} + 1 \right) \tag{2.29}$$

$$T_f^{\mathrm{avg}} = T_0 + \frac{4}{\pi} \left(\Delta T_s^{\max} + \Delta T_f^{\max} \right) \tag{2.30}$$

where:

$\Delta T_s^{\mathbf{max}}$ is the greatest rise of the temperature of the surfaces above that of the inlet,

$\Delta T_f^{\mathbf{max}}$ is the greatest rise of the temperature of the lubricant above that of the surfaces,

$T_f^{\mathbf{avg}}$ is the average lubricant temperature in the contact,

μ is the average friction coefficient within the contact,

PT is the thermal coefficient of both surfaces (in the present work, they are always made of the same material), defined as:

$$\mathrm{PT} = \sqrt{\rho_s C_s K_s} \tag{2.31}$$

where:

ρ_s is the volumic mass of the surfaces,
C_s is the specific heat of the surfaces,
K_s is the thermal conductivity of the surfaces.

a is the Hertzian half-width of the contact,

η is the viscosity of the lubricant at the average surface temperature and pressure conditions within the contact,

β is the thermoviscosity coefficient of the lubricant at the average surface temperature and pressure conditions within the contact.

It is thus seen that the average lubricant temperature T_f^{avg} and the average friction coefficient μ are dependent upon each other in the equation and are not known *a priori*. This deadlock may be broken by the simultaneous consideration of the rheological equation—in the case of the present work, Equation (2.16).

Consider first that the friction coefficient between the gear teeth is:

$$\mu = \frac{\overline{\tau} \cdot 2ab}{f_n b} \tag{2.32}$$

where $\overline{\tau}$ is the average shear stress within the lubricant oil.

Thus, a manipulation of Equation (2.16) yields:

$$\mu = \frac{2a\tau_L}{f_N}\left[1 - \exp\left(-\frac{\eta\dot{\gamma}}{\tau_L}\right)\right] \tag{2.33}$$

where η is determined at the average lubricant temperature T_f^{avg} and the average Hertzian pressure $(\pi/4)\, p_0$.

Conjointly with Equation (2.30), this is enough to determine the friction coefficient within the contact, albeit in an approximate manner, using an iterative scheme of computation.

A more sophisticated scheme could be adopted where temperature and friction coefficient would be computed not as averages but at each point in the contact, but this was deemed an unnecessary complication that would yield doubtful improvements on the scheme presented here.

2.4.2 Boundary film lubrication

As has already been pointed out, the other extreme in the possible behaviour of the lubricated contact is one in which the lubricant does not play any part in bearing the pressure between the surfaces.

This happens when the roughness of the surfaces in contact is so great, when compared to the theoretical smooth EHL film thickness, or when the speed is so small that the conditions for generating a lubricating film are not met. This is the boundary film regime of lubrication, which will be referred to as boundary film lubrication from now on.

The mathematical description of this contact problem is indistinguishable from that of the dry contact problem as regards the determination of the pressure distribution. It is obtained by discarding completely the Reynolds equation and retaining only the equation of the elastic deformation of the surfaces in contact (Equation (2.19)) and its limiting conditions (Equations (2.20)–(2.21)), recalled here:

$$h(x) = h_0(x) - \frac{2}{\pi E_{eq}} \int_{-\infty}^{+\infty} \ln\left|\frac{x - x'}{L}\right| p(x')\ dx' \tag{2.34}$$

$$f_N = \int_{-\infty}^{+\infty} p(x)\ dx \tag{2.35}$$

$$\forall x: \quad h(x) \geq 0 \ \wedge\ p(x) \geq 0 \tag{2.36}$$

Note that $h_0(x)$ includes both the overall geometric shape of the surfaces and their roughness in their initial, undeformed, state.

There is no analytical solution for this problem, only numerical ones. For this reason the exposition of the solution of this problem is deferred until a later section of this work.

This lubrication regime has until now been treated as if the lubricant had no effect. While it is true in the case of the pressure distribution, it is very much otherwise when it comes to determining the surface shear stress (the "pressure"

tangent to the surface). It has been found that, in boundary lubrication, although the lubricant film fails to form, the surfaces are nevertheless coated with lubricant molecules that ease the mutual sliding of the surfaces. Furthermore, it has been found that the resultant friction coefficient μ^{BDR} is constant over variations of geometry, load and velocities. It is typically to be found in the range of values 0.08–0.15 [20].

Thus the boundary coefficient of friction is essentially a property of the lubricant oil and of the roughness of the surfaces and must be determined experimentally.

2.4.3 Mixed film lubrication

Though no surface is ever perfectly smooth, in practice an identical response is shown by pairs of surfaces whose combined roughness is sufficiently small when compared to the lubricant film thickness. From this stems the notion of the specific film thickness, Λ. This specific film thickness is given by the equation:

$$\Lambda = \frac{\text{film thickness of the ideal EHL contact}}{\text{composite } R_q \text{ of the surfaces}} \tag{2.37}$$

The specific film thickness is the parameter used to classify a contact as full-film EHL or otherwise. Thus, and according to Vergne [21], the contacts are classified as:

$\Lambda > 3$ full-film EHL lubrication,

$3 > \Lambda > 2$ nearly full-film EHL with occasional contact between asperities,

$2 > \Lambda > 1$ mixed lubrication in which the load is borne partly by direct naked contact between the surfaces and partly by a lubricating film,

$1 > \Lambda$ boundary lubrication in which the load is borne almost exclusively by direct contact, although the friction coefficient is not so high as that found in the case of dry contact because of the presence of lubricant molecules that cling to the surfaces.

It is not to be supposed that these values are "set in stone": they are merely indicative and some slight variation in the position of the boundaries is found in the literature.

It must also be noted that it is not entirely accurate to say that an EHL contact whose Λ is 3 is the same as the ideally smooth EHL contact. The roughness will cause fluctuations in the contact pressure field. Nevertheless, those fluctuations of rarely more than 50% may be neglected when put alongside those of a contact in the mixed film regime. In illustration of this, it is useful to compare typical values of surface pressure for the various regimes of lubrication. Thus, the maximum surface pressure in an ideal EHL contact is typically in the order of 1 GPa, that in a full-film EHL contact with $\Lambda = 3$ can attain half as much and that in a mixed regime is easily in the order of 5 GPa.

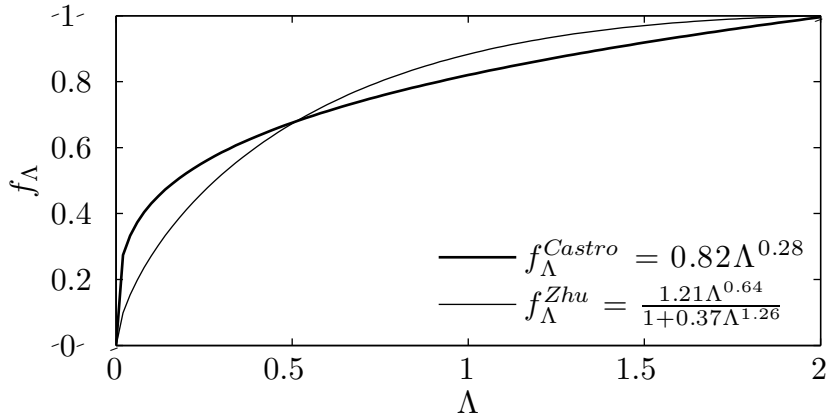


Figure 2.3: Examples of load sharing function taken from [1].

In the case of the meshing of gear teeth, the specific film thickness will find itself somewhere between 0.2 and 2, depending on the operating conditions and even on the precise instant in the meshing. Thus and in order to be able to deal with realistic cases of gear lubrication, it is necessary to develop a mixed film lubrication model.

The model adopted here follows closely that presented by Castro [22] with some modifications.

The contact load f_N may be divided between that borne by the lubricant f_N^{EHL} and that borne by the direct contact between the surfaces f_N^{BDR}, so that:

$$f_N = f_N^{\mathrm{EHL}} + f_N^{\mathrm{BDR}} \tag{2.38}$$

where:

$$f_N = \int_{-\infty}^{+\infty} p \, dx \tag{2.39}$$

$$f_N^{\mathrm{EHL}} = \int_{-\infty}^{+\infty} p^{\mathrm{EHL}} \, dx \tag{2.40}$$

$$f_N^{\mathrm{BDR}} = \int_{-\infty}^{+\infty} p^{\mathrm{BDR}} \, dx \tag{2.41}$$

Furthermore, the parameter f_Λ defined as:

$$f_\Lambda = f_\Lambda(\Lambda) = \frac{f_N^{\mathrm{EHL}}}{f_N} \tag{2.42}$$

is assumed to be a function dependent only on the specific film thickness Λ of the smooth EHL problem. This load sharing function is distinct for each combination of lubricant oil, gear material and roughness orientation (transversal, axial, or isotropic). In Figure 2.3 two examples of load sharing function are shown [1].

Thus:

$$f_N^{\mathrm{EHL}} = f_\Lambda \cdot f_N \tag{2.43}$$

$$f_N^{\mathrm{BDR}} = (1 - f_\Lambda) \cdot f_N \tag{2.44}$$

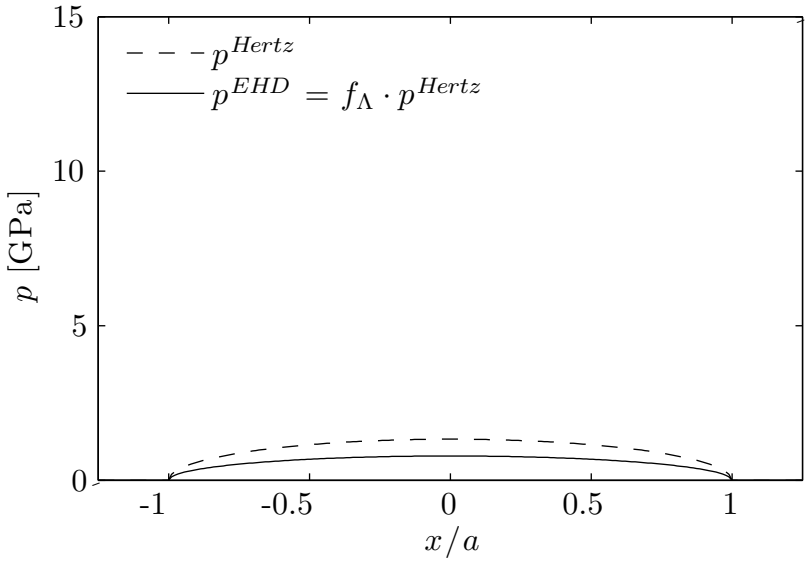


Figure 2.4: Hertzian and EHD surface pressure distribution. The smooth EHD part of the pressure is obtained from the Hertzian pressure distribution for the full load: $\int p^{Hertz}\, dx = f_N$.

Here the present model diverges from that proposed by Castro, for whom $p^{\mathrm{EHL}}/f_\Lambda$ is taken to be the pressure distribution of the EHL contact when $\Lambda = 3$. That pressure distribution was then obtained with the help of the method of Ai and Cheng [23].

Instead, this pressure distribution is obtained as follows (see Figure 2.4). The Hertzian pressure distribution p^{Hertz} and half-width a are obtained for the full load f_N. The pressure distribution borne by the lubricant is then taken as:

$$p^{\mathrm{EHL}} = f_\Lambda \cdot p^{\mathrm{Hertz}} \tag{2.45}$$

The determination of the friction coefficient of the lubricant borne part of the contact μ^{EHL} is then calculated by substituting f_N^{EHL} for f_N in Equations (2.28)–(2.33).

Similarly, the boundary lubrication contact problem is solved for the full load f_N, from which the pressure distribution $p^{\mathrm{BDR,T}}$ is obtained (see Figure 2.5). Then:

$$p^{\mathrm{BDR}} = (1 - f_\Lambda)\, p^{\mathrm{BDR,T}} \tag{2.46}$$

It is now a simple matter of recombining the partial pressure distributions to obtain the mixed film distribution (see Figure 2.6 and 2.7):

$$p = p^{\mathrm{EHL}} + p^{\mathrm{BDR}} \tag{2.47}$$

$$\mu = f_\Lambda \cdot \mu^{\mathrm{EHL}} + (1 - f_\Lambda)\, \mu^{\mathrm{BDR}} \tag{2.48}$$

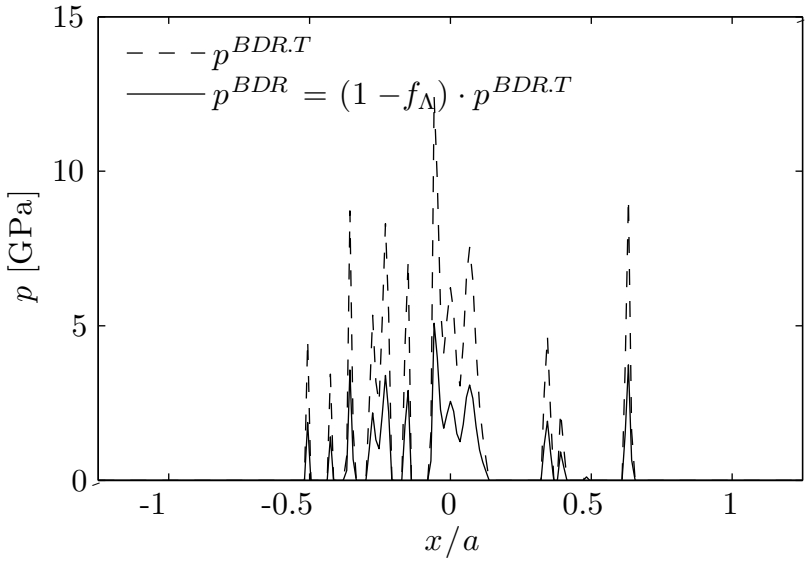


Figure 2.5: Full load and partial load boundary film surface pressure distribution. The boundary film part of the pressure is obtained from the boundary film pressure distribution for the full load: $\int p^{BDR.T}\, dx = f_N$.

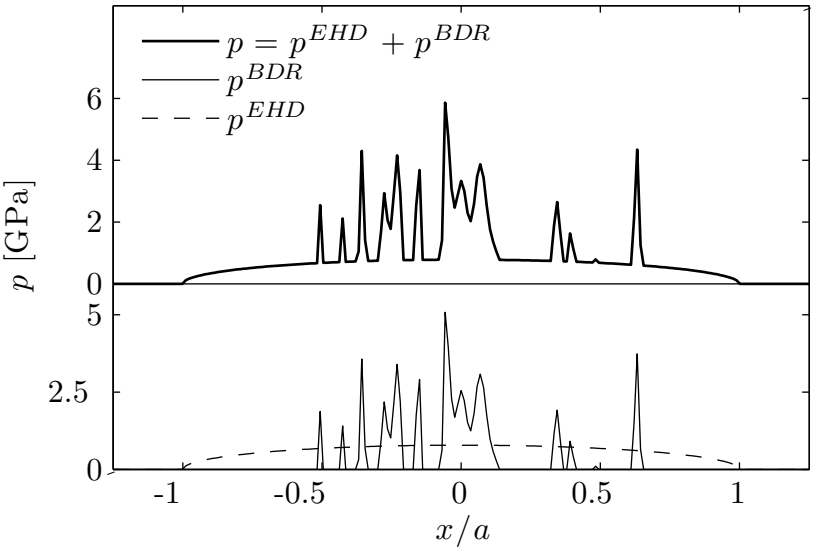


Figure 2.6: The final mixed film surface pressure distribution is the sum of the partial ones.

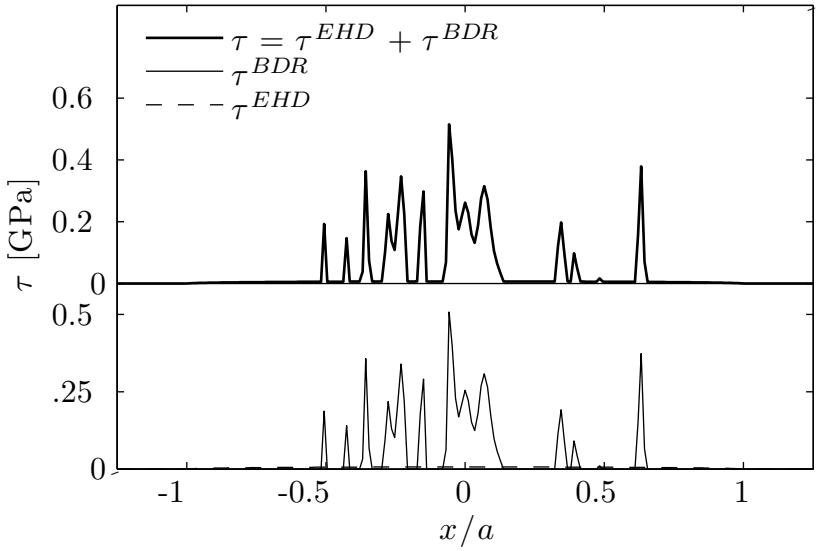


Figure 2.7: The final mixed film surface shear distribution is the sum of the partial ones.

3 Stresses, fatigue and damage in spur gears

3.1 Stresses in a spur gear

The stress state of a spur gear tooth flank may be decomposed thus:

$$[\sigma] = [\sigma_{\text{ini}}] + [\sigma_{\text{ela}}] + [\rho] \tag{3.1}$$

This is the sum of initial stresses (σ_{ini}), present before the application of any load, of elastic stresses (σ_{ela}) caused by the load and of re-centring stresses (ρ) that account for the plastic yield and the stress variations at microscopic level.

3.1.1 Initial Stresses

During its production process, a typical spur gear undergoes cutting, a surface treatment at high temperatures and grinding. All of these steps induce plastic deformations that, upon unloading, translate into important residual stresses, easily attaining orders of magnitude of several hundred MPa.

These initial stresses accumulate with those induced by the loads, and must be taken into account.

In his thesis [24], Batista gives a detailed account of residual stresses, which includes considerations on initial stresses: their definition, causes and the methods employed to measure them—with a particular emphasis on the X-ray diffraction technique.

3.1.2 Elastic stresses

In an elastic half-space in plane strain, it is well known [12] [25] that the elastic stresses are obtained from the surface stresses (see Figure 3.1) thus:

$$\sigma_{xx}(x,z,t) = -\frac{2}{\pi}\int_{-\infty}^{+\infty} p(\xi,t)\,\frac{(x-\xi)^2 z}{\left((x-\xi)^2+z^2\right)^2} + \tau(\xi,t)\,\frac{(x-\xi)^3}{\left((x-\xi)^2+z^2\right)^2}\,d\xi \tag{3.2}$$

$$\sigma_{zz}(x,z,t) = -\frac{2}{\pi}\int_{-\infty}^{+\infty} p(\xi,t)\,\frac{z^3}{\left((x-\xi)^2+z^2\right)^2} + \tau(\xi,t)\,\frac{(x-\xi)\,z^2}{\left((x-\xi)^2+z^2\right)^2}\,d\xi \tag{3.3}$$

$$\sigma_{xz}(x,z,t) = -\frac{2}{\pi}\int_{-\infty}^{+\infty} p(\xi,t)\,\frac{(x-\xi)\,z^2}{\left((x-\xi)^2+z^2\right)^2} + \tau(\xi,t)\,\frac{(x-\xi)^2 z}{\left((x-\xi)^2+z^2\right)^2}\,d\xi \tag{3.4}$$

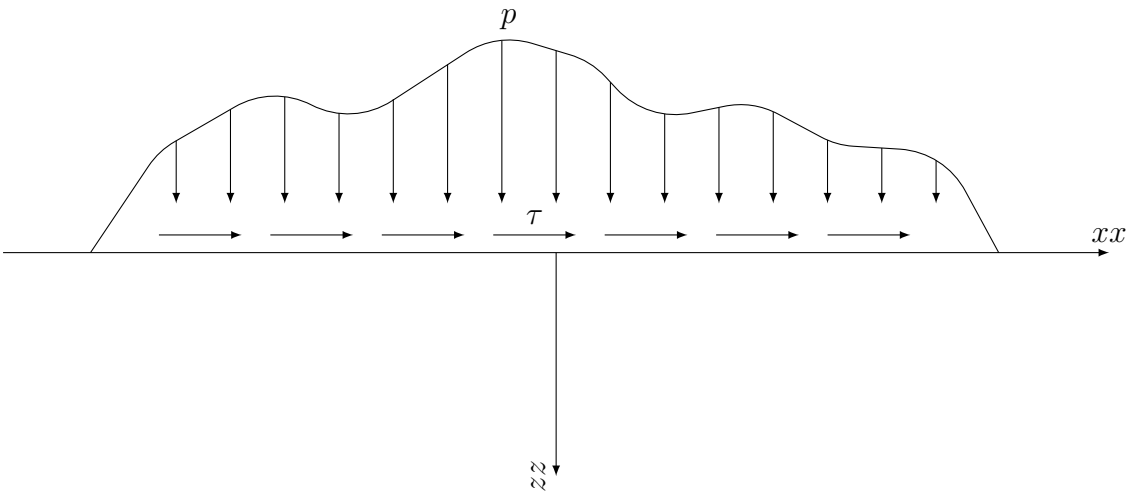


Figure 3.1: Elastic half-space coordinates and surface loads: p is the surface pressure, τ is the surface shear stress.

$$\sigma_{yy}(x,z,t) = \nu(\sigma_{xx} + \sigma_{zz}) \tag{3.5}$$
$$\sigma_{xy} = \sigma_{yz} = 0 \tag{3.6}$$

The principal stresses are:

$$\sigma_1 = \frac{\sigma_{xx} + \sigma_{zz}}{2} + \sqrt{\left(\frac{\sigma_{xx} - \sigma_{zz}}{2}\right)^2 + \sigma_{xz}^2} \tag{3.7}$$
$$\sigma_2 = \frac{\sigma_{xx} + \sigma_{zz}}{2} - \sqrt{\left(\frac{\sigma_{xx} - \sigma_{zz}}{2}\right)^2 + \sigma_{xz}^2} \tag{3.8}$$
$$\sigma_3 = \sigma_{yy} \tag{3.9}$$

The principal shear stresses are:

$$\tau_1 = \frac{\sigma_1 - \sigma_3}{2} \tag{3.10}$$
$$\tau_2 = \frac{\sigma_1 - \sigma_2}{2} \tag{3.11}$$
$$\tau_3 = \frac{\sigma_2 - \sigma_3}{2} \tag{3.12}$$

The maximum shear stress is the largest of these.
The octahedral shear stress is:

$$\tau_{\text{oct}} = \frac{1}{3}\sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2} \tag{3.13}$$

Figures 3.2 and 3.3 show the octahedral shear stress in the case of a Hertzian stress distribution and of a stress distribution resulting from a rough contact, both applied on a spur gear tooth. It is striking to note that, while the τ_{oct} are very

different on the surface, they are virtually the same in the sub-surface. This shows that the influence of the surface roughness is limited to the immediate depths beneath the surface. This has important implications when discussing what type of damage will appear on a gear, or indeed any lubricated mechanical part.

3.1.3 Plastic and mesoscopic stresses

Shakedown

During meshing, the individual gear teeth undergo what are essentially cyclic loads, with each meshing of a particular tooth constituting a cycle. The surface pressures are very high, so high as several GPa, well above the yield strength of the steel from which the gears are made. Thus, plastic deformations are bound to occur. Under these high cyclic loads, one of four things may happen [26] [27] [28]:

1. After a sufficient number of cycles, the stress state of the tooth may converge to an elastic cyclic oscillation around a "central", constant, stress field, and this phenomenon is called elastic shakedown.

2. Another possibility, is that the stress state will converge to a plastic oscillation around a central stress field: this is called plastic shakedown.

3. The strains may increase without bounds with each cycle, and this is called ratcheting.

4. Finally, immediate plastic failure may occur when the loads are overwhelmingly high.

In the case of both elastic and plastic shakedown, the central plastic stress field associated with the central plastic strain field, is in fact a self-balancing residual stress field that endures even after the loads are removed, and only oscillating stresses balance the external load [29].

Note that the elastic shakedown is a global state of the body, because once plastic shakedown sets in the equations of pure elasticity are no longer valid. Nevertheless, considering that the areas affected are always extremely small when compared to the typical dimensions of a tooth, it is legitimate to go on using the elasticity equations even when the conditions for elastic shakedown have been violated at some material points. Thus, in practice, every point in the body under consideration need not be in the same regime of cyclic loading: one point may be undergoing elastic shakedown while another undergoes plastic shakedown and yet another ratcheting [29].

The boundary between a long or infinite life to failure and a short one—the limit between high cycle fatigue (HCF) and low cycle fatigue (LCF)—is located at the transition from an elastic shakedown state to a plastic one [26]. In the case of a gear that must undergo a large number of cycles—many millions—the component

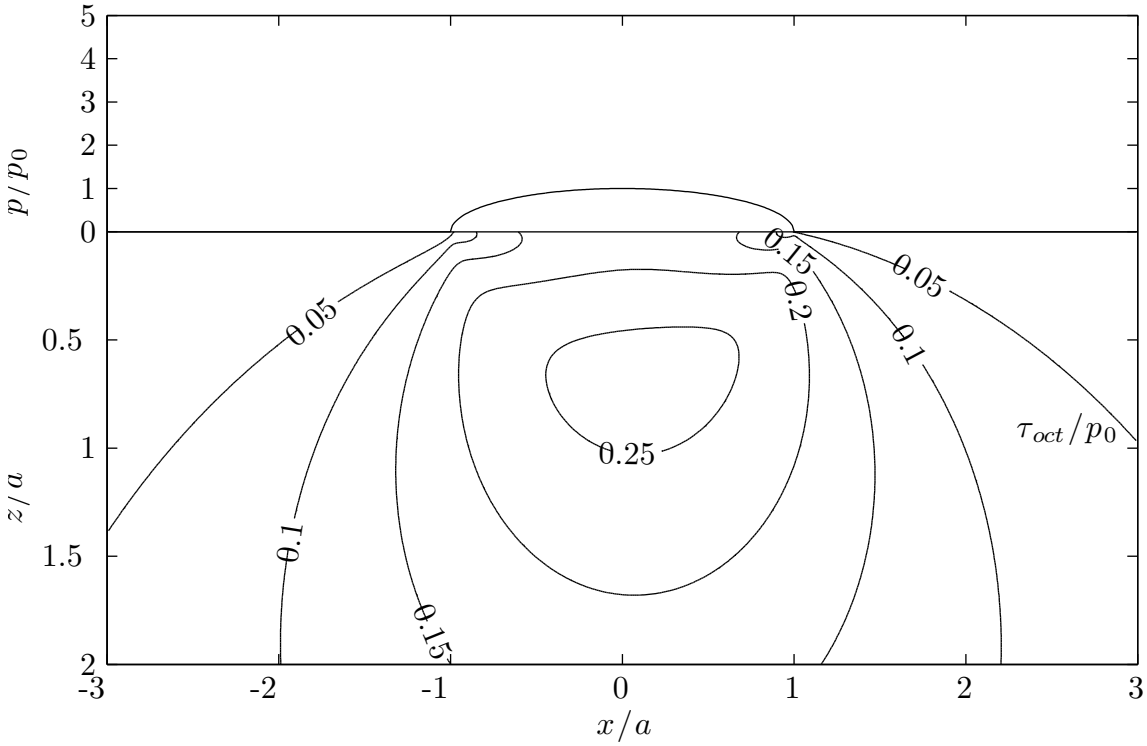


Figure 3.2: τ_{oct}/p_0 in a tooth submitted to a Herzian pressure distribution with a friction coefficient of 0.05.

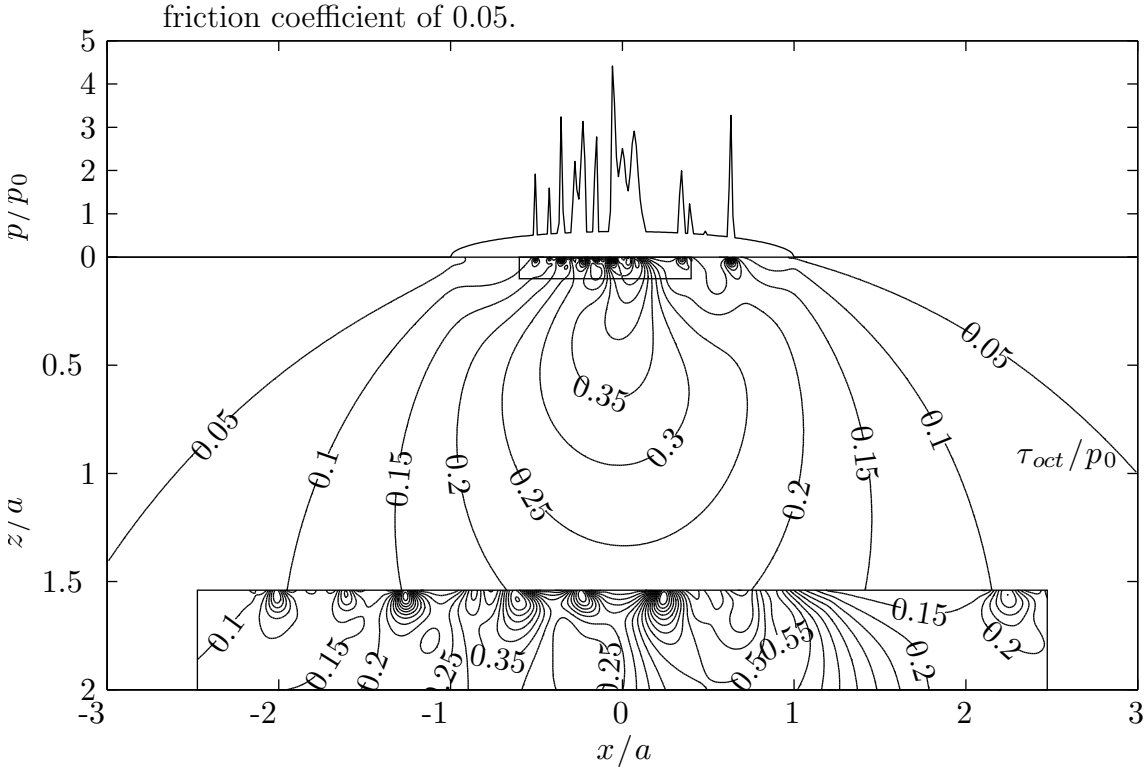


Figure 3.3: As in Figure 3.2 but under mixed film lubrication.

must be targeted to HCF. So, supposing that a material point in a gear tooth will converge to elastic shakedown, the stresses will be, in tensor notation:

$$\tilde{\sigma} = \left(\tilde{\sigma}^{\text{ini}} + \tilde{\sigma}^{\text{ela}}\right) + \tilde{\rho} \tag{3.14}$$

where $\tilde{\rho}$ is the residual stabilized stress field tensor, $\tilde{\sigma}^{\text{ela}}$ the elastic one and $\tilde{\sigma}^{\text{ini}}$ the initial one.

Considering an elastic-perfectly plastic material subjected to the Von Mises yield criterion, Melan's shakedown theorem states that, provided that an admissible stress tensor $\tilde{\rho}$ exists such that, when added to the elastic stress tensor it is placed on the yield surface, the body will shakedown elastically and the shakedown residual stress tensor will be $\tilde{\rho}$. For more complex material behaviour, such as plastic strain and kinematic hardening, shakedown can still occur, but Melan's theorem must be altered to take this into account. In principle, this obviates the need to follow the plastic deformation of the body cycle by cycle, and allows instead to obtain directly the more important converged state.

The paper by Chinh [30] includes a thorough mathematical treatment of the shakedown theorem and extends it. The paper by Constantinescu, Dang Van et al. [26] discusses in detail the relation between shakedown and fatigue. Shakedown analyses of a Hertzian rolling/sliding contact are presented in [29], [27] and [28].

Von Mises yield criterion

It is convenient at this point to recall the equations of the Von Mises plastic yield criterion. A stress tensor $\tilde{\sigma}$, be it macroscopic or mesoscopic, elastic or plastic, must obey the law of the balance of forces:

$$\text{div}\tilde{\sigma} = 0 \tag{3.15}$$

where it is supposed that both the body and inertial forces are negligible, a reasonable simplification when dealing with the high contact stresses of the meshing.

As a consequence, the tensor is symmetric and its invariants are:

$$\begin{aligned} I_1 =& \text{tr}\tilde{\sigma} \\ =& \sigma_{xx} + \sigma_{yy} + \sigma_{zz} \\ =& \sigma_1 + \sigma_2 + \sigma_3 \end{aligned} \tag{3.16}$$

$$\begin{aligned} I_2 =& \frac{1}{2}\Big((\text{tr}\tilde{\sigma})^2 - \text{tr}\left(\tilde{\sigma}^2\right)\Big) \\ =& \sigma_{xx}\sigma_{yy} + \sigma_{yy}\sigma_{zz} + \sigma_{xx}\sigma_{zz} - \sigma_{xy}^2 - \sigma_{yz}^2 - \sigma_{xz}^2 \\ =& \sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3 \end{aligned} \tag{3.17}$$

$$\begin{aligned} I_3 =& \det\tilde{\sigma} \\ =& \sigma_{xx}\sigma_{yy}\sigma_{zz} + 2\sigma_{xy}\sigma_{yz}\sigma_{xz} - \sigma_{xx}\sigma_{yz}^2 - \sigma_{yy}\sigma_{xz}^2 - \sigma_{zz}\sigma_{xy}^2 \\ =& \sigma_1\sigma_2\sigma_3 \end{aligned} \tag{3.18}$$

The octahedral shear stress is:

$$\begin{aligned}\tau_{\text{oct}} &= \frac{\sqrt{2I_1^2 - 6I_2}}{3} \\ &= \frac{1}{3}\sqrt{(\sigma_{xx}-\sigma_{yy})^2 + (\sigma_{yy}-\sigma_{zz})^2 + (\sigma_{zz}-\sigma_{xx})^2 + 6\left(\sigma_{xy}^2+\sigma_{yz}^2+\sigma_{xz}^2\right)} \\ &= \frac{1}{3}\sqrt{(\sigma_1-\sigma_2)^2 + (\sigma_2-\sigma_3)^2 + (\sigma_3-\sigma_1)^2}\end{aligned} \tag{3.19}$$

The stress tensor can be split into its volumetric and deviatoric part:

$$\tilde{\sigma} = \tilde{s} + p_H \cdot \tilde{I} \tag{3.20}$$

$$\tilde{s} = \text{dev}\tilde{\sigma} \tag{3.21}$$

$$p_H = \frac{1}{3}I_1 = \frac{1}{3}\left(\sigma_{xx}+\sigma_{yy}+\sigma_{zz}\right) = \frac{1}{3}\left(\sigma_1+\sigma_2+\sigma_3\right) \tag{3.22}$$

where p_H is the hydrostatic stress and $\tilde{s}$ the deviatoric stress tensor.

The hydrostatic part is responsible for the volumetric change of the body and the deviatoric part for its distortion. Because the first invariant of the deviatoric stress tensor is always null, its second invariant is:

$$\begin{aligned}J_2 &= I_2 - \frac{1}{3}I_1^2 \\ &= -\frac{1}{6}\Big((\sigma_{xx}-\sigma_{yy})^2 + (\sigma_{yy}-\sigma_{zz})^2 + (\sigma_{zz}-\sigma_{xx})^2 + 6\left(\sigma_{xy}^2+\sigma_{yz}^2+\sigma_{xz}^2\right)\Big) \\ &= -\left(\frac{s_{xx}^2+s_{yy}^2+s_{zz}^2}{2} + s_{xy}^2 + s_{yz}^2 + s_{xz}^2\right)\end{aligned} \tag{3.23}$$

On the other hand, and in the case of an isotropic elastic material, the volumetric distortional strain energy is given by:

$$U_s = -\frac{1}{2G}\left(I_2 - \frac{1}{3}I_1^2\right) \tag{3.24}$$

The Von Mises criterion states that the material yields when a limit to the elastic distortional strain energy is reached. Comparing Equations (3.19), (3.23) and (3.24), it follows that:

$$\tau_{\text{oct}} = \sqrt{-\frac{2}{3}J_2} \tag{3.25}$$

$$U_s = -\frac{J_2}{2G} = \frac{\tau_{\text{oct}}^2}{3G} \tag{3.26}$$

Thus imposing a limit to the distortional strain energy, to the second invariant of the deviatoric stress tensor or to the octahedral shear stress are equivalent applications of the Von Mises yield criterion. For instance, for a material point *not* to yield, its stress tensor must obey the equation:

$$-J_2 - K^2 \leq 0 \tag{3.27}$$

where K is a material property. This can be rewritten as:

$$\left(\frac{s_{xx}^2 + s_{yy}^2 + s_{zz}^2}{2} + s_{xy}^2 + s_{yz}^2 + s_{xz}^2\right) - K^2 \leq 0 \tag{3.28}$$

It is important to note that, if one maps $s_{xx}/\sqrt{2}$, $s_{yy}/\sqrt{2}$, $s_{zz}/\sqrt{2}$, s_{xy}, s_{yz}, s_{xz} to a six-dimensional space, the yield surface becomes a hypersphere.

Mesoscopic stresses

As outlined by Constantinescu, Dang Van et al. [26], three scales must be distinguished when discussing fatigue: the microscopic scale of dislocations within crystals, the mesoscopic scale of crystal grains and the macroscopic scale of the studied component as a whole, with typical distances of:

scale	typical lengths	
macroscopic	10^{-3} m	nominal geometry
mesoscopic	10^{-6} m	chrystal grain
microscopic	10^{-10} m	interatomic distance

At the macroscopic scale, the material properties vary smoothly and the material is continuous. At the mesoscopic scale, the material properties cannot be said to vary smoothly because of the difference in orientation of the grains. Nevertheless, the material may still be considered continuous because a grain contains a sufficiently large number of atoms and it thus still makes sense to talk of stresses and strains. At the microscopic scale, the material becomes a discrete aggregate of atoms, an no continuity of any kind exists.

The stresses discussed in this chapter up to this point are all at the macroscopic scale. On the other hand, the initiation of a fatigue crack is widely held to be a consequence of the nucleation of dislocations within crystal grains (see the internet website [31] for a succinct discussion of crystal defects and micro-plasticity; for a complete mathematical treatment of macro and micro plasticity, see [32]). It is therefore necessary to be able to evaluate the mesoscopic stresses. It must be said that the macroscopic stresses may be though of, in some sense, as an average of the mesoscopic stresses over a representative volume element (RVE) of the size of many grains. Thus, the mesoscopic stresses are the sum of the mesoscopic stress and of a perturbation due to the difference in elastic properties caused by the difference in grain orientation.

Dang Van, in an article describing his fatigue criterion [33], proposed a method of obtaining the mesoscopic stresses based on the elastic shakedown concept. While this method was presented as an integral part of the criterion, determining mesoscopic stresses is useful in itself. This is the reason why it is presented here independently of the fatigue criterion.

In the previous section, it was said that the macroscopic stabilized stress tensor could be obtained directly, without performing the computationally costly step

of following the intervening stages. While this is true in principle, it involves "guessing" the actual stabilized residual stress tensor and verifying it in the light of Melan's theorem. In practice, the analyst of the system makes assumptions about probable residual stress tensors and obtains upper and lower bounds of the shakedown limit within which the loads induce elastic shakedown (see [34] for a good explanation of this). Such an approach is detailed in [30] and [29]. Dang Van's method of determining the *mesoscopic* stabilized residual stress bypasses this step by allowing one to directly get at the mesoscopic stresses.

In the general case, the mesoscopic stress tensor is related to the macroscopic stress tensor thus:

$$\tilde{\Sigma} = \tilde{\tilde{A}} : \tilde{\sigma} + \tilde{\rho} \tag{3.29}$$

where $\tilde{\Sigma}$ is the mesoscopic stress tensor, $\tilde{\sigma}$ the macroscopic one, $\tilde{\rho}$ is the stabilized mesoscopic residual stress tensor, constant in time, and $\tilde{\tilde{A}}$ is the fourth order localization elastic tensor, that accounts for the differences in the elastic properties of the grains from the macroscopic ones due to their orientation.

As argued by Dang Van, it is possible to eliminate the localization tensor from Equation (3.29) by making a few reasonable assumptions. Namely, it is supposed that within the RVE centred on each material point considered, at least one grain is so unfavourably oriented that it will slip under the stresses, and that the material suffers isotropic and kinematic hardening at the mesoscopic level. Finally, it is supposed *a priori* that at the local mesoscopic level the material around each point will shakedown elastically, in a manner similar to the global elastic shakedown. Thus Equation (3.29) reduces to:

$$\tilde{\Sigma} = \tilde{\sigma} + \tilde{\rho} \tag{3.30}$$

This happens because, as the cycle progresses, the yield surface grows and shifts to encompass all the stress states through which the considered material point has travelled (see Figure 3.4). Once more, this is valid because of the assumption that elastic shakedown *does* occur locally—this is similar to conducting elastic calculations to check for yield—and that the material, at the mesoscopic scale, undergoes isotropic and kinematic hardening—a very general description of the plastic behaviour of a material and therefore a very reasonable one.

The yield criterion used is the Dang Van criterion, and therefore the yield surface is a hypersphere in the axes $s_{xx}/\sqrt{2}$, $s_{yy}/\sqrt{2}$, $s_{zz}/\sqrt{2}$, s_{xy}, s_{yz}, s_{xz}, whose radius expands—this is isotropic hardening—and whose centre moves— this is kinematic hardening. The final position and radius of the yield hypersphere is then such that the hypersphere is the smallest one that encompasses all the stress states in the cycle for the point under consideration. The stabilized mesoscopic residual stress associated with the local shakedown ($\tilde{\rho}$) is then the centre of the yield sphere. From this process of obtaining $\tilde{\rho}$ it follows necessarily that it is a purely deviatoric stress tensor. Mathematically, the yield limit is determined by solving the optimization

problem:

$$\begin{aligned} K^2 &= \min_{\tilde{\rho}'} \Big(\max_t \big(-J_2 \left(\tilde{\sigma}\left(t\right) + \tilde{\rho}' \right) \big) \Big) \\ \tilde{\rho} :\ K^2 &= \max_t \big(-J_2 \left(\tilde{\sigma}\left(t\right) + \tilde{\rho} \right) \big) \end{aligned} \tag{3.31}$$

Performing the mapping of the stresses to a six-dimensional space, as mentioned earlier, this becomes the well known geometric problem of the smallest enclosing ball.

As an illustration of the principle, consider a body in which the external loads only induce pure shear stress, with all the stress components null except σ_{xz} and σ_{yz}, at a given point in the body. The macroscopic elastic stress history of the point in the body may be conveniently represented in two dimensions, as in Figure 3.4. Because of this the yield surface collapses into a circle.

At the start of the cycle, the stress is equal to an initial stress σ_0 and the yield radius is the initial one. The cycle progresses to instant t_1, at which the elastic stress tensor is σ_1. Because in reaching this instant, the elastic stress tensor has moved outside the initial yield surface, it has pulled it along so that the yield surface centre has shifted to ρ_1 and its radius has dilated to K_1 in order to envelop both σ_0, σ_1 and every intermediate stress state. The very same thing happens when the cycle progresses to σ_2. Finally, when the full cycle has been gone through, the yield surface settles into its final shape and position and need no longer change with the application of further cycles.

3.2 Dang Van multi-axial high-cycle fatigue criterion

The mesoscopic stresses evaluated by the method presented in the previous section are those necessary to ensure the existence of elastic shakedown locally—in fact, to ensure near infinite life to fatigue. Whether the material has the capacity to sustain these stresses or not is another matter. The Dang Van criterion is the tool used to check this possibility at each instant after the stabilization of the residual stress.

A fatigue crack in its initial stage usually propagates along a plane of maximum shear strain, which corresponds in the isotropic case to a plane of maximum shear stress. Thus, the maximum shear stress is a relevant parameter of the initiation of a fatigue crack. On the other hand, a negative hydrostatic stress—a hydrostatic pressure—has been observed to benefit the resistance to fatigue of materials. From these considerations, Dang Van formulated [33] the simplest possible law that relates these parameters:

$$\tau_{\max} + \alpha_{\mathrm{DV}} \cdot p_H \leq \beta_{\mathrm{DV}} \tag{3.32}$$

where $\tau_{\max}$ and p_H are the maximum shear stress and the hydrostatic stress—not pressure—of the mesoscopic stress and α_{DV} and β_{DV} are fatigue material properties.

The material properties β_{DV} and α_{DV} are obtained respectively by performing reversed torsion and alternating bending tests, in order to obtain two points on the

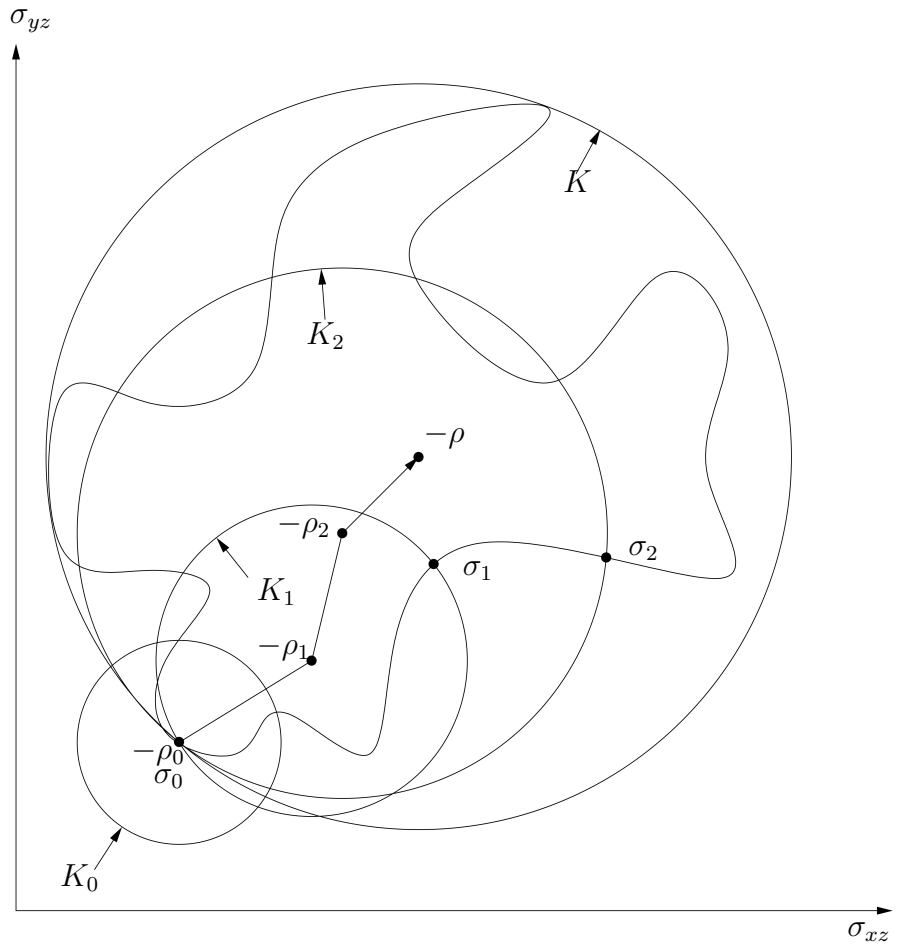


Figure 3.4: Stress cycle and hardening of a material point in pure shear stress.

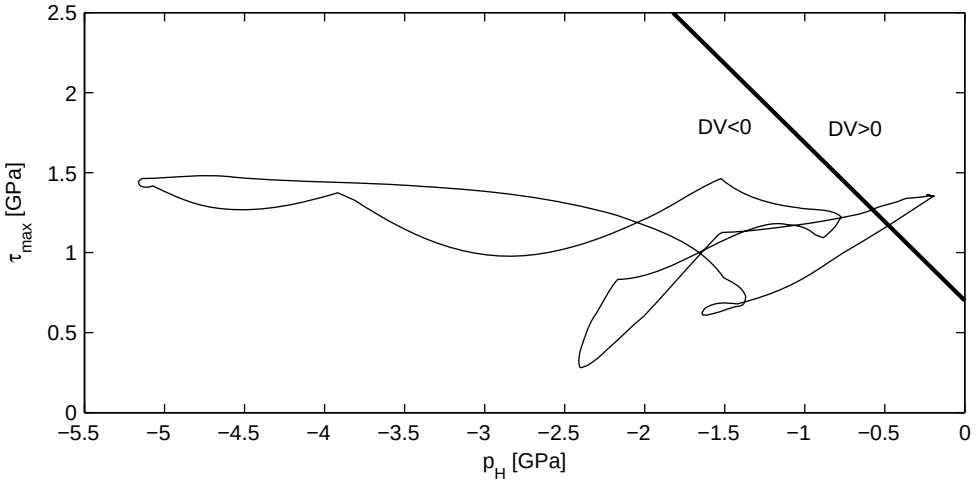


Figure 3.5: Position of the mesoscopic stress state of a material point during a load cycle on the $p_h/\tau_{\max}$ plane. DV $= \frac{\tau_{\max}+\alpha_{\mathrm{DV}}\cdot p_H}{\beta_{\mathrm{DV}}} - 1$

line described by Equation (3.32).

Because the maximum shear stress occurs in two mutually perpendicular planes, the orientation of the crack as it initiates can be either one, or even both.

Another interpretation can be given to the criterion, derived from the manner of obtaining β_{DV}. One can think of the stress cycle at each point as equivalent to a reversed torsion stress cycle such that its maximum shear stress is:

$$\beta_{eq} = \max_t \left(\tau_{\max} + \alpha_{\mathrm{DV}} \cdot p_H\right) \tag{3.33}$$

Thus, fatigue cracks do not initiate if this equivalent shear stress is less than the value of β_{DV}.

$$\beta_{eq} \leq \beta_{DV} \tag{3.34}$$

This equation must be verified at each instant in the cycle. Any point at which this is not true must eventually be the origin of a fatigue crack: it passes into the domain of low cycle fatigue or even ratcheting.

Figure 3.5 shows the position of the mesoscopic stress state of a material point during a load cycle on the $p_h/\tau_{\max}$ plane. It is seen that the path of the stress crosses the straight line delimiting the safety zone. Interestingly, the most critical instant occurs when p_H is at its lowest absolute value, in other words, when the external loads are null and the only hydrostatic stress comes from the initial stress. This is due to the fact that under loading, contact pressure induces high hydrostatic pressures that shift the graph to the left and under the limiting line. By the same token, a compressive initial stress has the same beneficial effect, as can be observed in Figure 3.5.

Numerous papers have been published that apply the Dang Van criterion to rolling contact fatigue—spurred by applications in the railroad and roller bearing

industries—such as [35], [36], [37] and [38]. Notably, in the last reference, Desimone et al. proposed a refinement of the criterion that will be discussed in Chapter 5.

3.3 Types of tooth flank surface damage

Faure, in [9], presents a comprehensive review of the types of damage that appear on gears. Thus he distinguishes several types of damage:

wear The progressive removal of matter from the surface of a tooth with use. This term covers a multiplicity of phenomena:

- **running-in** This is not strictly a damage. New gears, when used with light loads will see their roughness decrease in the first few hours. This has a beneficial effect on the duration of the gears.
- **current wear** The unavoidable wear that comes from the mutual sliding of surfaces. The exact mechanism is not known, but it is suspected that a combination of brittle fracture, plastic failure and fatigue come into play as roughness features collide.
- **scoring** Salient roughness features of one tooth dig into the opposite contacting tooth leaving a score mark. This is not relevant in the case of spur gears because their roughness is markedly anisotropic and the direction of greater roughness is perpendicular to that of sliding.
- **adhesive wear** The tooth surfaces contact at high pressure and micro-welding or adhesion occurs, after which material is torn form the surface. Faure includes in this class of damage both hot and cold scuffing.
- **three bodies wear** scratches and abrasion caused by the presence of particles in the lubricant.

corrosion Chemical phenomenon that originates from the presence of water or other corrosive agents in the lubricant.

overheating or burning An excessive temperature in operation can cause an accidental heat treatment that lowers the hardness of the gear material.

erosion by cavitation In some cases, in particular under high alternating loads, cavitation of the lubricant can occur. Both the implosion of the cavitated bubbles and the projection of high speed droplets cause shock-waves that produce impact craters with circumferential cracks. Due to the near-instantaneous duration of these events, the material response is very brittle so that no real progressive fatigue is involved.

electric erosion Removal of matter through the application of electric arcs that are produced by the high friction of the surfaces.

plastic deformation This covers the permanent deformations of the surfaces caused by excessive contact pressure.

contact fatigue The damages caused by the cyclic nature of the loads on the surface of a tooth. They are:

case crushing When the Hertzian maximum shear stress is located below the cemented or nitrurated layer, a fatigue crack may develop that, upon reaching the surface, causes the surface treated layer to be removed. The roughness of the surfaces plays little or no part in its creation since the stresses of a rough tooth at the initiation depth are nearly the same as in the case of a smooth Hertzian contact, as was shown in Figures 3.2 and 3.3.

spalling A crack originates at the depth of Hertzian maximum shear and propagates to the surface. In consequence, a flake is removed from the tooth, living a large crater of several hundred microns in depth and width. As in the case of case crushing, and for the same reason, the roughness has no influence on this type of damage.

pitting A crack originates on the surface and propagates downwards to a depth of over 100 μm before rejoining the surface, at which point a pit is formed. Since the origin of the crack is on the surface, the complex stress state near the surface caused by the roughness is of great importance in its occurrence.

micropitting As in the case of pitting, a crack originates from the surface and propagates downward but only to a depth of around 20 μm. Thus, the only difference from a pit is one of scale. The mechanisms at play must be those of pitting. Figure 3.6 shows a picture of a section of a tooth that exemplifies well the size and shape of micropits.

Fatigue is relevant only to the phenomena of normal wear and contact fatigue. By restraining the analysis of a tooth to the first 20 μm of depth, it is possible to isolate the initiation of fatigue damage related to pitting and micropitting from the other types of damage by contact fatigue. Because the phenomenon of pitting may be seen as a more advanced stage of micropitting, there is no need to distinguish between them when modelling the initiation of cracks. The precise separation of wear and micropitting is not so easy: some of the wear may be caused by fatigue cracks on the surface. The distinction must be made on the end results of the analysis: it is expected that wear is relatively uniform on large patches of the surface, while micropitting consists of small craters of a few microns of width on the surface.

When performing a longitudinal cut of a gear tooth, the surface cracks always intersect the plane of cut in well defined directions, as shown in Figure 3.7: they progress downward in the direction opposite to sliding at a shallow angle ($< 30°$).

It has been observed that micropitting is concentrated mainly on the flank of the pinion below the primitive line. This is consistent with the widely held view that the sliding $\vec{U}_2 - \vec{U}_1$ is most harmful when in the direction of rolling. Figure 3.8, where the picture of the micropitted surface of a tooth is shown, is an example of this.

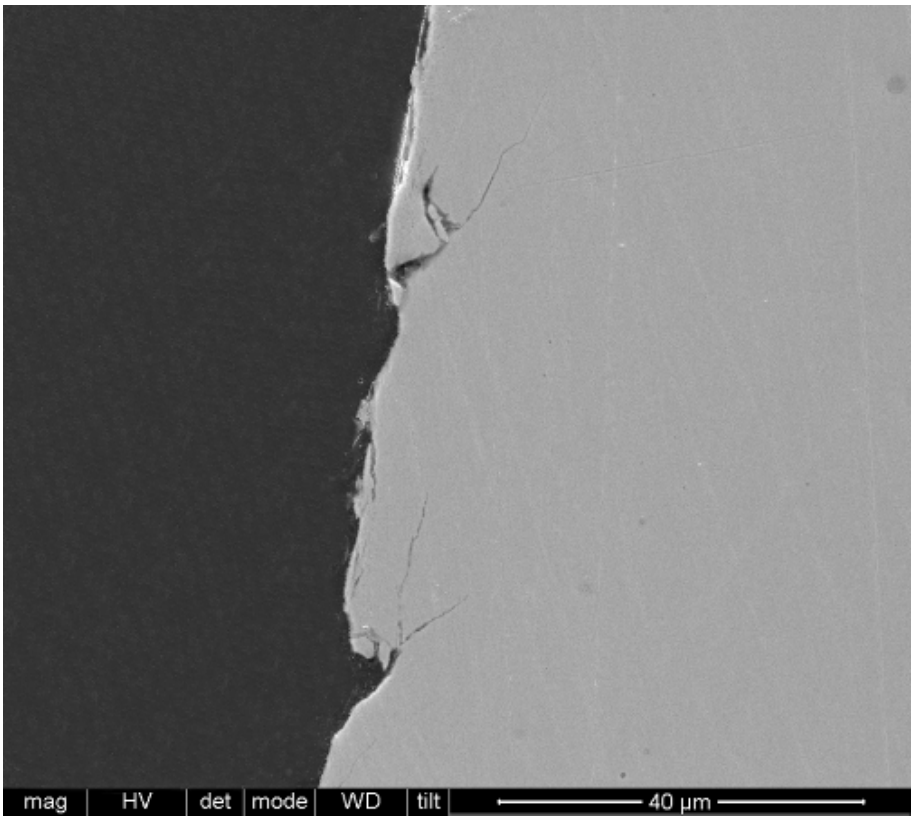


Figure 3.6: Photograph of a cut of a gear tooth where micropitting has occurred.

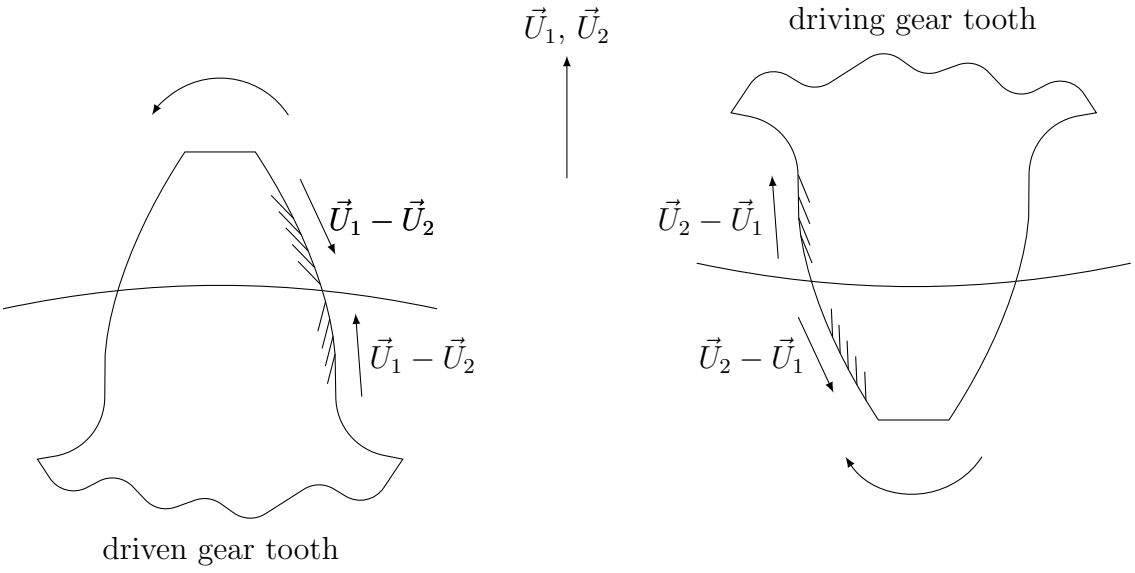


Figure 3.7: Orientation of surface fatigue cracks according to their position on the teeth. The directions of the rolling and sliding velocities and of the rotations are also shown.

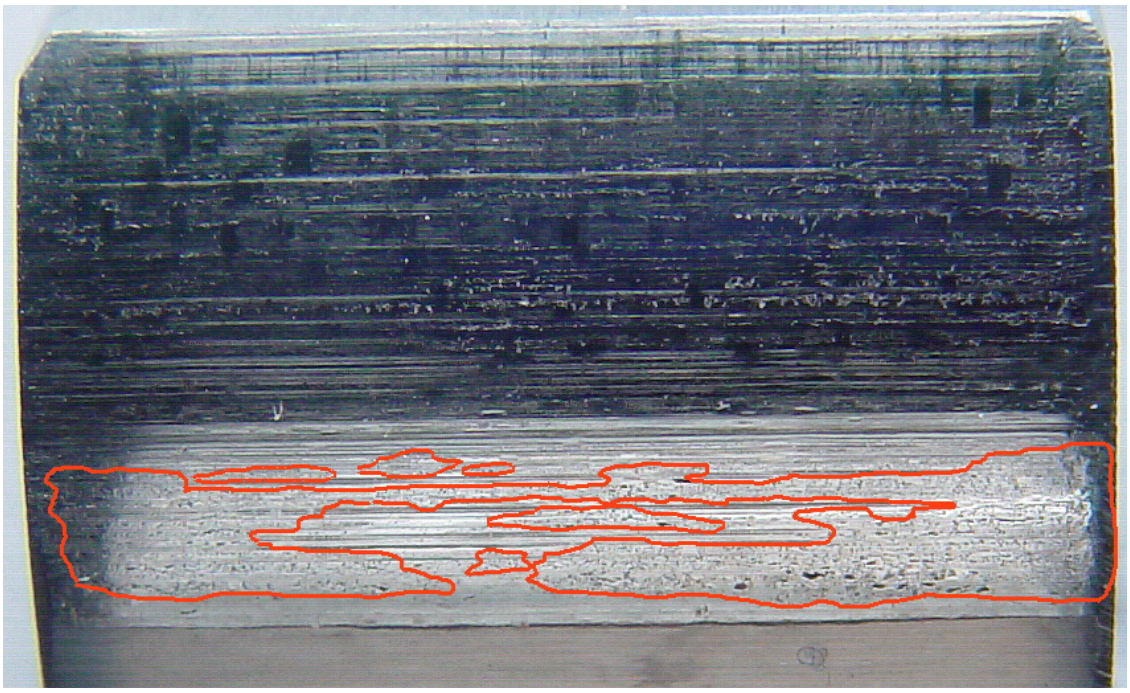


Figure 3.8: Photograph of the surface of a gear tooth where micropitting has occurred. The micropitted areas are surrounded by a red line. It is seen that micropitting is mainly restricted to the part of the flank below the pitch line.

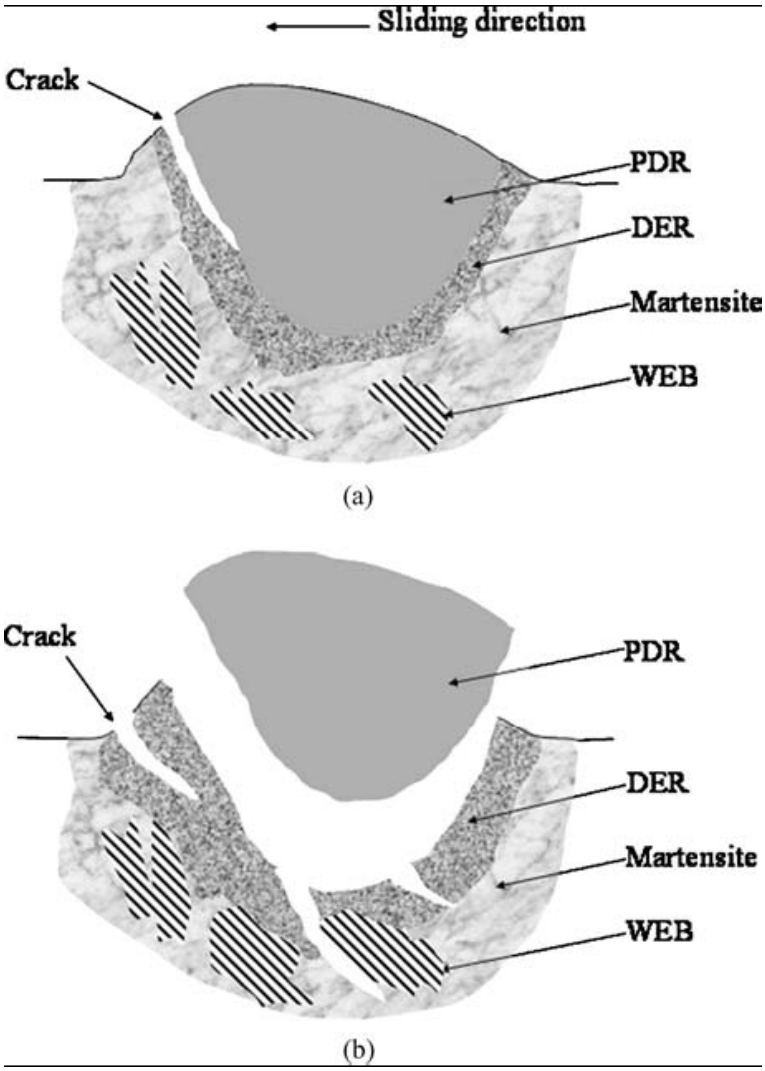


Figure 3.9: Shematic representation of micropitting initiation (a) and micropitting popagation (b) (taken from [2, Figure 11]).

Table 3.1: Factors that influence micropitting and suggested remedies.

influencing factor	range of influence	suggested remedy
grear surface roughness	$1 \rightarrow 3$ (from 6 μm to 3 μm)	reduce to 0.3 μm
material, heat treatment	$1 \rightarrow 2.8$	retained austenite
lubricant viscosity	$1 \rightarrow 2$	use highest practical vicosity
lubricant additive chemistry	$1 \rightarrow 2$ (equal viscosity of base oil)	use properly selective additives
coefficient of friction	$1 \rightarrow 1.7$	reduce the coefficient of friction
speed	$1 \rightarrow 1.3$	run at high speed
oil temperature	$1 \rightarrow 1.3$ ($\Delta T = 20$ K)	reduce oil temperature

In their articles [5] and [2], Oila and Bull made an extensive study of the phenomenon of micropitting with particular emphasis on its crystallographic aspects. In particular, in [2], they proposed a mechanism of initiation and propagation of micropitting cracks that is related with the formation of a phase produced by intense plastic deformations on the immediate surface of a tooth, coined plastic deformation region (PDR) by Oila and Bull, as shown in Figure 3.9, that was borrowed from [2].

The factors that influence micropitting, their range of influence and suggestions for preventing it, according to Cardis and Webster [39], are listed in Table 3.1 in descending order of importance. It is interesting to notice that the factors can be grouped in three main categories with some overlap:

specific film thickness: the surface roughness, the lubricant viscosity, the speed and oil temperature;

fatigue behaviour: the material, the heat treatment—or, in other words, the initial stresses in the gear flank surface;

friction coefficient: overall friction coefficient, speed, oil temperature, additive chemistry, which is known to be mainly felt at the level of boundary film lubrication.

It can be verified that the mixed film lubrication model and the contact fatigue initiation model address all of these issues.

The table shows that the roughness is of primordial importance. This is in accordance with the work of Gonçalves [40], which is in many ways a precursor to this one, who demonstated the influence of the roughness profiles in the results that she obtained with her own micropitting model.

Is is interesting to note a conspicuous absence from Table 3.1: the maximum Hertzian pressure. This is significant because in their paper [5] Oila and Bull singled out the maximum Hertzian pressure as the most important factor in the initiation of micropitting—their definition of initiation is different than that used throughout this work: they take it to mean the instant when the flank surface area where micropitting has occurred reaches 1.5% of the total flank surface area.

4 Numerical model

4.1 Overview of the model

The numerical model developed is divided into two major parts: the mixed film lubrication and the surface fatigue damage models. The general scheme is shown in Figure 4.1.

The first part is a loop over time that goes through an entire meshing cycle in order to obtain the elastic stresses at each instant. The box *geometry, kinematics, loads* corresponds to Chapter 1, where the details of the determination of the geometry of the surfaces, the velocities, the contact loads and the Hertzian contact are detailed.

The boxes *load sharing function* $f_\Lambda(\Lambda)$, *smooth EHD lubrication*, *rough boundary lubrication* and *mixed regime lubrication* correspond to Section 2.3.2, where a proposed approximate solution to the contact problem in the mixed film lubrication regime is presented.

The box *elastic stresses* corresponds to Section 3.1.2, where the theory behind the determination of the elastic stresses is laid down. Note that the stresses are calculated only to a depth of 20 μm in order to isolate micropitting from other types of fatigue failure. Furthermore, the stresses are only obtained for the driving gear tooth, since it has been found that micropitting is more prevalent on these.

The second major part concerns itself with the application of the Dang Van criterion. The calculations are no longer performed in a time loop. Thus knowing the initial stresses and the elastic stresses, the *mesoscopic residual stresses* are obtained by applying the method detailed in Section 3.1.3. The *mesoscopic stresses at every instant* are then easily obtained by adding:

$$\Sigma(x,z,t) = \sigma^{\text{ini}}(x,z) + \sigma^{\text{elas}}(x,z,t) + \rho(x,z) \tag{4.1}$$

Finally, the *Dang Van fatigue criterion* is applied as explained in Section 3.2. This calculation is performed only on the driving gear tooth, for reasons already explained.

Note that a number of difficulties regarding numerical implementation have been ignored until now. The most salient ones are discussed in what follows.

4.2 Numerical implementation

In order to perform the calculations listed in Figure 4.1, the driving gear tooth is discretized in the xz plane, where x is the arc distance on the surface of the tooth

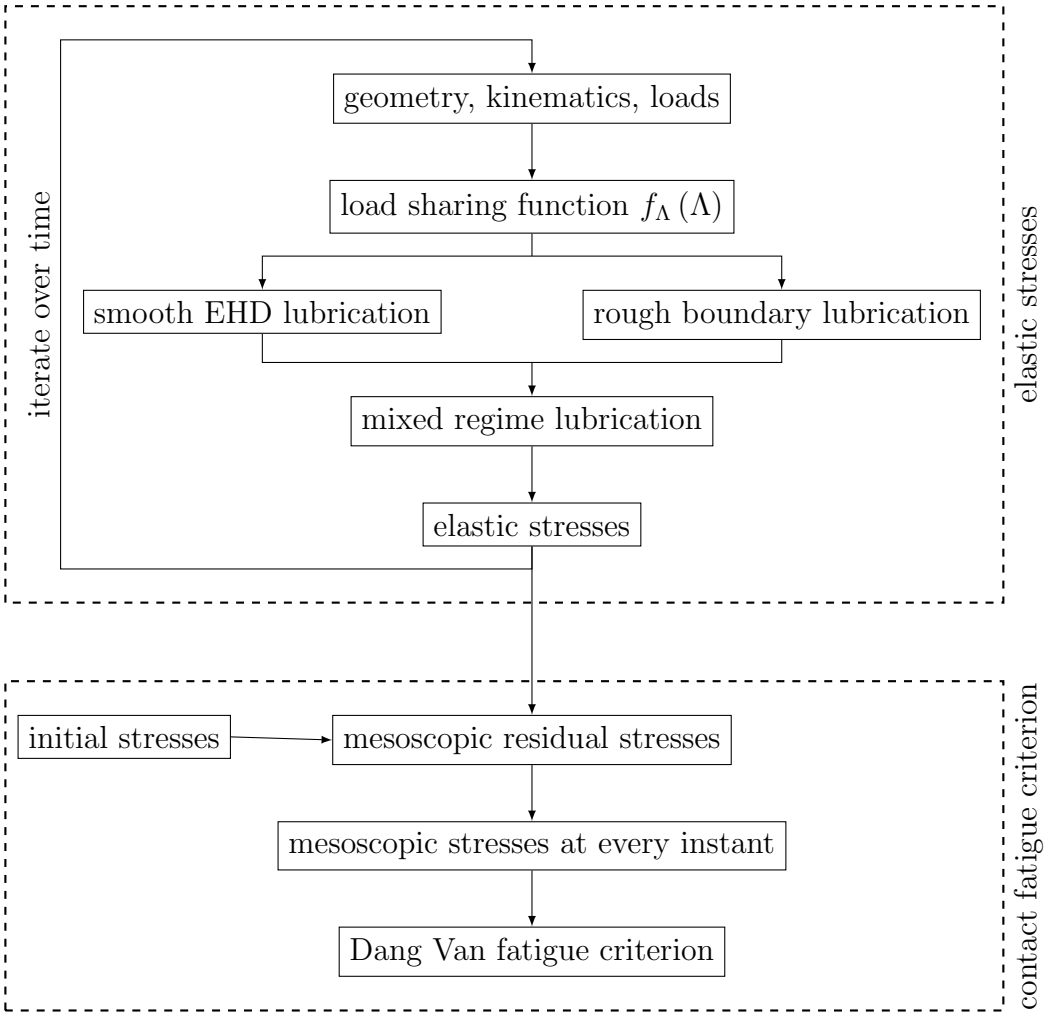


Figure 4.1: Scheme of the model for the prediction of micropitting crack initiation.

and z is the depth. Thus time is discretized as a succession of instants in which the theoretical Hertzian initial contact travels from nodal point to nodal point on the x axis.

Most of the steps in Figure 4.1 are trivial to implement numerically. Nevertheless some numerical difficulties were encountered during the implementation of the model. They are now discussed.

4.2.1 The rough boundary lubrication contact problem

This section deals with the problem put in Section 2.4.2. More precisely, the problem is to compute the pressure and the final separation of the surfaces.

The equations are here recalled:

$$h(x) = h_0(x) - \frac{2}{\pi E_{eq}} \int_{-\infty}^{+\infty} \ln \left| \frac{x - x'}{L} \right| p(x') \, dx' \tag{4.2}$$

$$f_N = \int_{-\infty}^{+\infty} p(x) \, dx \tag{4.3}$$

$$\forall x : \big(h(x) \geq 0 \ \wedge \ p(x) = 0\big) \vee \big(h(x) = 0 \ \wedge \ p(x) \geq 0\big) \tag{4.4}$$

The unknown are the vertical separation of the surfaces, $h(x)$, the contact pressure, $p(x)$ and the integration constant L, which can be interpreted as the distance form an infinitesimal line load at which the vertical displacement vanishes ($L = |x - x'| \Rightarrow \ln(|x - x'|/L) \cdot p(x') = 0$).

By discretizing the x axis at equal intervals of Δx and assuming a constant pressure p_j within each segment $[x_j - \Delta x/2; x_j + \Delta x/2]$ (see Figure 4.2), the equations become:

$$h_i = h_{0,i} + L' - \sum_j p_j \cdot K_{|i-j|} \tag{4.5}$$

$$f_N = \Delta x \sum_j p_j \tag{4.6}$$

$$\forall i : \big(h_i \geq 0 \ \wedge \ p_i = 0\big) \vee \big(h_i = 0 \ \wedge \ p_i \geq 0\big) \quad \text{(see Figure 4.3)} \tag{4.7}$$

where:

$$K_i = \frac{2}{\pi E_{eq}} \Delta x \left[\left(\frac{1}{2} + i \right) \ln \left| \frac{1}{2} + i \right| + \left(\frac{1}{2} - i \right) \ln \left| \frac{1}{2} - i \right| \right] \tag{4.8}$$

$$L' = \frac{2}{\pi E_{eq}} f_N \left(1 + \ln \frac{L}{\Delta x} \right) \tag{4.9}$$

Thus, the nodal pressures p_i and separations h_i and the constant L' become the unknowns.

Initially, the algorithm to solve this optimization problem was that detailed in the work by Seabra and Berthe [41]. It was found that under certain circumstances, the

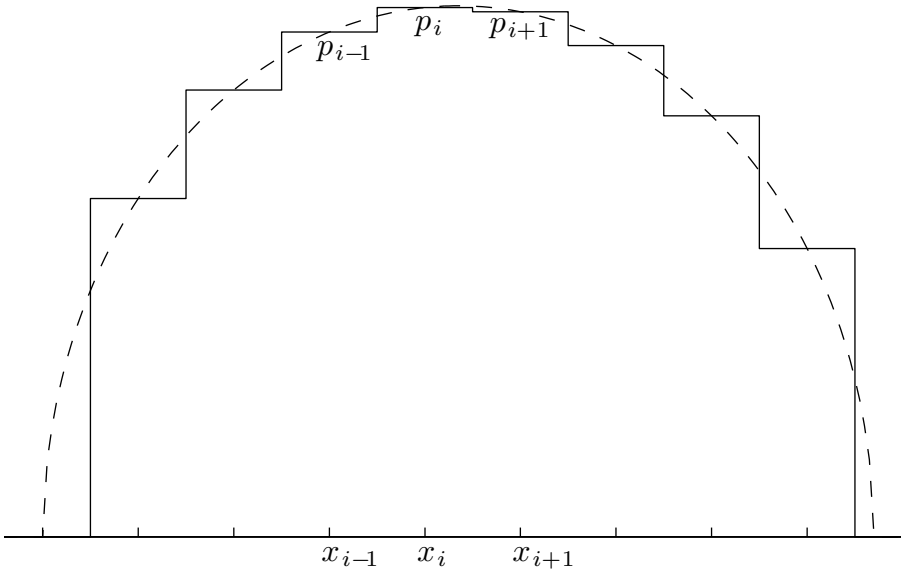


Figure 4.2: Discretization of the xx axis and of the surface pressure for the rough boundary film contact problem. The discretization is exaggerated for a better illustration.

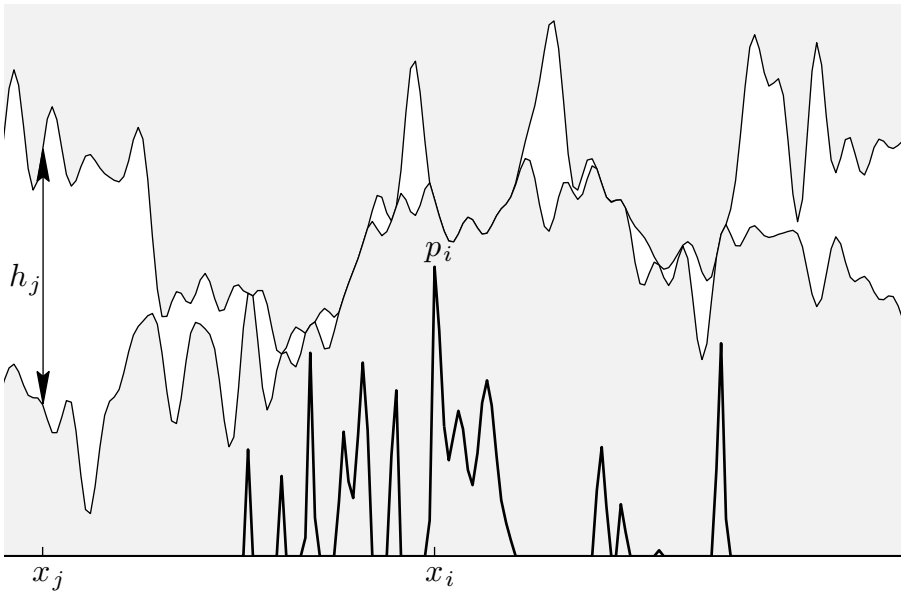


Figure 4.3: An example of valid deformed surfaces and surface pressures in a boundary film lubrication problem. As an example see the two discretization points i and j: $h_j > 0$ and $p_j = 0$; $h_i = 0$ and $p_i > 0$.

algorithm failed to converge on a solution. This is inconvenient in this case because thousands of contact problems must be solved in order to obtain a complete cycle with a sufficiently tight discretization grid. Another possibility would be the use of linear programming, but it was found to be excessively costly from a computational point of view. Because of this, a fast alternative had to be found, suitable to run automatically without any human intervention.

The algorithm presented by Polonsky and Keer in [42] was found to meet these requirements: it is fast and unconditionally stable. The algorithm as presented in [42] is described for a point contact, so that an adaptation had to be made to make it suitable for the line contact case. The algorithm is described here with a notation that is coherent with the previous equations and adapted to the two dimensional case.

This method is based on the conjugate gradient method, an optimization method usually used on unconstrained problems, modified to accommodate the constraints of Equations (4.6) and (4.7).

The process is started by attributing an arbitrary value to the nodal pressures, so long as it satisfies Equation (4.6). The parameter δ is initialized at 0 and G_{old} at 1. Their meaning shall be discussed later. The process then enters an iterative stage to be run until the error ϵ reaches a satisfactory value. The iterative stage is described mathematically as:

1. The nodal points that belong to the contact set I_c, the gaps h_i and the new "penetration" L' are determined. Notice that L' is chosen so that the average gap h within the contact set I_c is null. This is a weak form of enforcing the non negativity of the gap that ensures that the final solution will be correct within a tolerance.

$$I_c \leftarrow \{i : \ p_i > 0\} \tag{4.10}$$

$$N_c \leftarrow \text{size of } I_c \tag{4.11}$$

$$\forall i : \ h_i \leftarrow h_{0,i} - \sum_j K_{|i-j|} \cdot p_j \tag{4.12}$$

$$L' \leftarrow -\frac{1}{N_c} \sum_{i \in I_c} h_i \tag{4.13}$$

$$\forall i : \ h_i \leftarrow h_i + L' \tag{4.14}$$

2. The new conjugate gradient t_i is determined. It is seen now that G/G_{old} is the proportion in which the previous gradient enters in the composition of the new one and δ is 0 when the conjugate gradient becomes that of maximum

descent and 1 otherwise.

$$G \leftarrow \sum_{i \in I_c} h_i^2 \tag{4.15}$$

$$t_i \leftarrow \begin{cases} h_i + \delta \cdot \dfrac{G}{G_{\text{old}}} t_i & ; i \in I_c \\ 0 & ; i \notin I_c \end{cases} \tag{4.16}$$

$$G_{\text{old}} \leftarrow G \tag{4.17}$$

$$\tag{4.18}$$

3. The gradient is translated into the pressure domain and the non negativity of the pressures is enforced. If no nodal points will need to be added to the contact in the next iteration δ is set to 1. Otherwise it is set to 0.

$$\forall i : \; r_i \leftarrow \sum_j -K_{|i-j|} t_j \tag{4.19}$$

$$\bar{r} \leftarrow \frac{1}{N_c} \sum_{i \in I_c} r_i \tag{4.20}$$

$$\forall i : \; r_i \leftarrow r_i - \bar{r} \tag{4.21}$$

$$\tau \leftarrow \left(\sum_{i \in I_c} h_i t_i \right) \left(\sum_{i \in I_c} r_i t_i \right)^{-1} \tag{4.22}$$

$$\forall i : \; p_i^{\text{old}} \leftarrow p_i \tag{4.23}$$

$$\forall i \in I_c : \; p_i \leftarrow p_i - \tau t_i \tag{4.24}$$

$$\forall i : \; p_i < 0 \Rightarrow p_i \leftarrow 0 \tag{4.25}$$

$$I_{\text{ol}} \leftarrow \{ i : \; p_i = 0 \; \wedge \; h_i < 0 \} \tag{4.26}$$

$$\delta \leftarrow \begin{cases} 1 & ; I_{\text{ol}} = \emptyset \\ 0 & ; I_{\text{ol}} \neq \emptyset \end{cases} \tag{4.27}$$

$$\forall i \in I_{\text{ol}} : \; p_i \leftarrow p_i - \tau h_i \tag{4.28}$$

$$\tag{4.29}$$

4. The new pressure vector is scaled to balance the overall load and the error is computed.

$$P \leftarrow \Delta x \sum_j p_i \tag{4.30}$$

$$\forall i : \; p_i \leftarrow \frac{f_N}{P} p_i \tag{4.31}$$

$$\epsilon \leftarrow \frac{\Delta x}{f_N} \sum_i \left| p_i - p_i^{\text{old}} \right| \tag{4.32}$$

When the solution has sufficiently converged, the nodal values h_i and p_i as well as L' have been obtained.

Note that Polonsky and Keer advocate the use of a multi-level multi-summation scheme for the computation of $K_{|i-j|} \cdot p_j$ in their paper. It was found that in the two dimensional case treated here, this would introduce too much computational overhead in the programme and a simple direct matrix multiplication is used instead.

4.2.2 The computation of elastic stresses

In Section 3.1.2, the integral Equations (3.2), (3.3) and (3.4) for the calculation of σ_{xx}, σ_{zz} and σ_{xz} were presented without reference to an actual computation method, which is presented now.

No generally valid analytic formula exists for the computation of the aforementioned equations, so that numerical methods must be employed. Thus, the elastic half-space, whose coordinate system is shown in Figure 3.1, must be discretized. The xx axis maintains the same discretization as in the previous section. On the zz axis, z_i coordinates are positionned with a constant distance Δz from one another. This coordinate system is fixed on the tooth: it does not change as the contact travels on the tooth surface. As before, the surface pressure and shear stress p and τ are considered constant within each segment of the xx axis $[x_j - \Delta x/2; x_j + \Delta x/2]$.

Thus, the equations for the determination of the elastic stresses become:

$$\sigma_{xx}^{ij} = \sum_k p_k K_{i,j-k}^{21} + \tau_k K_{i,j-k}^{30} \tag{4.33}$$

$$\sigma_{zz}^{ij} = \sum_k p_k K_{i,j-k}^{03} + \tau_k K_{i,j-k}^{12} \tag{4.34}$$

$$\sigma_{xz}^{ij} = \sum_k p_k K_{i,j-k}^{12} + \tau_k K_{i,j-k}^{21} \tag{4.35}$$

where:

$$\sigma_{xx}^{ij} = \sigma_{xx}\left(x_j, z_i\right) \tag{4.36}$$

$$\sigma_{zz}^{ij} = \sigma_{zz}\left(x_j, z_i\right) \tag{4.37}$$

$$\sigma_{xz}^{ij} = \sigma_{xz}\left(x_j, z_i\right) \tag{4.38}$$

and:

$$\begin{aligned} K_{ij}^{03} = &- \frac{1}{\pi}\left(\arctan\frac{(1/2-j)\,\Delta x}{z_i} + \arctan\frac{(1/2+j)\,\Delta x}{z_i}\right) - \\ &- \frac{1}{\pi}\left(\frac{(1/2-j)\,\Delta x z_i}{\left[(1/2-j)\,\Delta x\right]^2 + z_i^2} + \frac{(1/2+j)\,\Delta x z_i}{\left[(1/2+j)\,\Delta x\right]^2 + z_i^2}\right) \end{aligned} \tag{4.39}$$

$$K_{ij}^{12} = - \frac{1}{\pi}\left(\frac{z_i^2}{\left[(1/2-j)\,\Delta x\right]^2 + z_i^2} - \frac{z_i^2}{\left[(1/2+j)\,\Delta x\right]^2 + z_i^2}\right) \tag{4.40}$$

$$K_{ij}^{21} = -\frac{1}{\pi}\left(\arctan\frac{(1/2-j)\,\Delta x}{z_i} + \arctan\frac{(1/2+j)\,\Delta x}{z_i}\right) + \\ + \frac{1}{\pi}\left(\frac{(1/2-j)\,\Delta x z_i}{\left[(1/2-j)\,\Delta x\right]^2 + z_i^2} + \frac{(1/2+j)\,\Delta x z_i}{\left[(1/2+j)\,\Delta x\right]^2 + z_i^2}\right) \tag{4.41}$$

$$K_{ijk}^{30} = \frac{1}{\pi}\left(\frac{z_i^2}{\left[(1/2-j)\,\Delta x\right]^2 + z_i^2} - \frac{z_i^2}{\left[(1/2+j)\,\Delta x\right]^2 + z_i^2}\right) + \\ + \frac{1}{\pi}\ln\frac{\left[(1/2-j)\,\Delta x\right]^2 + z_i^2}{\left[(1/2+j)\,\Delta x\right]^2 + z_i^2} \tag{4.42}$$

In view of this, the determination of the stresses could be performed as a series of matrix multiplications. The reason this is not the method adopted in this work is the staggering amount of data that needs to be treated. A surface discretization of a tooth with an 8 mm surface length with $\Delta x = \Delta z = 2\mu m$ yields a total of $3 \cdot 4001 \cdot 21 \cdot 4001 \approx 10^{12}$ elastic components σ_{xx}, σ_{zz} and σ_{xz} to be obtained in the course of a complete cycle. This is by far the most computationally intensive part of the model. For this reason, fast alternatives to direct matrix multiplication had to be found.

One possibility is suggested by the convolution theorem that states that the Fourier transform of a convolution of two function is equal to the product of the Fourier transforms of each of these functions:

$$\mathcal{F}\left(\int_{-\infty}^{+\infty} f\left(x - x'\right) g\left(x'\right)\, dx'\right) = \mathcal{F}\left(f\right) \cdot \mathcal{F}\left(g\right) \tag{4.43}$$

Thanks to the existence of very fast algorithms for obtaining the discretized Fourier transforms—the FFT algorithm prominent among them—it is less costly to obtain the Fourier transforms of the functions, multiply them and then perform the inverse Fourier transformation on the result than it is to multiply the matrices. Unfortunately, this is partly offset by the need to extend the domain of the functions in order to avoid a distortion of the results caused by the periodic nature of the discretized Fourier transform.

Another possibility, and the one used in this work, is that of a multi-level multi-integration scheme, as described by Polonsky and Keer in [42] as well as Venner in [43]. For a comprehensive theoretical overview of the method, the reader is advised to consult the latter reference. The method is described as a mean of obtaining the surface deflection under a surface pressure but it can equally well be used for obtaining the stresses within the tooth.

In broad terms, the method is based on the Saint-Venant principle, which states that, at a sufficient distance from the external loads, any statically equivalent load produces the same stress state. This law is exploited to compute the stresses in a discretization grid coarser than the targeted one and then transfer the values onto ever finer grids, with corrections as needed, until the stresses are obtained for the targeted grid.

The computed matrix multiplication is not exact, but it's error can be made inferior to that introduced by the discretization of the continuous integral expressions.

The transfer of values from one grid to another is effected by Lagrangean polynomial interpolations, whose coefficients are:

$$\forall i : 1 \leq i \leq 2t : \; s_i^t = \prod_{k=1;k\neq i}^{2t} \frac{2\,(t-k)+1}{2\,(i-k)} \tag{4.44}$$

To ensure that the error never exceeds that of the discretization, a correction region is defined whose width is:

$$m_c = 0.7tN_x^{1/t} - 1 \tag{4.45}$$

where $2m_c - 1$ are the number of nodes in the correction region.

Thus, both the size of the correction region and the order of the Lagrange polynomial derive from the definition of t. In this work, t has always been chosen in order to minimize m_c.

In order to allow the reader to implement the algorithm, the computation of the part of σ_{zz} caused by p is detailed as an example. In all that follows, an index c will denote a coarse grid and an index of f a fine one. Additionally, all indexes i,j, etc... are numbered from 0. Thus, all grid nodes will have coordinates (x_j, z_i) such that:

$$0 \leq i \leq N_z - 1 \tag{4.46}$$

$$0 \leq j \leq N_x - 1 \tag{4.47}$$

First of all, the surface pressure matrix p_i is transfered to successively coarser grids iteratively. A coarse grid is obtained from a finer one by doubling the node spacing in xx:

$$\Delta x^c = 2\Delta x^f \tag{4.48}$$

and the number of nodes in the coarser grid is:

$$N_x^c = N_x^f/2 + 2t - 1 \tag{4.49}$$

The pressures of the fine grid are lumped on the nodes of the coarse grid:

$$p_m^c = p_{2(m-t+1)}^f + \sum_{k=1}^{2t} s_k^t \cdot p_{2(m-t+1)+2(k-t)-1}^f \tag{4.50}$$

Figure 4.4 shows the passage from a fine grid to a coarser one. Note that the domain of discretization grows with the coarseness of the grid. This is to ensure that the same Lagrangean function can be used at the limits of the discretization grids as in the interior of the grid for the transfer. This effect is greatly exaggerated in the figure due to the small number of nodes of the original grid. Nevertheless, a careful reader might enquire if this doesn't void the purported advantage of this method over an FFT scheme. This is not so, because although the domain is enlarged, the

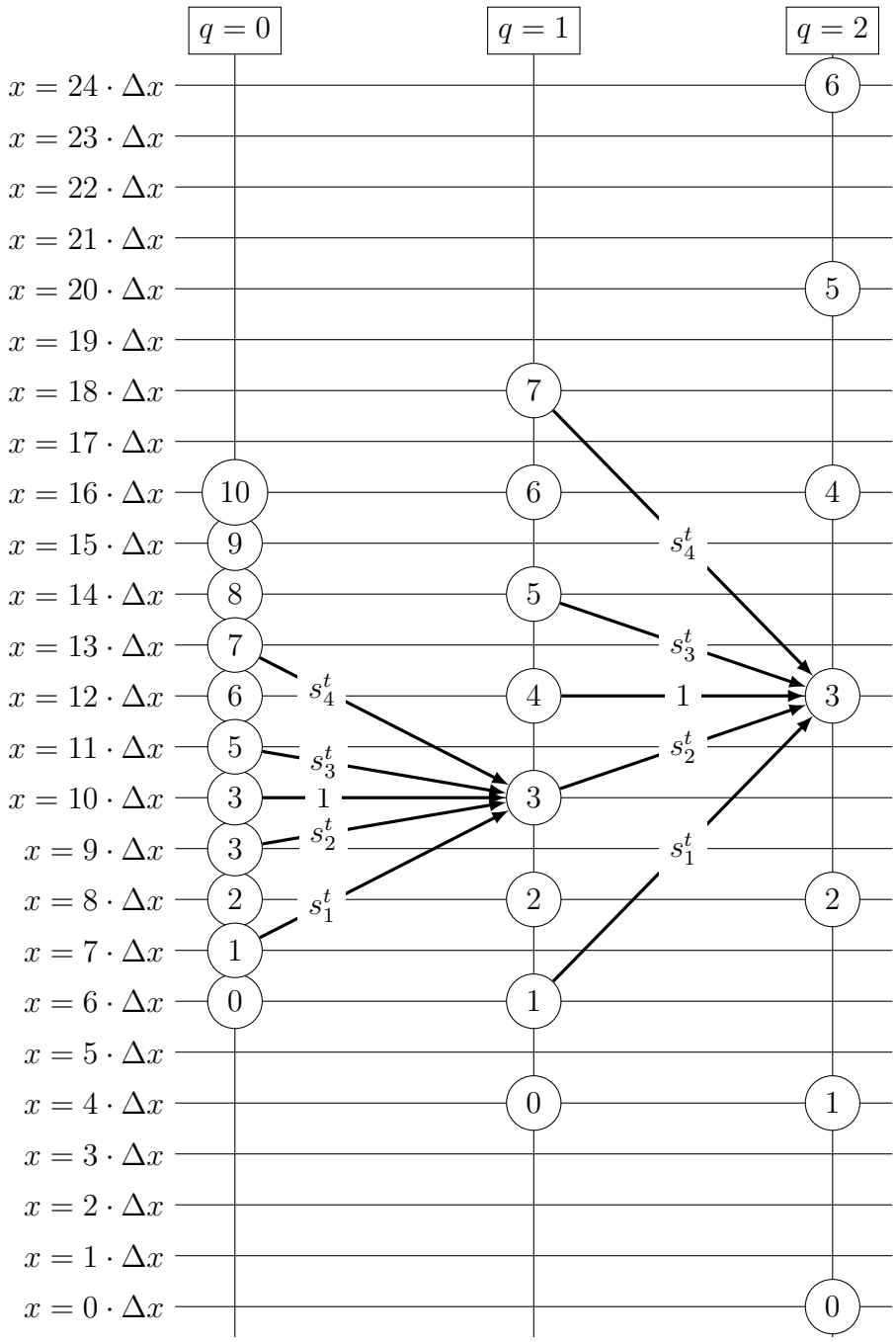


Figure 4.4: An illustration of the transfer of surface pressure from a fine grid to a coarser one when $t = 2$. The left column of nodes shows the position of the nodes of the finest grid—the original one: $q = 0$—on the xx axis. The middle column the position of the nodes of the second finer grid ($q = 1$). The right column the position of the nodes of the next grid ($q = 2$). As an example, the arrows show how the pressure on the nodes $i = 3$ of the grids $q = 1$ and $q = 2$ are obtained from the pressures on the next finer grids.

number of nodes decreases as the grid coarsens. Figure 4.5 shows the grids and pressures obtained from the application of the process shown in Figure 4.4. Only the discretization in the xx axis varies. A similar scheme could have been applied to the zz direction but the increase in complexity in the algorithm cannot be justified, since the depth of tooth analysed in this work is typically two orders of magnitude smaller than length. Note that the pressures from a fine grid are lumped in the next coarser on, so that the segment of application of a constant pressure still has a width of Δx. This ensures that the kernel doesn't need not be recalculated for each level of discretization.

This procedure is repeated until a grid is arrived at that is sufficiently coarse to allow for an inexpensive direct multiplication. The kernel matrix of the coarsest grid is obtained from the original one as follows:

$$\left(K^{03}\right)^{c}_{i,j} = K^{03}_{i,2^q j} \tag{4.51}$$

where q is the level of the coarsest grid—$q = 0$ corresponds to the original grid. And the stress is thus:

$$(\sigma_{zz})^{c}_{ij} = \sum_{k} p_k \cdot \left(K^{03}\right)^{c}_{i,j-k} \tag{4.52}$$

The computed stress must now be transferred sequentially to successively finer grids. This is effected by an iterative process that "climbs back" from the coarsest grid to the original one. Each iteration deals with two grids: a coarse grid and the next finer one. The coarse grid of one iteration is the fine grid of the previous one.

The stress in the coarse grid is first corrected. The correction coefficients are computed from the kernel of the next finer grid for every i and $-m_c \leq j \leq m_c$:

$$C^{(1)}_{ij} = \begin{cases} 0 & ;\text{ even } j \\ -K^{f}_{ij} + \sum_{k=1}^{2t} s^{t}_{k} \cdot K^{f}_{i,j-2(k-t)+1} & ;\text{ odd } j \end{cases} \tag{4.53}$$

Then the correction is applied to the stresses in the coarse grid:

$$(\sigma_{zz})^{c}_{ij} \leftarrow (\sigma_{zz})^{c}_{ij} + \sum_{k=-m_c}^{m_c} C^{(1)}_{ik} \cdot p^{f}_{2(j-t+1)-k} \tag{4.54}$$

Next, the stress on the fine grid is computed by interpolation (see Figure 4.6):

$$(\sigma_{zz})^{f}_{ij} = \begin{cases} (\sigma_{zz})^{c}_{i,(j+1)/2+t-1} & ;\text{ even } j \\ \sum_{k=1}^{2t} s^{t}_{k} \cdot (\sigma_{zz})^{c}_{i,(j+1)/2+k-2} & ;\text{ odd } j \end{cases} \tag{4.55}$$

From this point on, the coarse grid is abandoned and all computation is done on the fine grid. First, as was done for the coarse grid, a correction step is applied, whose coefficients are computed for every i and $-m_c \leq j \leq m_c$:

$$C^{(2)}_{ij} = \begin{cases} 0 & ;\text{ even } j \\ -K^{f}_{ij} + \sum_{k=1}^{2t} s^{t}_{k} \cdot K^{f}_{i,j+2(k-t)-1} & ;\text{ odd } j \end{cases} \tag{4.56}$$

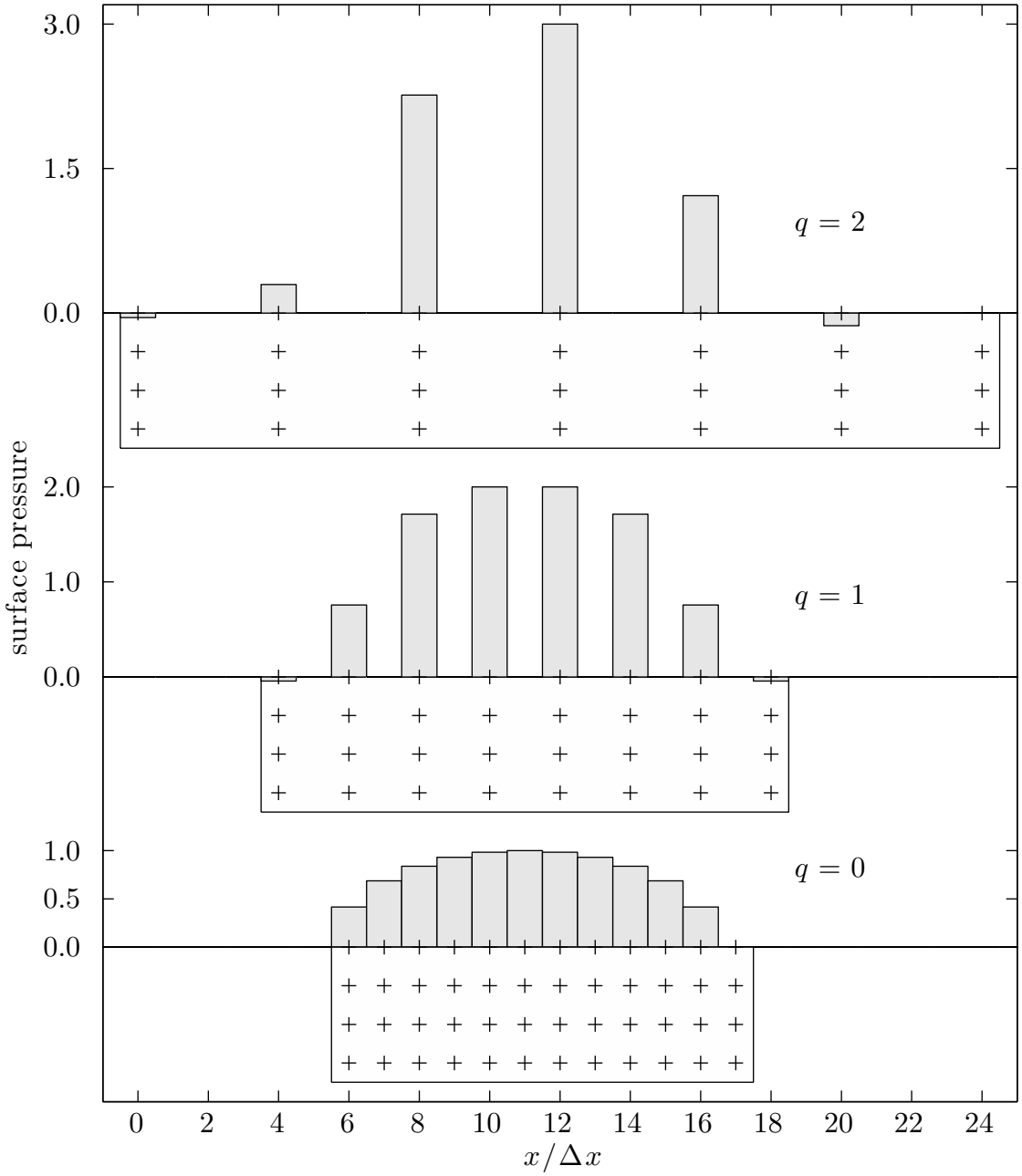


Figure 4.5: Transition from fine to coarse grids. The change in the node interval and size of the discretization domain is shown, as well as the lumping of the surface pressures. Note that the discretization of the zz axis doesn't change from one grid to another.

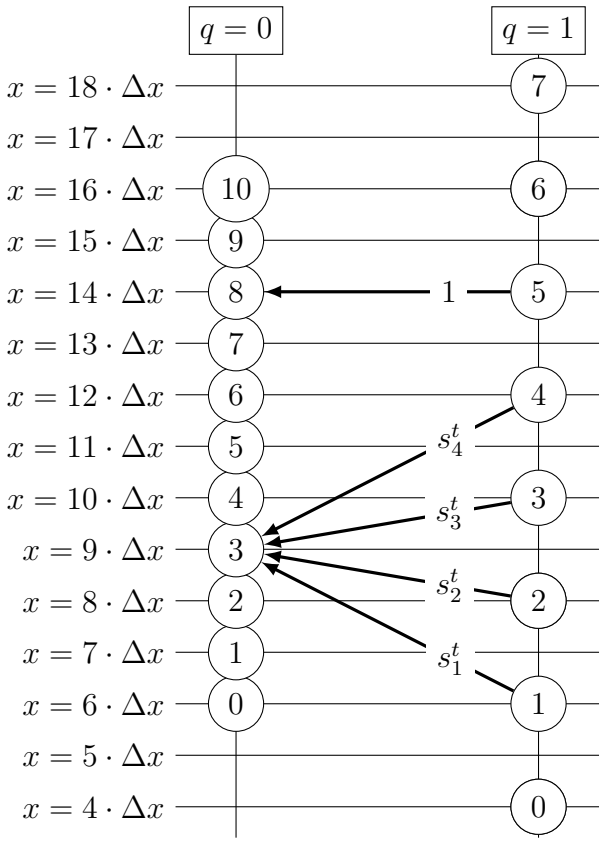


Figure 4.6: An illustration of the transfer of stresses from a coarse grid to a fine one when $t = 2$. The stresses in the even nodes of the fine grid are transferred from the coarse grid as node 8 is. The stresses in the odd nodes are transferred from the coarse grid as node 3 is.

Then the correction is applied to the fine stress:

$$(\sigma_{zz})^f_{ij} \leftarrow (\sigma_{zz})^f_{ij} + \sum_{k=-m_c}^{m_c} C^{(2)}_{ik} \cdot p^f_{j-k} \tag{4.57}$$

Operations (4.53) to (4.57) constitute an iteration. At the end of the iteration, if the fine grid is not the original, targeted grid, then it is passed on to the next iteration as its coarse grid. Thus, the iteration is applied until the fine grid is the original one, at which point the stresses have been computed.

This example only describes how to obtain the portion of σ_{zz} caused by p, but it is easy to generalize the method in order to obtain all three stress components caused by both p and τ simultaneously. Indeed, by a judicious fusing and reshaping of the stress and kernel matrices, it is possible to deal only with two kernels and one stress matrix with a corresponding increase in the speed of the method.

4.2.3 The computation of the mesoscopic residual stress tensor

As discussed in Section 3.1.3, the calculation of the mesoscopic stresses hinges on the computation of a mesoscopic residual stress tensor. This optimization problem was stated in Equation (3.31) reproduced here:

$$\begin{aligned} K^2 &= \min_{\rho'} \Big(\max_t \big(-J_2 \left(\sigma \left(t \right) + \rho' \right) \big) \Big) \\ \rho : \; K^2 &= \max_t \big(-J_2 \left(\sigma \left(t \right) + \rho \right) \big) \end{aligned} \tag{4.58}$$

Replacing J_2 by its explicit expression and recalling that ρ' must necessarily be a purely deviatoric stress tensor:

$$\begin{aligned} -J_2 \left(\sigma \left(t \right) + \rho' \right) = &\left(\frac{s_{xx} - \rho'_{xx}}{\sqrt{2}} \right)^2 + \left(\frac{s_{yy} - \rho'_{yy}}{\sqrt{2}} \right)^2 + \\ &+ \left(\frac{s_{zz} - \rho'_{zz}}{\sqrt{2}} \right)^2 + \left(s_{xy} - \rho'_{xy} \right)^2 + \left(s^2_{yz} - \rho'_{yz} \right) + \left(s_{xz} - \rho'_{xz} \right)^2 \end{aligned} \tag{4.59}$$

where, as before, s_{xx}, s_{yy}, s_{zz}, s_{xy}, s_{yz} and s_{xz} are the components of the deviatoric macroscopic stress tensor. Note that this macroscopic stress tensor is the sum of the initial stresses with the elastic stresses.

In appendix 1 of his paper [33], Dang Van described an iterative scheme for the approximate calculation of ρ. The same procedure is described by Ciavarella et al. in [37].

This method, that will not be described in detail here, consists in applying numerically the physical reasoning presented in Section 3.1.3, page 51, and in Figure 3.4 to the successive macroscopic stress tensors at discrete intervals in the cycle. From a numerical point of view, the outstanding feature of this method is that the quality of the provided solution is controlled by the definition of a parameter χ. Dang Van

recommends a value of 0.05 for χ, although he also states that the error decreases with lower χ at the expense of a higher computation time.

Dang Van's algorithm was initially implemented in the course of this work but was later abandoned because of difficulties with its accuracy and convergence. A numerical experiment was run in which the stress state over time remained the same for a tooth, but χ and the discretization of the tooth were varied—note that a finer discretization of a tooth entails also a larger number of discrete instants in the cycle.

It was found that the asymptotic rate of convergence of the method is very low, so that the convergence tolerance TOL must be set very low when using a fine discretization grid: otherwise, the algorithm could stop when detecting a false convergence. Moreover, the rate of convergence decreases notably with an increase in the fineness of the discretization and the value of χ must be lowered accordingly. Even worse, it was found that at very small values of χ, the algorithm becomes entirely erratic and the converged solution is composed wholly of random noise. A case in point was a simulation attempted with a discretization step on the surface of 0.5 μm, in which no solution could be found. Even in the cases where a solution can be obtained , the interplay between the discretization and the pair of parameters (TOL, χ) is complex and proper values for these parameters must be found by trial and error on a case by case basis. This is obviously undesirable when, as is the case here, an automatic calculation method—one without user input—is needed.

It has already been stated that Equation (4.59) amounts to the smallest enclosing ball problem when mapping $s_{xx}/\sqrt{2}$, $s_{yy}/\sqrt{2}$, $s_{zz}/\sqrt{2}$, s_{xy}, s_{yz} and s_{xz} to a six-dimensional space. In the specialized literature on computational geometry, two solution procedures were found for this problem.

The first, by Kumar et al., is presented in [44]. A *Matlab* implementation of the method is available at the web-site [45]. While very reliable, the computation time of the programme proved excessive for the needs of this work.

The second solution procedure, and that used in this work, is presented by Gärtner in [46]. Conveniently, a *C* computer language implementation of the algorithm is provided at the web-site [47]. The method will not be discussed here in detail, as it falls somewhat far from the core concerns of this work and the author only has a very superficial grasp of its inner workings. Suffice to say that it has proved to be fast, robust and reliable. The solution is always exact relatively to the data set provided to the programme: thus it is completely insensitive to the discretization of the problem.

5 Simulation of a gear micropitting test

This chapter presents a simulation of a real gear micropitting test performed by Cardoso [8] as part of his Master's Thesis work.

5.1 Description of the test

In his work [8], Cardoso performed an extensive series of test in order to characterize the mechanical performance of several gear oils. Of interest to the present work is his micropitting test preformed with carburized gears lubricated with a mineral oil. Note that this test was not designed with the simulation of micropitting in mind, but was later co-opted for this purpose by the author of the present work, so that many important measurements—important from the point of view of obtaining data for the simulation, that is—were omitted from the tests. This is no criticism of Cardoso's excellent work, but serves instead to underline the need for dedicated tests in the future.

5.1.1 FZG test rig and gears

The gear micropitting test was conducted on an FZG test rig, shown in Figure 5.1

Cardoso presents in his work a detailed explanation of the operation of the machine that will therefore be omitted here.

More immediately relevant for the present work are the geometric parameters of the gears used in the test, FZG type C gears, listed in Table 5.1. A photograph of the gears is shown in Figure 5.2.

5.1.2 Lubricant oil and gear material

The gear oil is an ISO VG 150 paraffinic mineral oil that contains additives to increase its resistance to micropitting. Its properties are reproduced from [8, Table 1.1] in Table 5.2. As can be seen, it gives no information about the parameters to introduce into the Roelands viscosity equation, or about the limiting shear stress. This issue will be addressed in a later section.

The material of which the gears were made is a DIN 20 MrCr 5 carburizing steel. The gears were carburized, quenched and annealed before the final grinding operation. The heat treatment generated a curburized layer of a depth of approximately

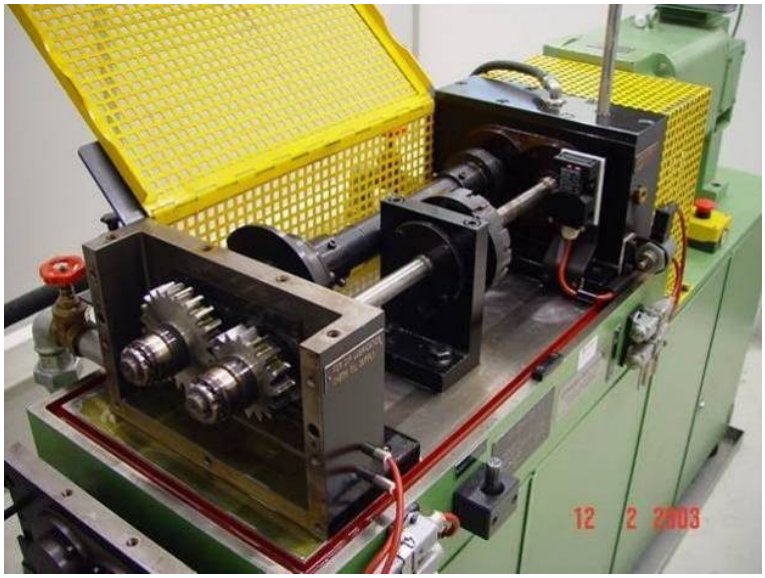


Figure 5.1: Photograph of an FZG test rig.

Table 5.1: Geometric and kinematic parameters of an FZG type C gear.

parameter	**driving gear**	**driven gear**
face width	$b = 14\ mm$	
operating centre distance	$a' = 91.5$ mm	
number of teeth	$Z_1 = 16$	$Z_2 = 24$
rotational speed	$n_1 = 2250$ RPM	$n_2 = 1500$ RPM
base radius	$R_{b1} = 33.829$ mm	$R_{b2} = 50.7435$ mm
addendum radius	$R_{a1} = 41.3175$ mm	$R_{a2} = 59.2715$ mm
distances on contact path, see Figure 1.9	$\overline{T_1A} = 4.2946$ mm $\overline{T_1B} = 10.4374$ mm $\overline{T_1C} = 13.9699$ mm $\overline{T_1D} = 17.5792$ mm $\overline{T_1E} = 23.7220$ mm $\overline{T_1T_2} = 34.9248$ mm	

Table 5.2: Gear oil properties.

Parameter	**Method**	**Desig.**	**Lubricating Oils**
Lubricant maker Country			EXXON MOBIL USA
Oil type			Industrial gear oil
Base oil	DIN 51451		Paraffinic mineral oil

Chemical Content

Zinc	ASTM D 4927	Zn	-
Calcium	ASTM D-4927	Ca	40 ppm
Phosphor	ASTM D-4927	P	175 ppm
Sulphur	ASTM D-4927	S	15040 ppm

Physical properties

Density at 15 °C	DIN 51757	ρ_{15}	$0.894\,\mathrm{g\,cm^{-3}}$
Kinematic Viscosity at 40 °C	DIN 51562	ν_{40}	153.6 cSt
Kinematic Viscosity at 100 °C	DIN 51562	ν_{100}	14.4 cSt
Viscosity Index	DIN ISO 2909	VI	96
Pour point	DIN ISO 3106		−27 °C

Viscosity and density equations

Viscosity $\ln(\ln(\nu + a)) = m - nT$	ASTM D341	n m a	3.5268 9.1413 0.7 cSt
Thermoviscosity $\beta = \frac{m}{T}\frac{(\nu+a)\ln(\nu+a)}{\nu}$		β_{40} β_{100}	0.0570 cSt/°C 0.0269 cSt/°C
Piezoviscosity $\alpha = (a_1 + a_2 T + (b_1 + b_2 T)\,p)^{-1}$		a_1 a_2 b_1 b_2	367 bar 2.96 bar/°C $9.98 \cdot 10^{-3}$ $2.07 \cdot 10^{-4}$ °C^{-1}
Thermal expansion		α_ν	$8.26 \cdot 10^{-4}$ g/cm^3/°C

Wear properties

KVA weld load	DIN 51350-2		2200 N
KVA wear scar (1h/300N)	DIN 51350-3		0.32 mm
Brugger	DIN 51347-2		68 N/mm^2
FZG rating	DIN 51354	K_{FZG}	> 12

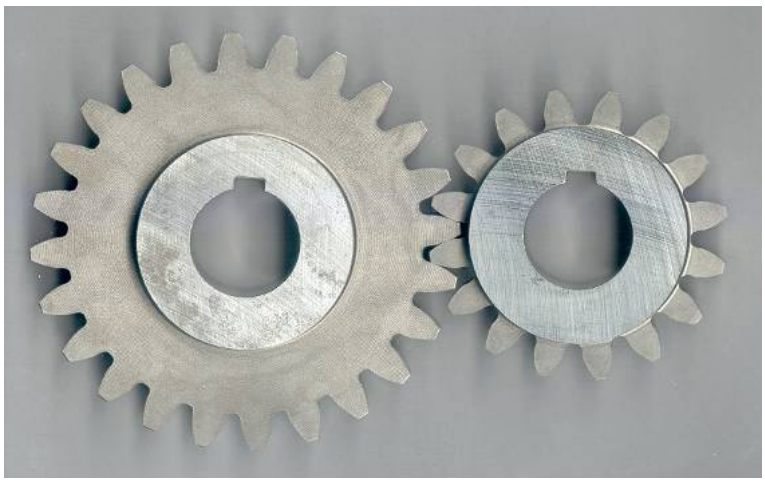


Figure 5.2: Photograph of a pair of FZG type C gears.

Table 5.3: Physical properties of the gear steel.

property	**value**
Young's modulus of elasticity	210 GPa
Poisson ratio	0.3
Density	$7850\,\mathrm{kg\,m^{-3}}$
Thermal conductivity	$42\,\mathrm{W\,m^{-1}\,K^{-1}}$
Specific heat capacity	$0.46\,\mathrm{kJ\,kg^{-1}\,K^{-1}}$

0.8 mm. The surfaces had a very high hardness, typically around 60 HRC, while the core kept a very high strength and a considerably high elongation. The relevant physical properties of the steel are listed in Table 5.3.

5.1.3 Gear micropitting test procedure

The test procedure based on an FZG specification (DGMK micropitting short test—GFKT–C/8.3/90) is as follows:

Table 5.4: Operating conditions of the micropitting test.

load stage	K3	K6	K8	K9
rotational speed	$n_1 = 2250$ RPM ; $n_2 = 1500$ RPM			
normal load F_N	851.34 N	2923.5 N	5072.6 N	6373.2 N
oil temprerature	80 °C	90 °C		
duration	1 h	16 h		
number of cycles undergone by the driving gear	$135 \cdot 10^3$	$2.16 \cdot 10^6$		

1. The test gears were cleaned in an utlrasonic bath, dried, weighed and their roughness were measured.
2. They were mounted on the FZG test rig and submitted to the load stage K3 (a running-in stage), during 1 hour.
3. Without removing the gears from the test rig, a lubricant sample was taken.
4. They were immediately submitted to load cycle K6 during 16 hours.
5. At the end of the K6 stage, lubricant samples were collected, the gears were dismounted, cleaned and weighed, the roughness of the driving gear was measured and the surfaces of the gears were photographed.
6. They were then remounted on the test rig and subjected to load stage K8 during 16 hours.
7. The same steps were performed as at the end of load stage K6.
8. Once more, the gears were mounted and subjected to load stage K9 during 16 hours.
9. A last round of measurements were performed as at the end of load stages K6 and K8.

Table 5.4 lists the operating conditions for each load cycle. Note that the gears were dip lubricated and the bath temperature of the oil was maintained constant.

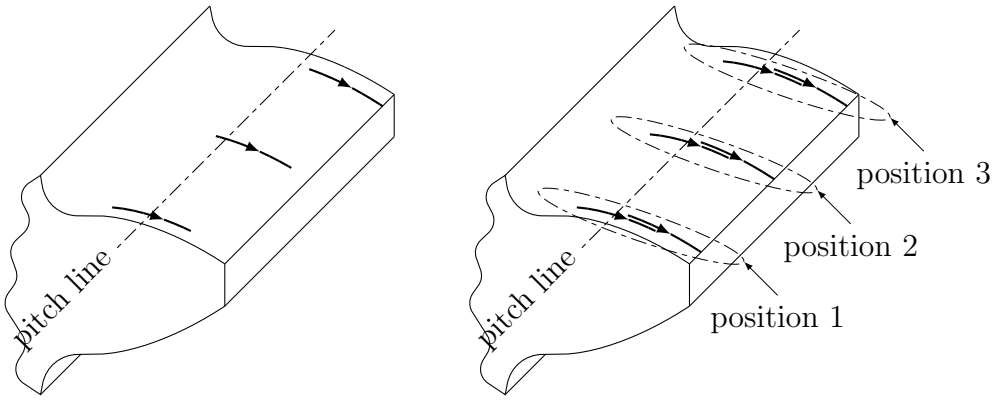
Figure 5.3 shows the approximate positions where the roughness measurements were made. Each series of measurements was performed on two distinct teeth. The left image shows the measurements performed before load stage K3. Each of the two teeth was submitted to three separate measurements for a total of six. The right image shows the measurements between successive load cycles (K6–K8 and K8–K9). Each tooth was measured thrice below the pitch line and thrice above it, for a total of twelve distinct measurements.

5.1.4 Test results

Some results of the test are shown in Figures 5.4 to 5.6, which were taken from [8].

Figure 5.4 clearly demonstrates that micropits became visible at the end of load stage K8. This is confirmed by the roughness measurements in Figure 5.6. It is interesting to note that the axial measurements demonstrate much more clearly the presence of micropitting than the radial ones. This is due to the fact that the roughness is much smaller in the axial direction and the micropits thus stand out vividly. Unfortunately, the model does not deal with the axial direction, since it is two-dimensional.

At the end of load stage K9, the topography of a 2 mm by 2 mm square portion of the driving tooth flank surface was measured by Cardoso. The results are illustrated



(a) Before load stage K3 (b) Between load stages

Figure 5.3: Positions of roughness measurement. a) the roughness measurements before stage K3. b) roughness measurements between load stages: each position includes one measurement below the pitch line and one above it.

(a) After load stage K3+K6 (b) After load stage K8 (c) After load stage K9

Figure 5.4: Photographs of a teeth taken after each load stage of the micropitting test. The areas were micropitting has occurred are delimited with red lines.

in Figure 5.5, where the dark "stains" are the outlines of the micropits whith depths more profound than 5 μm. It is significant that the micropits are longer in the horizontal direction than in the vertical one. The vertical direction corresponds to the zz axis in the model. This is due to the fact that the roughness is much less pronounced in the horizontal—or axial— direction.

5.2 Description of the simulation process

For each combination of a boundary friction coefficient $\mu^{\mathrm{BDR}} \in (0.07, 0.08, \ldots, 0.15)$ with a fatigue property $\beta_{\mathrm{DV}} \in (400, 420, \ldots, 900)$ and with a position of roughness mesurement (see Figure 5.3) a simulation was performed as follows:

1. Introduce the initial roughness into the model

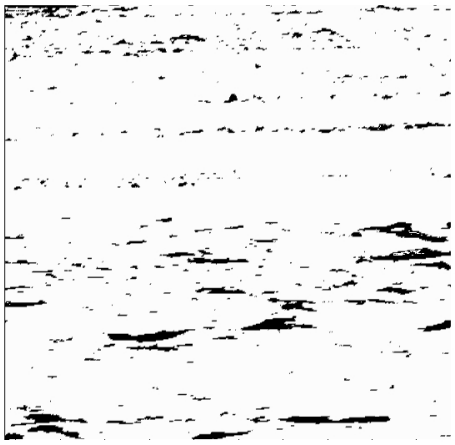


Figure 5.5: Micropits deeper than 5 μm on a 2 mm by 2 mm square portion of the gear flank surface at the end of load stage K9.

2. Run the simulation of load stage K6 as shown in Figure 4.1.

3. Introduce the roughness measured at the end of stage K6 into the model.

4. Run the simulation of load stage K8.

5. Introduce the roughness measured at the end of stage K8 into the model.

6. Run the simulation of load stage K9.

The properties that were considered are given explicitly in the next section.

5.3 Input data values and remarks concerning the realism of the simulation

5.3.1 Physical simplifications

Several simplifications in the representation of the physical phenomena involved in gear meshing were made. Their enumeration belongs perhaps more properly in the previous chapter, as most of them are features of the model. Nevertheless, it is convenient to place here this enumeration to enable an easy comparison with the testing conditions.

The simplifications may be divided into two categories: the first concerning intrinsic properties of the model, the second reflecting the particular conditions of the test.

The simplifications inherent to the model are:

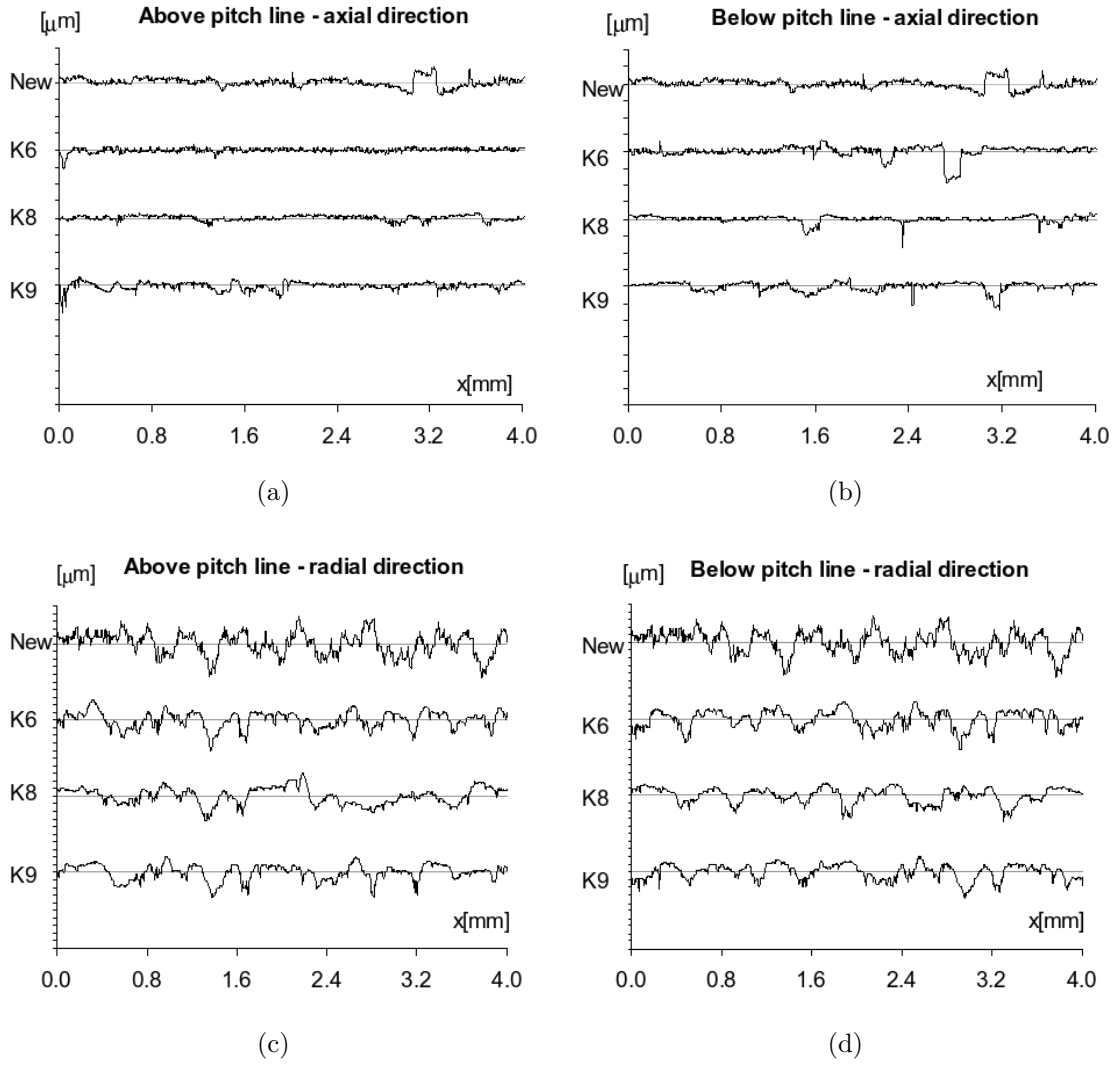


Figure 5.6: Evolution of the roughness during the micropitting test. a) and b) were measured in the axial direction—the direction perpendicular to those shown in Figure 5.3. c) and d) were measured in the directions shown in Figure 5.3.

- Because the roughness of a spur gear tooth is orthotropic and is so much more pronounced in the transversal direction than in the longitudinal one—about one order of magnitude—it was deemed acceptable to eliminate altogether the consideration of longitudinal roughness from the model. This allows for the treatment of the meshing of the teeth as a bi-dimensional one in plane strain.

- At the start of each cycle, the corresponding new roughness measurements are introduced into the model. Apart from that, no transfer of surface damage—for instance, invisible micro cracks—is effected from one load cycle to the next.

- The roughness change in the course of a load cycle is disregarded in the calculations of each load cycle. Thus wear and fatigue cracks occurring during a load cycle are only taken into account at the start of the next load cycle, when new roughness measurements are introduced into the model as input data.

The simplifications caused by the conditions of the test are:

- Any pair of driving gear teeth meshes with three driven gear teeth in the course of gearing. Thus, to study one driving gear tooth a full cycle should include the meshing with each of the three driven gear teeth with which it contacts. Because no measurement of driven gear tooth was performed at the end of the load cycles, it has been assumed here that the roughness of the driven gear teeth must be roughly the same as that of the driving gear teeth. This contradicts the established fact that driving gears suffer more wear and fatigue than driven ones. Nevertheless, the equality of the roughness had to be accepted as a work basis to enable the continuation of this work. There are twelve measurements of driving gear teeth roughness, thus at the start of each cycle, one of these measurements was affected to all driving gear teeth and another to all driven gear teeth. As a consequence, a load cycle contains only one distinct meshing of gear teeth.

- A roughness measurement does not cover the whole length of a tooth, so that distinct measurements where performed by Cardoso to cover the tooth flank below and above the pitch line. Because there is no way to position a measurement in relation to another precisely, the merging of these measurements in order to form a roughness image of the entire length of a tooth is unreliable at best.

- Because residual stresses were not measured in the tests performed by Cardoso, *initial* residual stresses were taken from the literature [48]. No alteration in the macroscopic residual stresses between successive load cycles were taken into account. Instead, the same initial residual stressses were applied at the start of each load cycle. There is a good argument in favour of this stance.

Table 5.5: Initial residual stresses due to the manufacturing processes (cutting, heat treatment, grinding).

depth [μm]	σ_{xx} [MPa]	σ_{yy} [MPa]	p_H [MPa]
0	−287	−287	−191
10	−366	−366	−244
30	−463	−463	−309
50	−164	−164	−109

Supposing that the actual residual stresses produced during a load cycle are closely approximated by the mesoscopic residual stress, then they can have no influence on the Dang Van criterion, since they are purely deviatoric. This leaves only the initial residual stresses produced by the manufacturing process as an influence on Dang Van fatigue. Thus, and provided that thermal loads are kept sufficiently small that they do not influence yield, as is the case here, this approach should be sufficiently accurate for practical purposes. The initial stresses used are listed in Table 5.5.

- No roughness measurement was made between load stages K3 and K6. Because of this it was decided that a simulation of load stage K6 would be run with the initial roughness, while disregarding completely the running-in load stage K3. This is problematic, as it is well known that roughness reduces considerably during running-in, with a considerable beneficial effect on the fatigue behaviour of the gears. Nevertheless, this was the lesser evil, as the alternative would be not to simulate the K6 load stage at all.

5.3.2 Properties related to mixed lubrication

It has already been said that the properties listed in Table 5.2 do not give sufficient information to determine the non-linear rheology of the lubricant oil. In fact, the data needed to model the gear oil is listed now:

- The oil parameters to be used in the Roelands viscosity Equation (2.9): η_0, S_0 and Z.
- The parameters to be used in the limiting shear stress Equation (2.17): τ_{L0}, α_{τ_L} and β_{τ_L} in order to use the Bair and Winer visco-plastic non-linear rheology Equation (2.16).

Additional data needed by the model are:

- The boundary lubrication regime friction coefficient μ^{BDR}.
- The load sharing function $f_\Lambda(\Lambda)$. A valid function must satisfy the conditions:

– with very high specific film thickness, there is only smooth EHL:

$$\lim_{\Lambda \to +\infty} f_\Lambda = 1$$

– when Λ is negligible, there is only boundary lubrication:

$$\lim_{\Lambda \to 0} f_\Lambda = 0$$

The family of functions of the form $\tanh\left(a_\Lambda \cdot \Lambda^{b_\Lambda}\right)$ automatically satisfies these requirements while retaining a wide descriptive range.

The viscosity parameters were obtained by fitting the parameters of the Roelands equation to the equations proposed in Table 5.2 under the heading of "Viscosity and density equations". An important point should be made at this stage: as discussed by Bair [49] and Bair et al. [50], the classic pressure-viscosity relations—and among them the Roelands equation—tend to grossly overestimate the viscosity at high pressure (in the range of > 1 GPa). This puts a bound on the precision of any calculations based on such rheological laws.

The remaining parameters were not so easy to determine, in the absence of the needed experimental data. An indirect course had to be taken to determine the parameters. In Chapter 5 of [8] is presented a power loss model for gearboxes, in which, among other things, a method for evaluating the friction coefficient is proposed. Thus, based on power loss tests, conducted on gears similar to those of the micropitting tests, and on the corresponding average friction coefficients, a series of simulations of the contact were run. The simulations, performed according to the principles described in Section 2.4.3 of this work, enabled the calculation of the average friction coefficient over a meshing cycle while varying parameters τ_{L0}, α_{τ_L}, β_{τ_L}, μ^{BDR}, a_Λ and b_Λ. For each value of μ^{BDR} in a range that spans the values 0.7–0.14 the remaining parameters that best fitted the power loss tests in the least square sense were determined.

It was found that the parameters related to the limiting shear stress vary within a narrow range and can be adequately taken as independent of μ^{BDR}, within the framework of these calculations.

Note that for every distinct μ^{BDR}, the correlation with the power loss test data was very high—over 98%. Therefore, it is impossible to decide which combination of boundary friction coefficient and load sharing function is the correct one. Thus, simulations must be repeated for each possible value of μ^{BDR}. It is not to be expected that the parameters thus determined are very exact but, in the absence of dedicated tests, this is the best that can be done at the moment. Figure 5.7 shows a few of the load sharing functions that were used and Table 5.6 the several combinations of parameters that were considered in the simulations.

5.3.3 Properties related to fatigue

Because the teeth have a carburized surface layer, and micropitting fatigue cracks originate on the surface, the Dang Van fatigue properties α_{DV} and β_{DV} should

Table 5.6: Parameters for the mixed film lubrication regime.

μ^{BDR}	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15
Load sharing function $f(\Lambda) = \tanh\left(a_\Lambda \cdot \Lambda^{b_\Lambda}\right)$									
a_Λ	1.110	1.318	1.484	1.614	1.718	1.800	1.868	1.925	1.975
b_Λ	0.277	0.270	0.258	0.243	0.228	0.213	0.199	0.187	0.177
Roelands viscosity									
T_0	363 K								
η_0	$1.56 \cdot 10^{-2}$ Pa $\cdot$ s								
S_0	1.28								
Z	0.608								
Limiting shear stress									
τ_{L0}	25 MPa								
α_{τ_L}	588^{-1} MPa^{-1}								
β_{τ_L}	0 K								

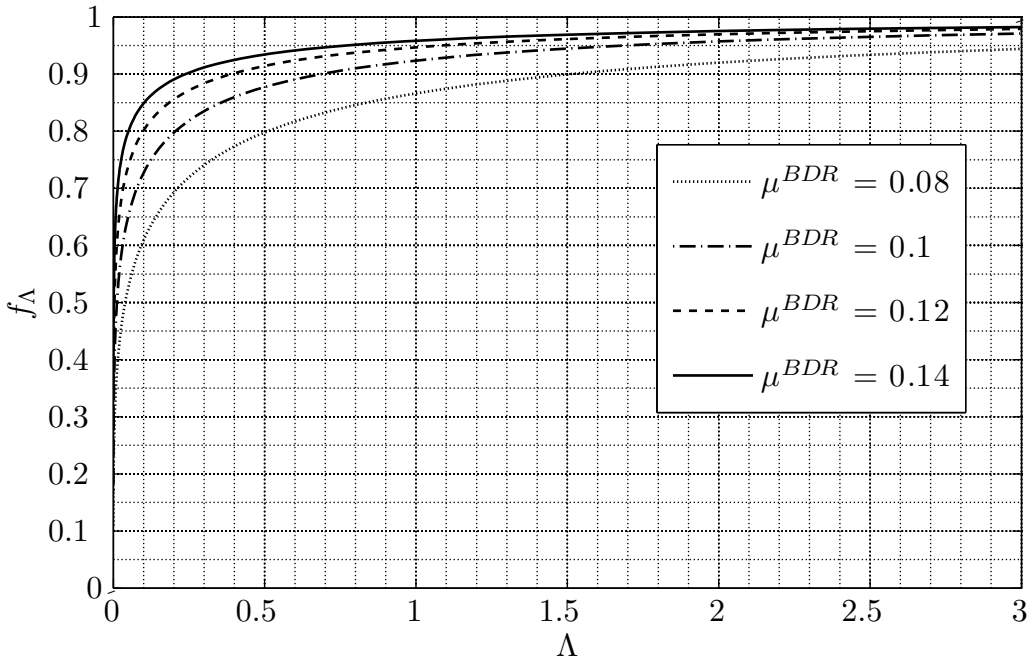


Figure 5.7: Plot of load sharing functions against the specific film thickness. The functions that combine with $\mu^{BDR} \in \{0.08, 0.1, 0.12, 0.14\}$ are shown.

ideally be obtained from fatigue tests on samples made wholly of carburized steel. This presents severe practical difficulties, both in obtaining such test samples—carburized layers are difficult to obtain at depths higher than 1 mm—and in testing such a brittle material. In practice, these data are not available on the open literature.

To bypass this difficulty, it was assumed that α_{DV} is the same in the carburized layer as in the substrate steel. On the other hand, the simulation was run for a range of β_{DV} spanning the values from 400 MPa to 900 MPa. This would allow patterns of gear behaviour to emerge from the simulation results that would go a long way to verify the model. In physical terms, this is equivalent to assuming that the relation between alternate torsion fatigue and pure bending fatigue is the same for the carburized steel layer and the substrate steel, while the actual alternate torsion fatigue stress may be altogether different. Although there is no actual physical evidence that such is the case, this seems a reasonable assumption that helps remove a barrier that, in the absence of the proper fatigue data, would prevent any simulation from being performed.

The value of α_{DV} of 0.987 used here was taken from [51]. Note that this value is far from being canonical for steel, as can be attested by the fact that, in their work [38], Desimone et al. reported a value of 0.52 for a quenched and tempered high strength steel.

5.4 Simulation results

It is not practical to show the results of all the simulations that were made since they total 26 (possible values of β_{DV}) $\times$ 9 (possible combinations of μ^{BDR} with f_Λ) $\times$ 3 (possible combination of tooth and roughness measurement position) = 702 cases, each with 3 load stages. Instead, this section will show in detail the result obtained for a simulation case that will be referred to as case 1 from now on. This simulation was performed using the central roughness measurements on the second measured tooth of the driving gear with the properties listed in Table 5.7.

In Figure 5.8, the values of $\beta_{eq} = \max\left(\tau_{\max} + \alpha_{DV} \cdot p_H\right)$ that violate the Dang Van fatigue criterion ($\beta_{eq} > \beta_{DV}$) at the end of load stage K8 are shown for the whole tooth. The figure was obtained by "stretching" the nominal profile of the tooth until it became a straight line while conserving the roughness variations. Thus, the xx coordinates measure the position on the surface of the tooth and the zz coordinates the depth within the tooth. Notable points in the meshing are marked on the abscissa as A, B, "pitch", D (see Figures 1.9 and 1.10): the point where the tooth nominally initiates contact with the opposite tooth on the driven gear is marked as A in the abscissa; the point where the tooth becomes the only one of the driving gear teeth in the meshing is marked as B etc...

It is interesting to note that the points where the fatigue criterion is violated come in patches. It is immediately apparent in the figure that the part of the tooth under the pitch line (from A to "pitch") suffers the most from contact fatigue,

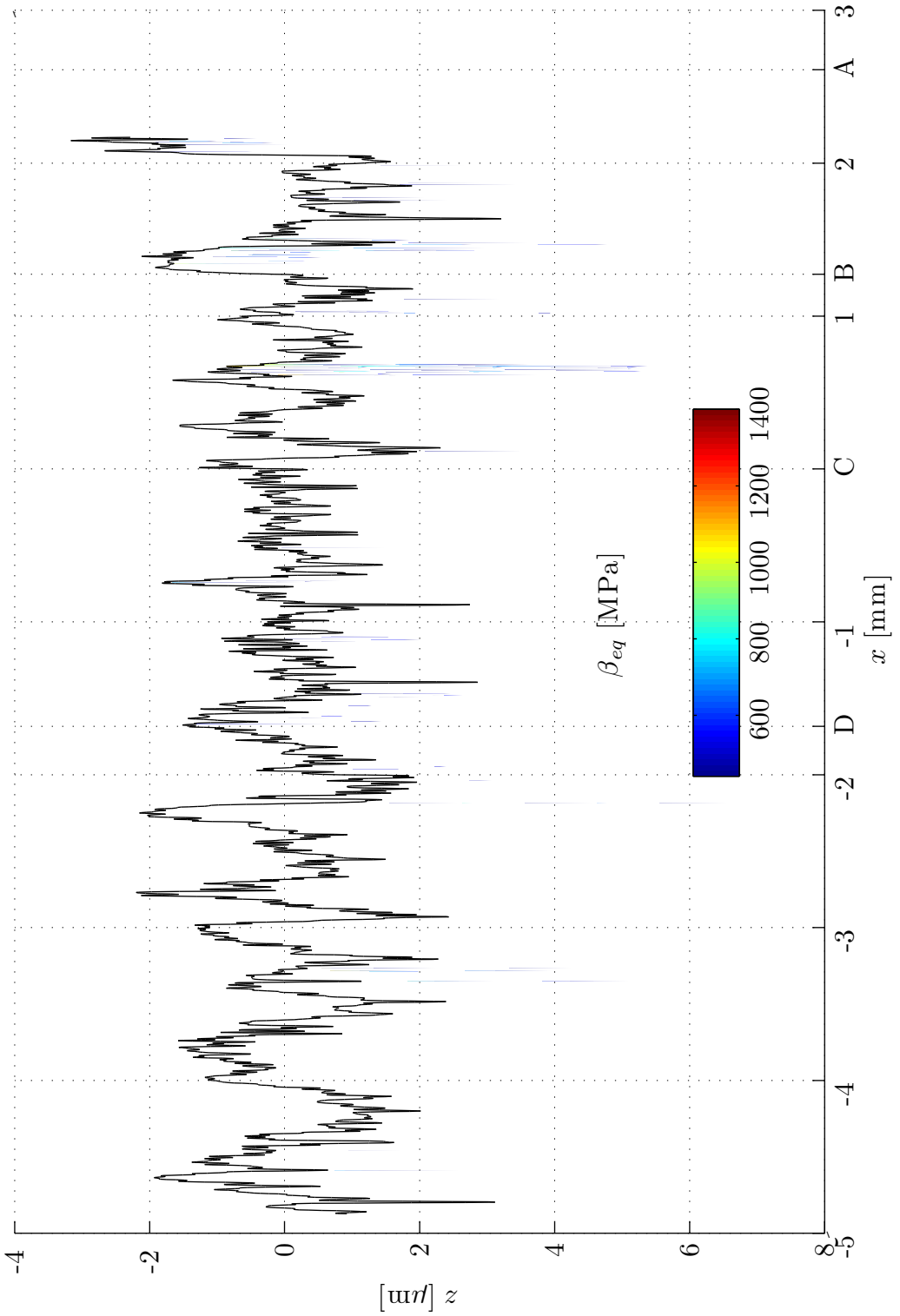


Figure 5.8: Simulation case 1, load stage K8: values of $\beta_{eq} > \beta_{DV}$ in the xz plane.

Table 5.7: Parameters of simulation case 1

Roelands viscosity			
T_0 [K]	η_0 [Pa · s]	S_0	Z
363	1.56×10^{-2}	1.28	0.608

limiting shear stress		
τ_{L0} [MPa]	α_{τ_L} [MPa^{-1}]	β_{τ_L} [K]
25	588^{-1}	0

μ^{BDR}	f_Λ	
	a_Λ	b_Λ
0.14	1.925	0.187

fatigue properties	
α_{DV}	β_{DV} [MPa]
0.987	440

according to the Dang Van criterion: it has a greater concentration of violation patches and these tend to have higher values of β_{eq}.

Figure 5.9 show the same information but only for the part of the tooth below the pitch line. It is puzzling at first to remark that not all salient roughness peaks give rise to severe contact fatigue initiation, as well as that some valleys do give rise to contact fatigue initiation. To explain this, one has to remember that the opposing driven gear tooth also has a roughness. This means that the film thickness above a driving gear tooth roughness peak is only shallow when a sufficiently salient driven gear tooth roughness feature opposes the driving gear tooth roughness peak. This is essentially a matter of chance, although the probability of fatigue initiation under a roughness peak is much higher than anywhere else, as can be seen in Figure 5.9.

Figure 5.10 shows with greater detail the patch around the point with higher β_{eq}. This patch is rather atypical in that it is very wide (around $0,1$ mm). It is composed of a number of patches that coalesced into one, as can be deduced from the fact that a number of local maxima are present on the surface.

In Figure 5.11 is shown another patch were the Dang Van criterion is violated. This is a simpler one, were a single maximum is present. This type of patch is far more frequent than the one shown in Figure 5.10. In fact, from the perusal of Figure 5.9, one can conclude that the more complex patches are an agglomeration of several simple patches such as this one caused by the proximity of several roughness peaks. Therefore, the patch shown in Figure 5.11, can legitimately be considered typical. It is interesting to note its dimensions: roughly 10 μm wide by 5 μm deep. This is very similar to the size of a micropit. The points Q and Q' are singled out for later use.

It is most instructive to see what happens in the area of the tooth around this typical patch of fatigue initiation, and the remainder of this section will be devoted

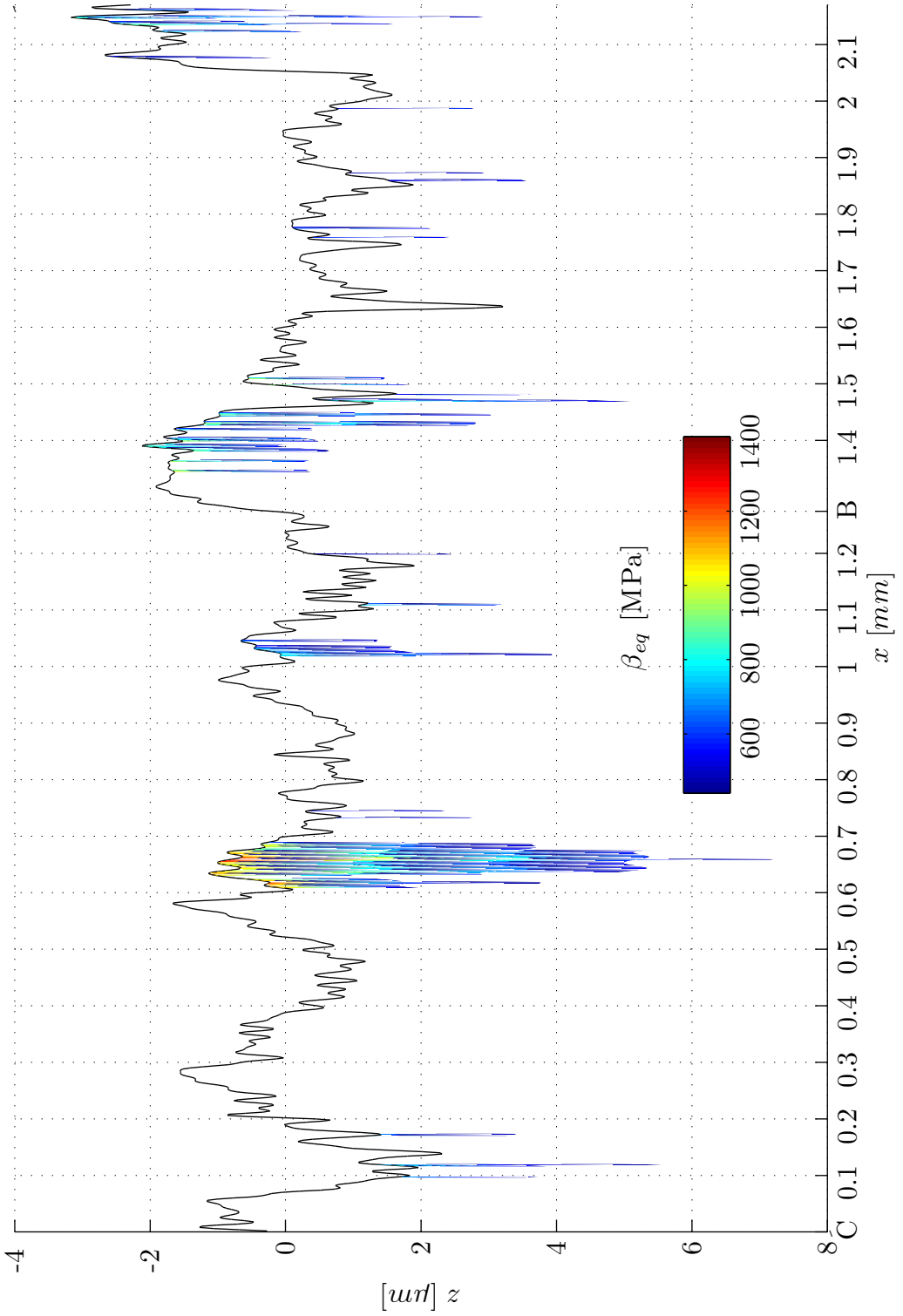


Figure 5.9: Simulation case 1, after load stage K8: $\beta_{eq} > \beta_{DV}$ in the part of the driving gear tooth under the pitch line.

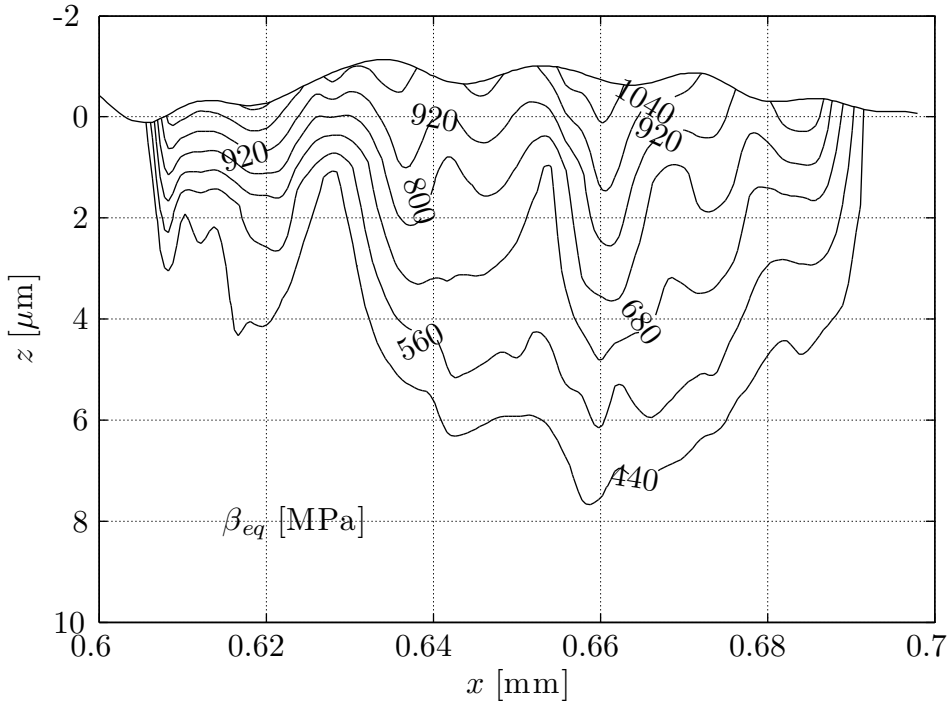


Figure 5.10: Simulation case 1, after load stage K8: contour plot of a detail of Figure 5.9.

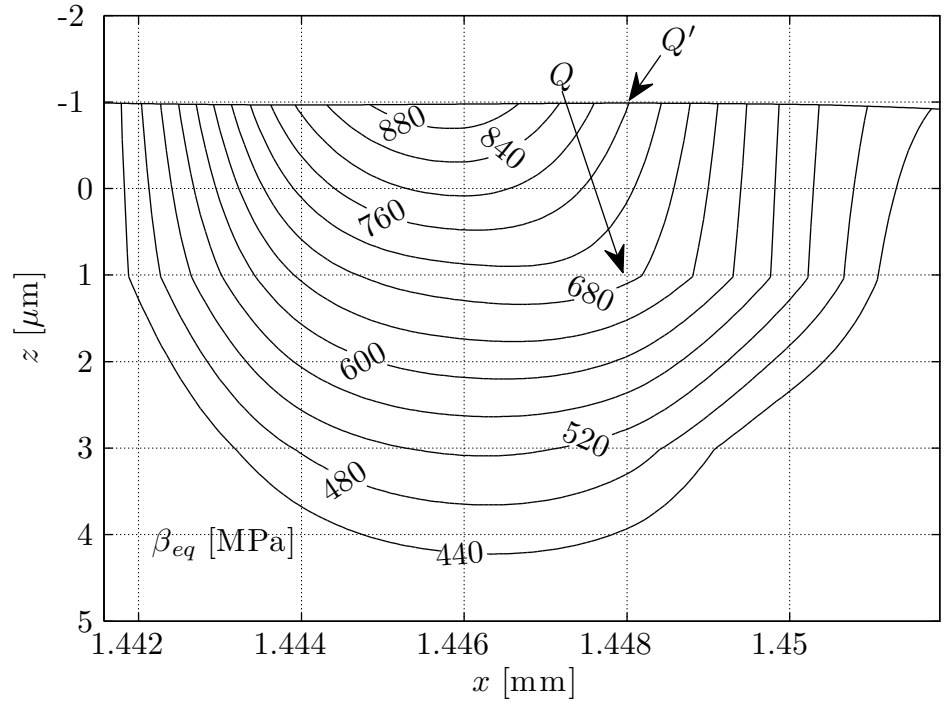


Figure 5.11: Simulation case 1, load stage K8: contour plot of another detail of Figure 5.9. Two points are singled out for later reference.

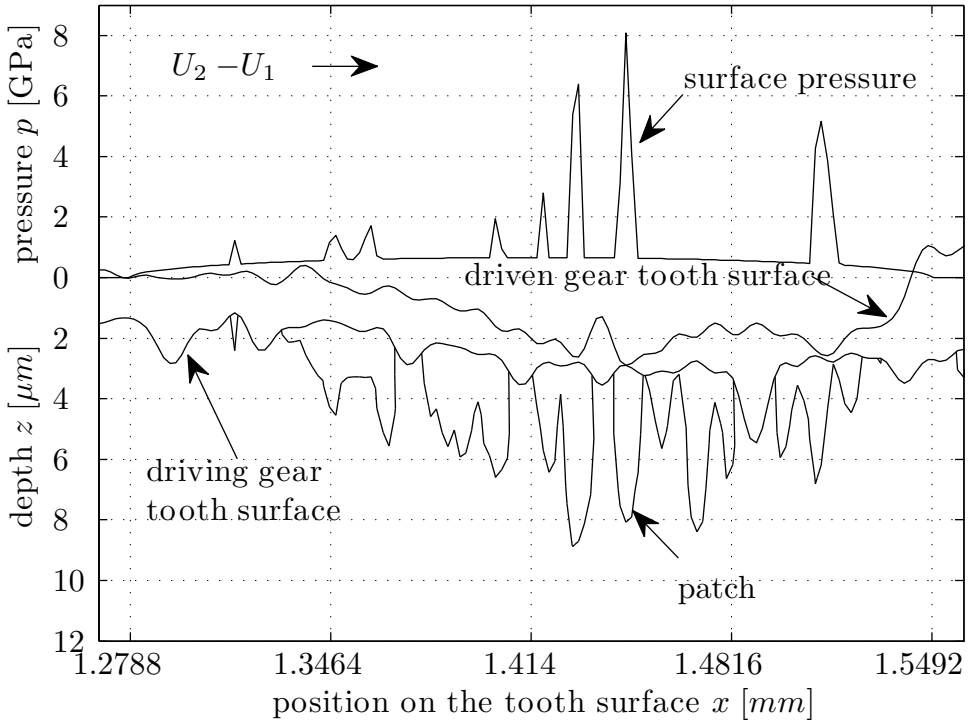


Figure 5.12: Surface pressure field when the patch of Figure 5.11 undergoes its most intense load.

to this.

The first step is to observe the pressure field when the patch of Figure 5.11 undergoes its most intense load. This is shown in Figure 5.12, where the outlines of the patches within which $\beta_{eq} > \beta_{DV}$ are represented along with the pressure distribution on the surface. Note that the tooth surfaces represented in the figure include both the nominal geometry of the teeth and the elastic deformations, unlike those in Figures 5.8–5.11. It is striking that the pressure peak over the patch is very intense (around 8 GPa) and very localized: no more than 10 μm in width. The surface shear is not represented in the figure. It is proportional to the surface pressure (the mixed film lubrication friction coefficient is around 4%) and has the same direction as the sliding velocity $U_2 - U_1$ depicted in the figure.

As an aside, the figure illustrates well the reason why some valleys in the driving gear tooth suffer from fatigue initiation: the roughness peaks of the opposite tooth can be so salient as to compensate for the lack of a jutting peak on the driving gear tooth.

It is interesting to turn ones attention to a single point and see what happens to it during the meshing cycle. The point in question is marked as Q in Figure 5.11. The point on the surface of the tooth on the vertical of Q is marked Q' on the same figure. The history of point Q is shown in Figures 5.13–5.18

In Figure 5.13 the loads on point Q' are presented through time. Note that the surface pressures and shear stresses shown are *not* a load distribution inside a

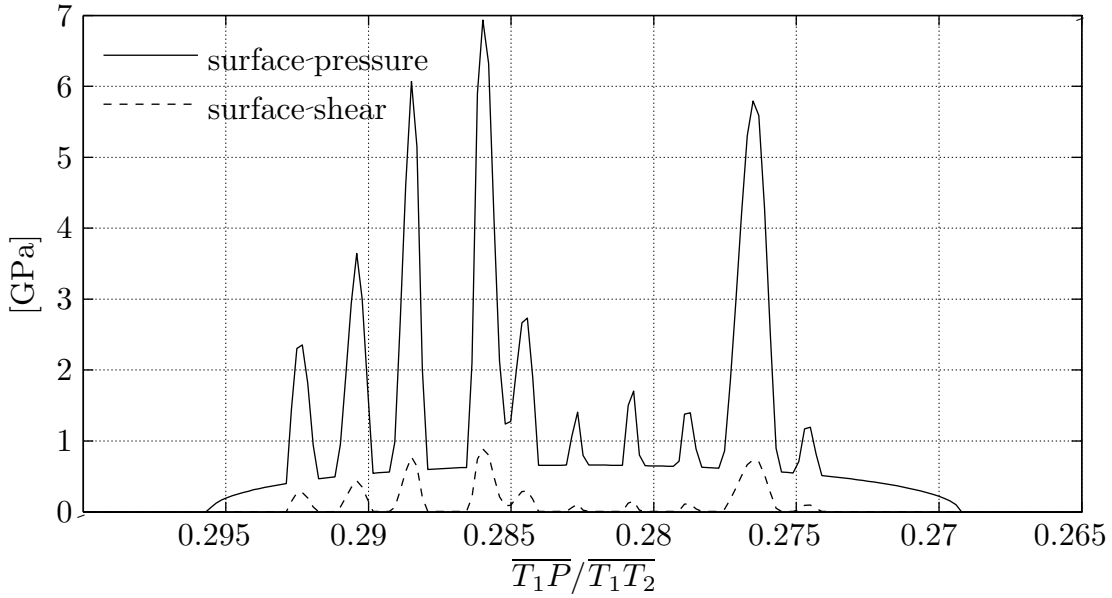


Figure 5.13: Surface loads on point Q' directly above the selected point Q inside the tooth plotted against the position of the contact in the meshing line at each instant.

contact: they are the surface pressure and shear stress applied on Q' through time. Because Q' is in the contact for only a short time during the meshing, only the relevant time interval is shown. Time is represented here as the non-dimensional position of the nominal contact point on the meshing line, and flows from the right of the figure to the left.

In Figure 5.14 the macroscopic elastic stress components at point Q are presented through time, in manner similar to that in Figure 5.13. It is interesting to compare this to Figure 5.15, where the mesoscopic stress components are plotted. This demonstrates the fact that the history of each elastic stress component is shifted as a whole vertically by the action of the initial stressses and of the mesoscopic residual stresses.

The mesoscopic principal normal stresses and maximum shear stress are shown in Figure 5.16. It is interesting to note that the maximum mesoscopic shear stress remains fairly constant throughout the meshing, with a few peaks that coincide with the pressure peaks of Figure 5.13.

In Figure 5.17, the history of the value of $\tau_{max} + \alpha_{DV} \cdot p_H$ in point Q is shown, as well as two horizontal lines corresponding to the values of β_{DV} and β_{eq}. It is, at first, surprising to notice that the most severe values of $\tau_{max} + \alpha_{DV} \cdot p_H$ do not occur when point Q is in the contact, but rather when it is outside it. In fact, $\tau_{max} + \alpha_{DV} \cdot p_H = \beta_{eq}$ for most of the meshing and only lowers when Q' enters the contact: it is actually at its lowest when the loads on Q' are at the highest. Of course, this is due entirely to the beneficial action of the hydrostatic pressure, that suffers a surge when Q' enters the contact, while the maximum shear

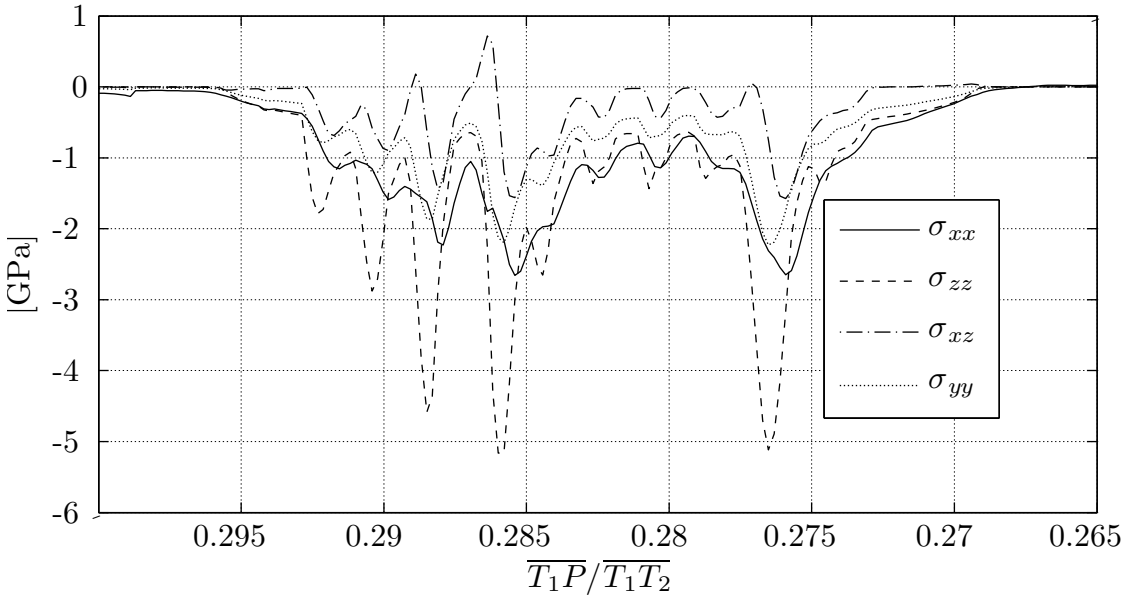


Figure 5.14: History of the macroscopic elastic stresses in point Q plotted against the position of the contact in the meshing line.

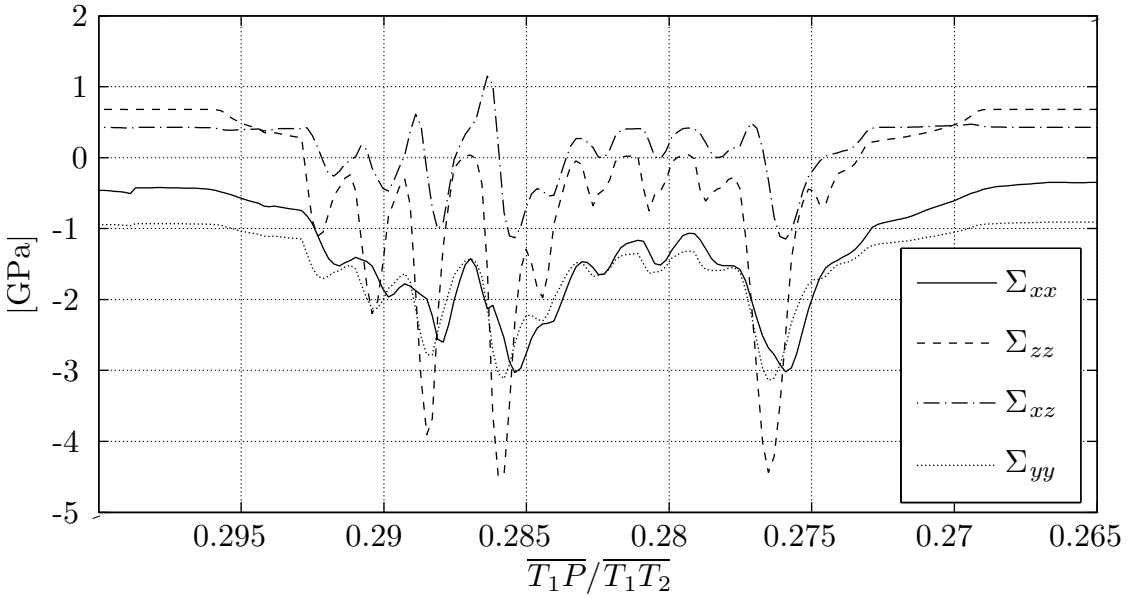


Figure 5.15: History of the mesoscopic stresses in point Q plotted against the position of the contact in the meshing line.

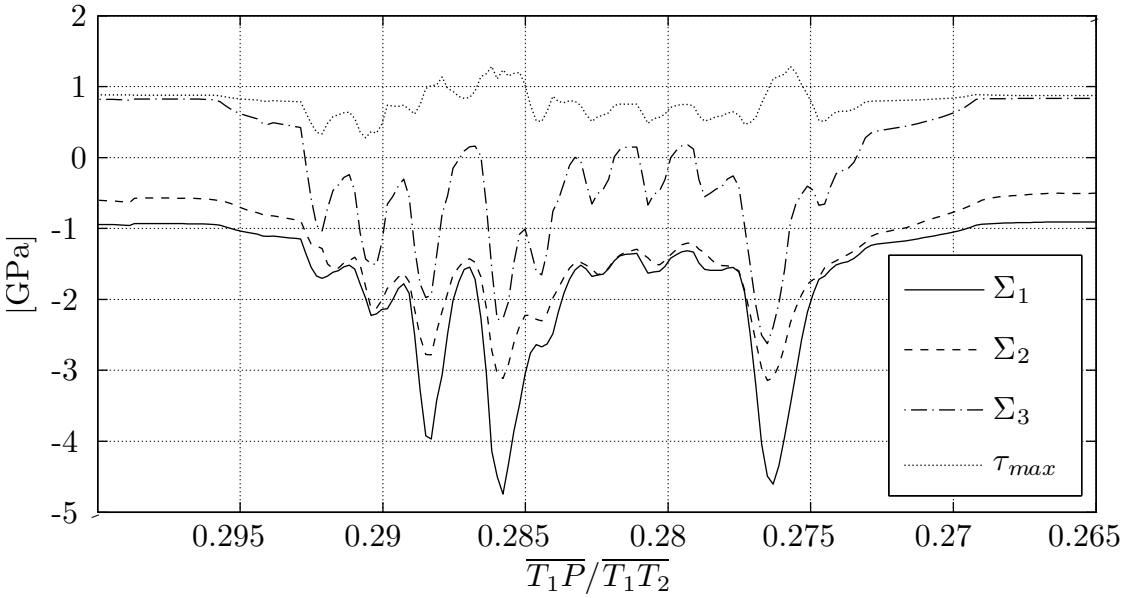


Figure 5.16: History of the mesoscopic principal normal stresses and maximum shear stress in point Q plotted against the position of the contact in the meshing line.

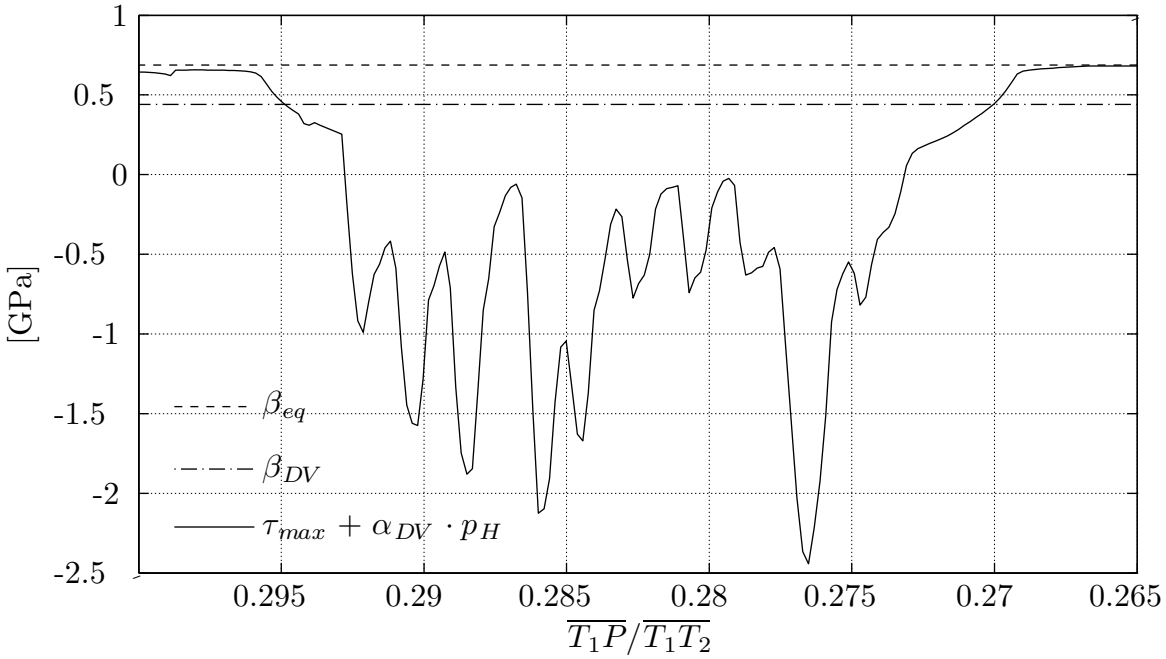


Figure 5.17: History of the value of $\tau_{max} + \alpha_{DV} \cdot p_H$ in point Q plotted against the position of the contact in the meshing line. Two horizontal lines corresponding to the values of β_{DV} and β_{eq} are also shown.

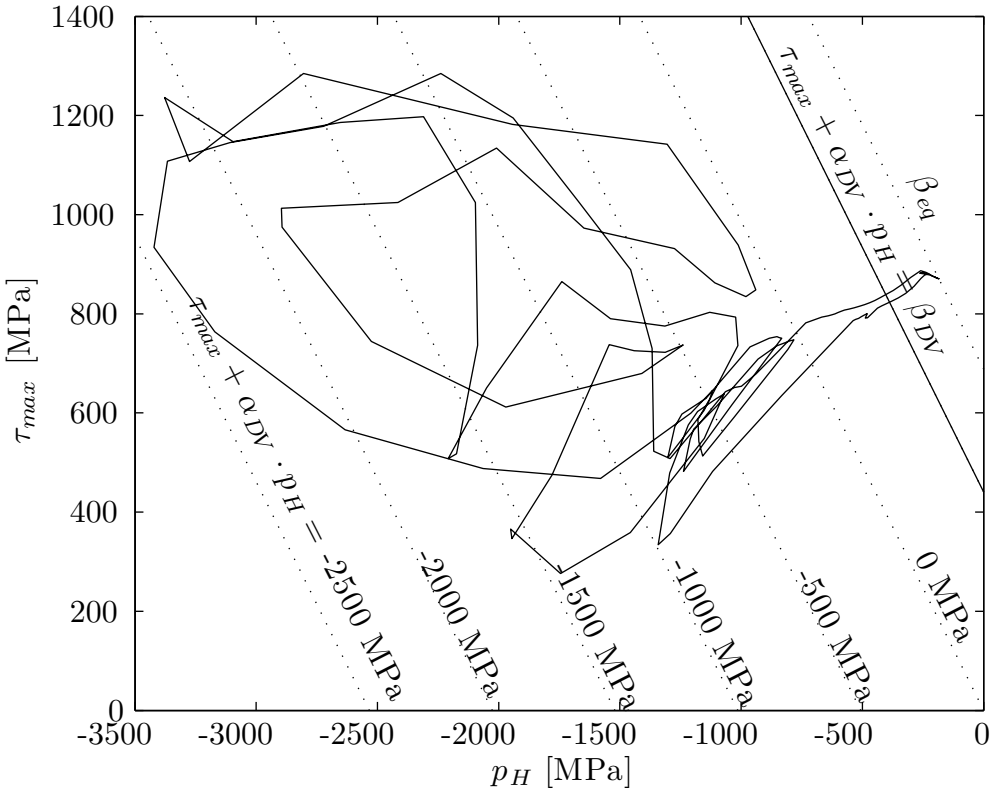


Figure 5.18: Map of the cycle undergone by point Q that plots the mesoscopic maximum shear stress against the hydrostatic stress for the whole meshing cycle.

stress fluctuates around the value at which it is kept when outside of the contact. Thus, one could evaluate β_{eq} with a high degree of certainty using only the sum of the mesoscopic residual stresses with the initial stresses, disregarding entirely the transitory effects of the elastic stresses.

This leads to a further conclusion: the only really beneficial hydrostatic pressure is that of the initial stresses (e.g. residual stresses due to heat treatment and/or grinding operation), since the mesoscopic residual stress tensor is deviatoric and therefore adds no hydrostatic pressure.

Dang Van points out in [33] that fatigue crack initiation occurs within grains that underwent excessive plastic deformation. It is easy to see that this plastic deformation is intimately related with the intensity of the mesoscopic residual stresses. Thus the figures confirm quantitatively what was known qualitatively.

All this underlines the fact that the residual stresses are the crucial consideration when dealing with fatigue initiation in gear teeth. It is interesting to mention that Batista [24] studied surface damage by contact fatigue in gears with the extensive use of the X-ray diffraction technique. His view that the residual stresses are a key element in contact fatigue is reached here through other means.

Finally, Figure 5.18 plots the path taken by the stresses on point Q in the p_H/τ_{max}

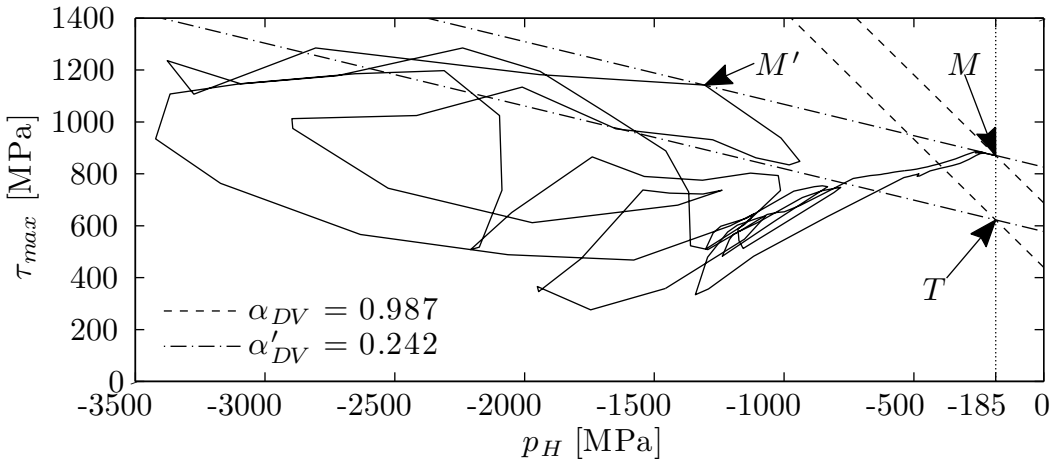


Figure 5.19: The cycle is shown as in figure5.18. The line of the Dang Van limit is rotated until it corresponds to a value of $\alpha_{\mathrm{DV}} = \alpha'_{\mathrm{DV}} = 0.242$ such that $\beta'_{\mathrm{eq}} - \beta'_{\mathrm{DV}} = \beta_{\mathrm{eq}} - \beta_{\mathrm{DV}}$. The construction lines and points marked illustrate the geometric reasoning.

plane. In the figure are also shown some straight lines that correspond to iso-values of $\tau_{max} + \alpha_{DV} \cdot p_H$, most notably the line where $\tau_{max} + \alpha_{DV} \cdot p_H = \beta_{DV}$, which divides the plane into a violation (on the right) and non violation (on the left) regions, in regard to the Dang Van criterion.

In Figure 5.19, the cycle is shown as in Figure 5.18, but this time in true proportion. The line of the Dang Van limit is the dashed line that goes through point T. The line where $\tau_{max} + \alpha_{DV} \cdot p_H = \beta_{eq}$ is the dashed line that goes through point M, the most unfavourable point of the cycle. Consequently, $\beta_{eq} - \beta_{DV} = \overline{MT}$. If the dashed line that goes through M is rotated until it touches another point on the cycle, the dash-dotted line that goes through M and M' is obtained. If another dash-dotted line is constructed parallel to the first but going through point T, a virtual Dang Van limit, to which correspond values of $\beta'_{DV} = 578$ MPa and $\alpha'_{DV} = 0.242$, is arrived at that would yield exactly the same value of $\beta_{eq} - \beta_{DV}$. Indeed, this is true of any line that goes through M and has values of $\alpha'_{DV} \in [0.242, 1]$.

There is three reasons that make this seemingly pointless argument become significant:

- The shape of the cycle path in the plane τ_{max}–p_H of a point that suffers, according to the Dang Van criterion, from contact fatigue initiation is always similar to that of Figure 5.19—particularly as regards "flatness".

- The worst point in the path is always very close to the point of no contact loads *provided that* $\alpha_{DV} \gtrapprox 0.25$. Consequently, its hydrostatic stress is nearly the same as the initial hydrostatic stress.

- The initial hydrostatic stress is nearly constant in the first 10 μm of depth—the depths at which initiation occurs, as can be seen in Figures 5.8–5.11. In

this particular case, it ranges from −191 to −244 MPa (see Table 5.5), a very small difference when considering the magnitudes that have been discussed.

As a consequence, in all the points in the first 10 μm below the surface of the tooth flank, the maxima of $\tau_{max} + \alpha_{DV} \cdot p_H$ will correspond to points gathered closely about a vertical line of abcissa $p_H \cong -200$ [MPa] in the plane τ_{max}–p_H.

From these considerations, the Dang Van criterion can be rewritten so:

$$\tau_{max} + \alpha_{DV} \cdot \left(p_H^{ini} - \overline{p}_H^{ini}\right) < \beta_{eq} - \alpha_{DV} \cdot \overline{p}_H^{ini} \tag{5.1}$$

where $\overline{p}_H^{ini}$ is an average value of the initial hydrostatic stress in the first 10 μm below the surface of the flank and p_H^{ini} the actual initial hydrostatic stress for the point in consideration. Because, generally $p_H^{ini} - \overline{p}_H^{ini}$ is small, it can be disregarded and thus the criterion becomes:

$$\tau_{max} < \beta_{eq} - \alpha_{DV} \cdot \overline{p}_H^{ini} \tag{5.2}$$

The initiation is dependent on only one material parameter!

What this long-winded exposition leads to, is the fact that it is sufficient to vary β_{DV} while keeping α_{DV} to get the complete range of possible fatigue responses. That is, of course, provided that $\alpha_{DV} \gtrapprox 0.25$.

It was mentioned in Section 3.2, page 54, that Desimone et al. have proposed in [38] a refinement of the Dang Van criterion. They attempted to delineate the locus of safe operation by conducting numerous simple tests. They concluded that the beneficial effect of hydrostatic pressure makes itself felt only up to a limit above which only the maximum shear stress influences fatigue initiation. This is equivalent to impose a horizontal line on Figure 5.18 as an additional boundary to the permitted stress states. In fact, for the region represented in Figure 5.18, the slanting straight line would not appear at all and the limit would only be represented as a horizontal line. In practice, this is equivalent to having $\alpha_{DV} = 0$. While the reasoning above only applies when $\alpha_{DV} \gtrapprox 0.25$, this is such a low value, compared to the 0.987 used here or the 0.52 mentioned by Desimone et al., that it is hard to imagine that setting α_{DV} to null should make a very great difference.

5.5 Comparison of the simulation with the gear test results

Essentially, two types of data can be—and were—obtained from Cardoso's experiments: gear mass loss data and roughness evolution data. The first kind is very global in scope in that it does not distinguish damage mechanisms: everything, from running-in to micropitting, is bundled into a few rotund numbers: mass loss at the end of a stage and counting of material particles residue in the lubricant. This is not very discriminating and needs to be complemented by the second kind of data, the roughness evolution. The identification of micropitting is effected by the

detection of pits at the surface of a tooth with depths in the order of 10 μm. As it is not practical to study the surface pit by pit, roughness parameters must be used to obtain meaningful information. A good choice would be the roughness parameters $R_{Z.DIN}$ (for overall amplitude of roughness features), R_{vk} (for reduced valley depths) and R_{pk} (for reduced peak heights). The traditional R_a and R_q roughness parameters convey very little information as to the existence of micropitting and are therefore not used.

From what has been said, the conclusion that only the results of material removal are observable is inescapable. This introduces a difficulty in the comparison of experimental data with the results of a numerical simulation based on the model presented here. Indeed, the present model is only intended as an initiation model: it does not tell anything of what might happen in subsequent stages of the history of cracks, namely their propagation and the consequent removal of material from the gear tooth surface.

A second issue that forces itself to one's attention is that the present model only describes the initiation of high cycle fatigue cracks: at no point is there an inbuilt bias toward the specific mechanism of micropitting formation, the idea being that micropitting should emerge and be recognized as such from the analysis of the results. In part, this is addressed by the fact that the analysis is restricted to a maximum depth of 20 μm. Nevertheless, this means that other surface damage mechanisms may become bundled in the results such as pitting and wear.

5.5.1 Mass loss

The simulation provides at each load stage the values of $\beta_{eq} = \max(\tau_{\max} + \alpha_{DV} \cdot p_H)$ for each material point near the surface of the tooth. As has already been seen, each of those points where $\beta_{eq} > \beta_{DV}$ should be the origin of a crack. But because these points come in patches (see Figure 5.11), the initiation of a crack at one of them should pre-empt the initiation of cracks at other points in the same patch, since the ensuing tension fields inevitably changes a great deal. Which point will actually originate a crack, what its trajectory will be and how much material will be removed are impossible questions to answer within the framework of the Dang Van Criterion. Strictly speaking, to go any further than this, a crack propagation model should be used. This is of no help in the present circumstances, since a fatigue crack propagation model is not forthcoming—indeed, it exceeds the ambitions of the present work and is a natural continuation of it.

The problem of obtaining an estimate for the mass loss of the driving gear still remains. An interesting clue is given by the experimental work by Oila and Bull [2], already mentioned in Section 3.3, page 59, from which Figure 3.9 was taken. It shows that the trajectory of a micropitting crack skirts around a nucleus of material having undergone heavy plastic deformation and originates at the surface on the downstream side from the sliding direction. Recall that, in effect, a point of fatigue crack initiation, according to the Dang Van Criterion, is a point that undergoes high plastic deformation counterbalanced by hydrostatic stress. The resemblance is

striking and is further reinforced when considering that the shapes and dimensions of the patch of crack initiation in Figure 5.11 and of the area of removed material in Figure 3.9 are very similar.

In view of this, it is reasonable to suppose that there must be a rough correspondence between the area affected by the Dang van Criterion and the mass of material removed by micropitting. It must be understood that no physical model is offered in support of this hypothesis and that, at most, a qualitative relation should be expected.

In order to compare the mass loss from the tests with the model results, a predicted mass-loss value is defined as follows:

$$m_{DV} = Z_1 \cdot A_{DV} \cdot b \cdot \rho_{\text{steel}} \tag{5.3}$$

where m_{DV} is the predicted mass loss for each stage, Z_1 the number of teeth of the driving gear, A_{DV} the predicted micropitted area in the xz plane for one tooth (obtained from the average of the three tooth profiles of roughness measurement positions 1 and 2 on tooth 2 and position 1 on tooth 1) , b the face width of a tooth and ρ_{steel} the density of the steel.

Table 5.8 shows the comparison of the measured mass loss in the driving gear at the end of each load stage with the predicted one. The predicted mass loss values are separated into mass loss below the pitch line and above it, in addition to the values for the whole teeth. Also listed are the specific values of mass loss (mass loss per surface length on a tooth). Note that, in addition to the load stages at which the test was actually performed, load stage K3 was simulated by itself and its predicted mass loss values are listed in Table 5.8 as well. Two modelled cases are shown: simulation case 1, with $\mu^{\text{BDR}} = 0.14$ and $\beta_{DV} = 440$ MPa (this is the case that was discussed at length in Section 5.4), and simulation case 2, with $\mu^{\text{BDR}} = 0.12$ and $\beta_{DV} = 560$ MPa.

As shown in the table, the test measurements are precise within 1 mg (0.5 mg on either side of the recorded value).

The first striking fact is that the predicted mass loss during load stages K3+K6 is very different from the measured one in case 1. Nevertheless, and even though there are no actual measurements against which to compare it, the value from the simulation of load stage K3 by itself is plausible, since it falls well below the value measured at the end of the consecutive run of load stages K3 and K6. This shows that the bad results of the simulation of load stage K3+K6 are probably attributable to the fact that it is incorrect to use the initial roughness with load stage K6 in the simulation: it deviates far too much from the actual conditions of the test. The same is observed in case 2.

Another interesting fact is that the specific predicted mass loss (mass loss per unit length of tooth), while being sensibly the same in stage K9 above and below the pitch line, is far higher below the pitch line than above it at stage K8. This is interesting because Cardoso reports that micropits did appear at stage K8, as can be seen in Figure 5.6, mostly below the pitch line (it is well known that micropitting

Table 5.8: Comparison of the measured mass loss at the end of each load stage with the predicted one. The predicted mass loss values are separated into mass loss below the pitch line and above it. Also listed are the specific values of mass loss (mass loss per length of tooth).

position	K3		K3+K6		K8		K9	
	abs [mg]	specific [mg/mm]	abs [mg]	specific [mg/mm]	abs [mg]	specific [mg/mm]	abs [mg]	specific [mg/mm]
model prediction, case 1, with $\mu^{\mathbf{BDR}} = 0.14$ and $\beta_{DV} = 440$ MPa								
below pitch line	0.624	0.2387	3.946	1.511	2.463	0.9429	1.52	0.5821
above pitch line	0.750	0.1381	5.571	1.026	3.309	0.609	3.403	0.6265
whole tooth	1.374	0.1707	9.517	1.183	5.771	0.7174	4.924	0.612
model prediction, case 2, with $\mu^{\mathbf{BDR}} = 0.12$ and $\beta_{DV} = 560$ MPa								
below pitch line	0.227	0.0868	2.171	0.8312	0.9359	0.3583	0.3976	0.1522
above pitch line	0.255	0.0469	3.024	0.5567	1.057	0.1946	1.137	0.2094
whole tooth	0.482	0.0599	5.195	0.6458	1.993	0.2477	1.535	0.1908
test measurements								
whole tooth	N/A	N/A	3 ± 0.5	0.373	6 ± 0.5	0.746	5 ± 0.5	0.6214

preferentially forms below the pitch line), with only a meagre increase in observed micropits at the end of load stage K9. This is the first indication that the model works as intended.

Notice as well that the predicted specific mass loss above the pitch line in load stages K8 and K9 is nearly constant in simulation case 1. One may wonder if these values might not measure the amount of wear caused by contact fatigue. Considering that the value of predicted specific mass loss above the pitch line is sensibly 0.6 in both load stages K8 and K9, and supposing an approximately equal degree of wear above and below the pitch line, the specific mass loss below the pitch line attributable to micropit generation would then be equal to $0.94 - 0.6 = 0.34$ mg/mm in load stage K8 and none in load stage K9. The same reasoning could be applied to simulation case 2. While this is a debatable conjecture, there is ground for further studies in this direction.

Finally, the most dramatic fact to emerge from the data can be better observed with the help of Figure 5.20, where the values of mass loss for the whole gear are also shown. The fact that the prediction of case 1 are spot-on in load stages K8 and K9 nearly jumps from the page. This should not be taken too literally as meaning that the proper values of the boundary friction coefficient, distribution function and Dang Van fatigue properties are those of case 1, too many uncertainties regarding the input data remain for one to be able to make such an assertion. In this regard, it is instructive to observe that the representation of the mass loss in simulation case 2, and indeed all other simulation cases not shown here, is nearly parallel to that in case 1, and therefore to the real mass loss in load stages K8 and K9. This advocates strongly for a very direct correlation of the Dang Van fatigue initiation area with the mass loss due to micropitting and mild wear.

As side note, it was found that in all simulation cases the predicted mass loss curves were always nearly parallel to those shown in Figure 5.20 and that their position was strongly affected by the value of β_{DV} and very little by that of μ^{BDR}. One should not read too much into this, because the insensitivity of the results to μ^{BDR} can simply be a numerical artefact that comes from the particular way in which each μ^{BDR} is linked with a different load sharing function—determined as described in Section 5.3.2.

In conclusion, the mass loss comparison, in spite of all the uncertainties regarding the input data and the chosen comparison method, seem to indicate that the model is adequate for the modelling of micropitting. This is a qualitative evaluation of the model. For a quantitative one to be made, a battery of experiments designed for the exclusive study of micropitting must be performed, in order to remove all the uncertainties that plagued these simulations.

5.5.2 Roughness evolution

We come now to the section of this chapter in which the results of the simulation are compared with the tests with regard to the evolution of the teeth flanks roughness.

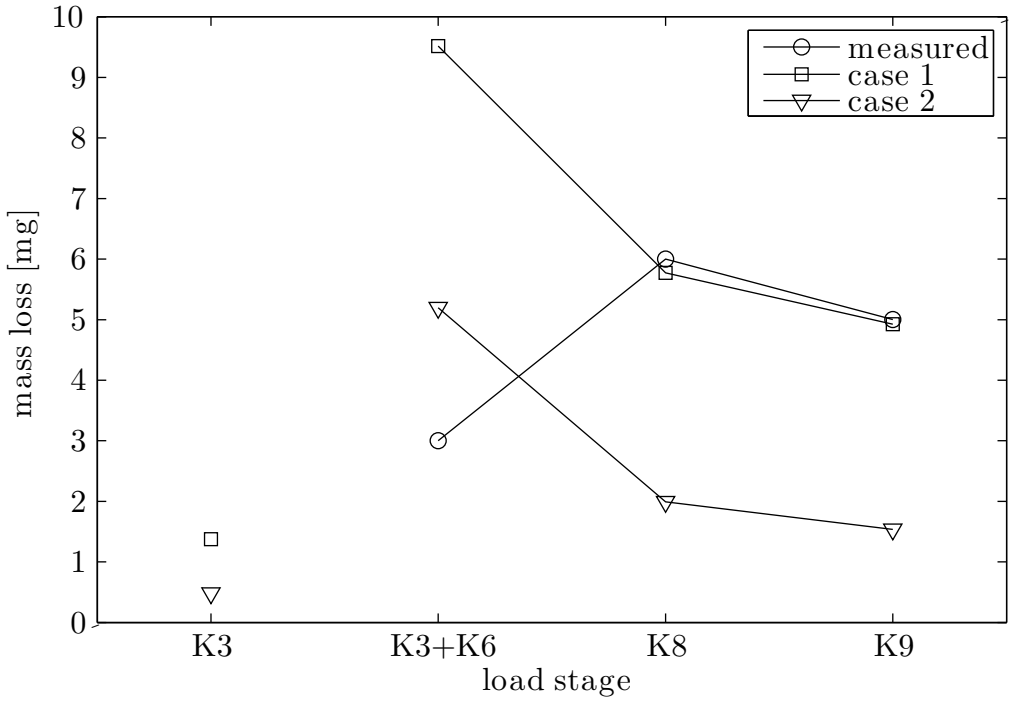


Figure 5.20: Mass loss of the driving gear in each stage: measured values and predicted ones for simulation case 1 ($\mu^{\mathrm{BDR}} = 0.14$ and $\beta_{DV} = 440$ MPa) and case 2 ($\mu^{\mathrm{BDR}} = 0.12$ and $\beta_{DV} = 560$ MPa).

In the previous section a prediction of the mass loss was evaluated by subtracting the patches of fatigue initiation, such as that shown in Figure 5.11, from the surfaces. In the same way, an attempt is made here to predict, for each load stage, the final roughness of the surface of the driving gear.

The results are shown for the same simulation cases used in Section 5.5.1. Each of these simulations gives results for three tooth profiles, called *tooth 1*, *tooth 2 position 1* and *tooth 2 position 2*, according to the position of the roughness measurement by Cardoso that gave rise to it: tooth 1 is the measurement above and below the pitch line on the first tooth in the central position shown in Figure 5.3; tooth 2 position 1 is the measurement above and below the pitch line on the second tooth in the leftmost position shown in Figure 5.3; tooth 2 position 2 is the measurement above and below the pitch line on the second tooth in the central position shown in Figure 5.3.

Evolution of the roughness below the pitch line in simulation case 1

In Figures 5.21–5.23, the evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for, respectively, tooth 1, tooth 2 position 1 and tooth 2 position 2, is shown.

The meaning of Figure 5.21 is explained as follows. The figure deals with the evolution of the roughness parameters for tooth 1 below the pitch line. In the top row, the evolution of the roughness parameter $R_{Z.DIN}$ is shown in two ways. On the left column, the values of the predicted $R_{Z.DIN}$ are plotted against the load stages as a full line with round markers, the values of the measured $R_{Z.DIN}$ are plotted against the load stages as a full line with square markers. On the right column, the values of the predicted $R_{Z.DIN}$ are plotted against the values of the measured $R_{Z.DIN}$. The same principle applies to the middle and bottom rows, which deal with R_{pk} and R_{vk}.

Figures 5.22 and 5.23 follow the same nomenclature applied to, respectively, tooth 2 position 1, and tooth 2 position 2.

Acceptable values of roughness parameters should fall within the range of measured roughness shown as boxes in the leftmost columns of the figures. Furthermore, the graphs of the rightmost columns should fall more or less on the diagonal that connects the bottom left corner to the upper-right corner of each picture.

Such is not the case.

Evolution of the roughness below the pitch line in simulation case 2

The evolution of the roughness parameters is represented in Figures 5.24–5.26 for the simulation case 2 in exactly the same manner as for simulation case 1. It is interesting to note that this simulation case, that performed poorly when comparing the predicted mass loss with the measured one, shows almost acceptable values of roughness parameters, if one discounts the results of load stage K3+K6, which are here, as in the mass-loss comparison, thoroughly off the mark. This shows

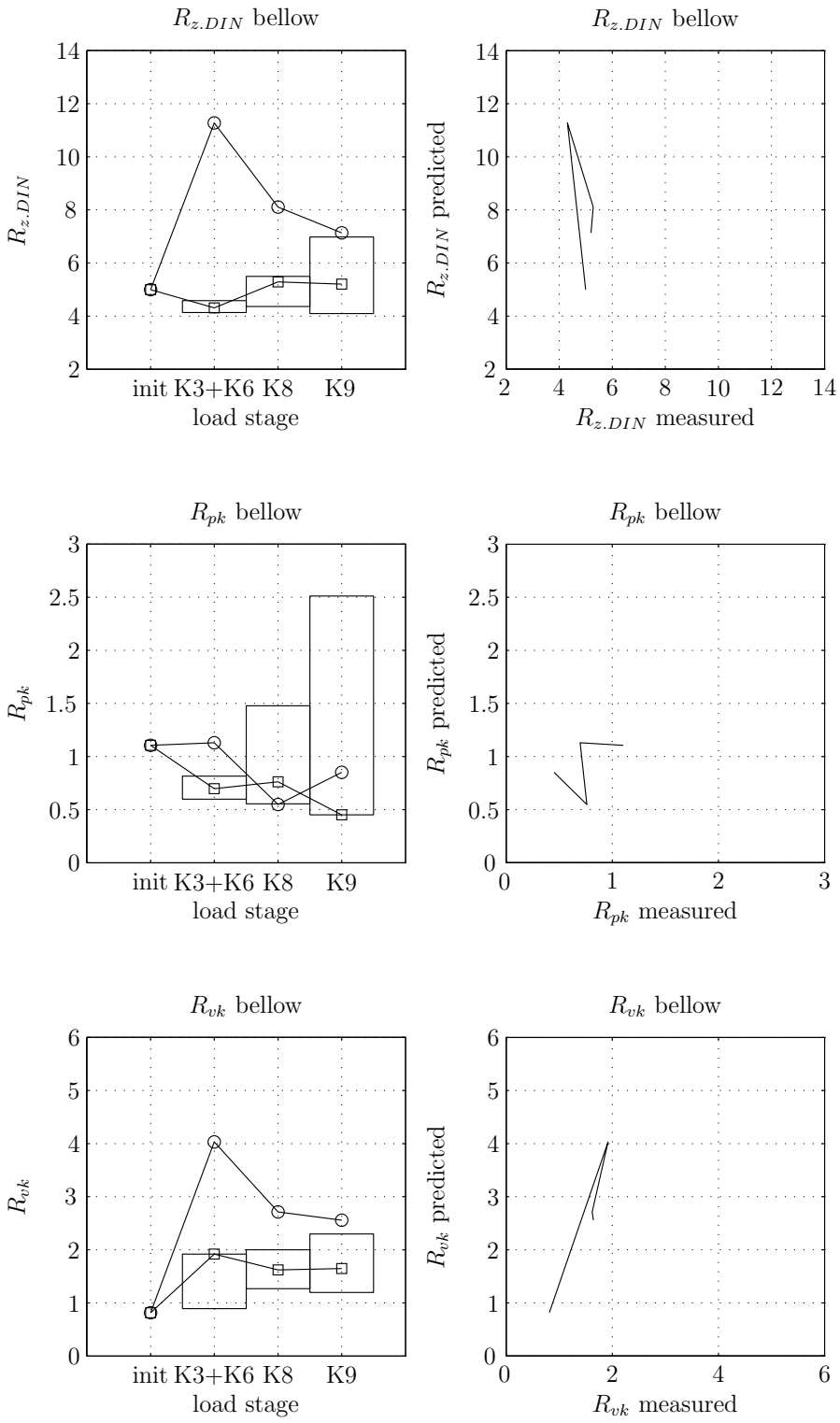


Figure 5.21: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for simulation case 1, tooth 1 and comparison with Cardoso's measurements. The rectangles mark the extrema of the measurements.

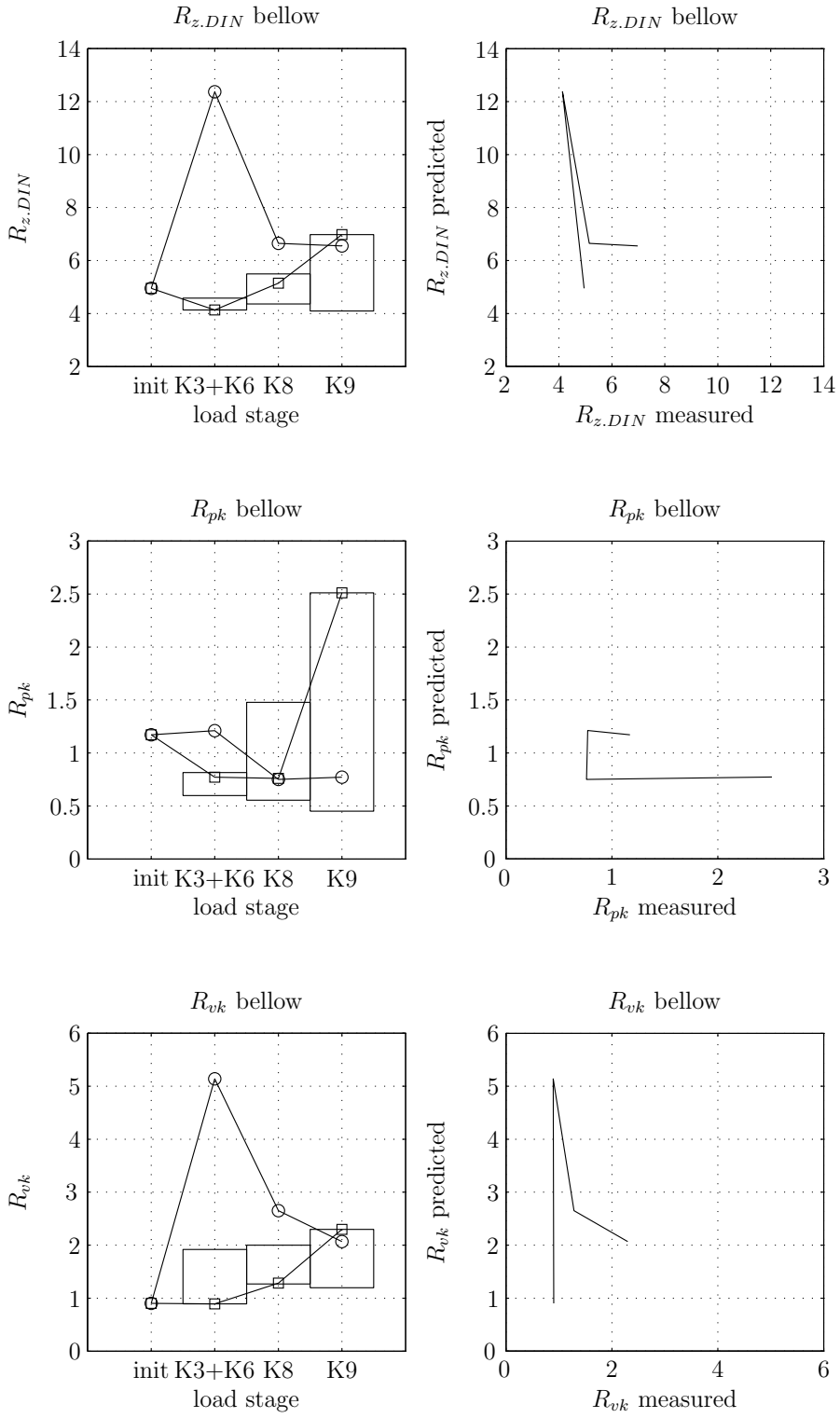


Figure 5.22: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for the simulation case 1 on tooth 2, position 1, and comparison with the measurements by Cardoso.

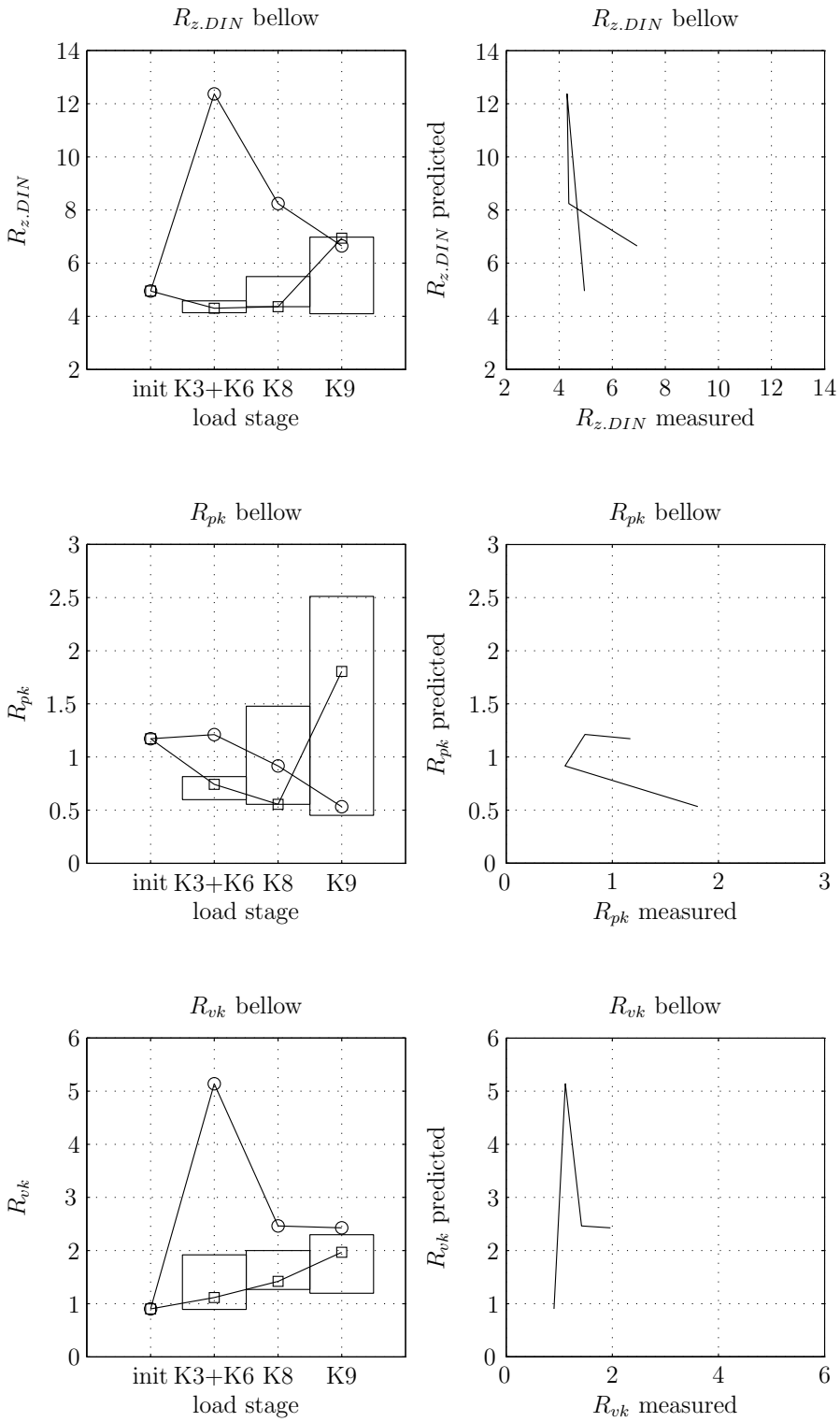


Figure 5.23: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for the simulation case 1 on tooth 2, position 2, and comparison with the measurements by Cardoso.

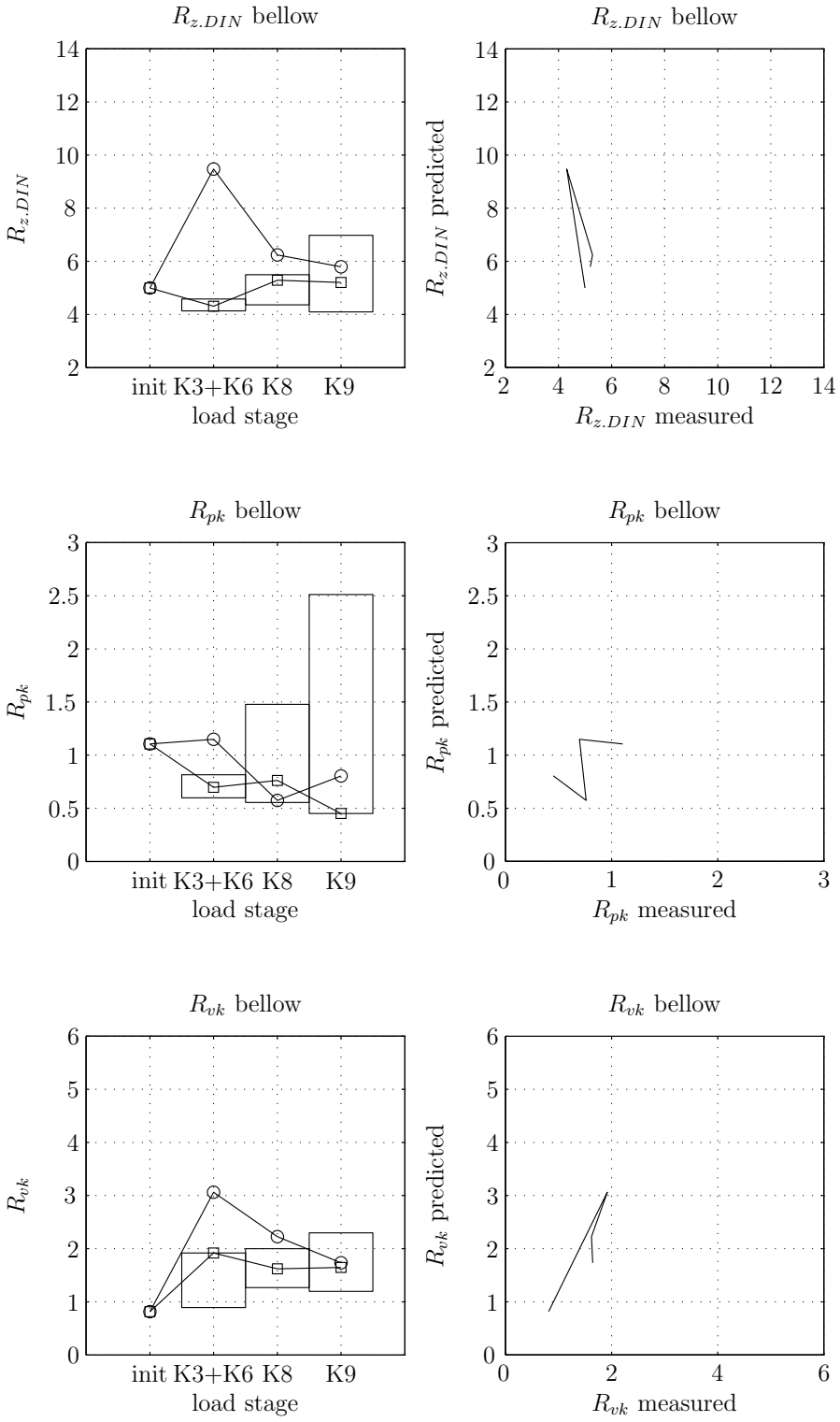


Figure 5.24: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for the simulation case 2 on the first tooth and comparison with the measurements by Cardoso.

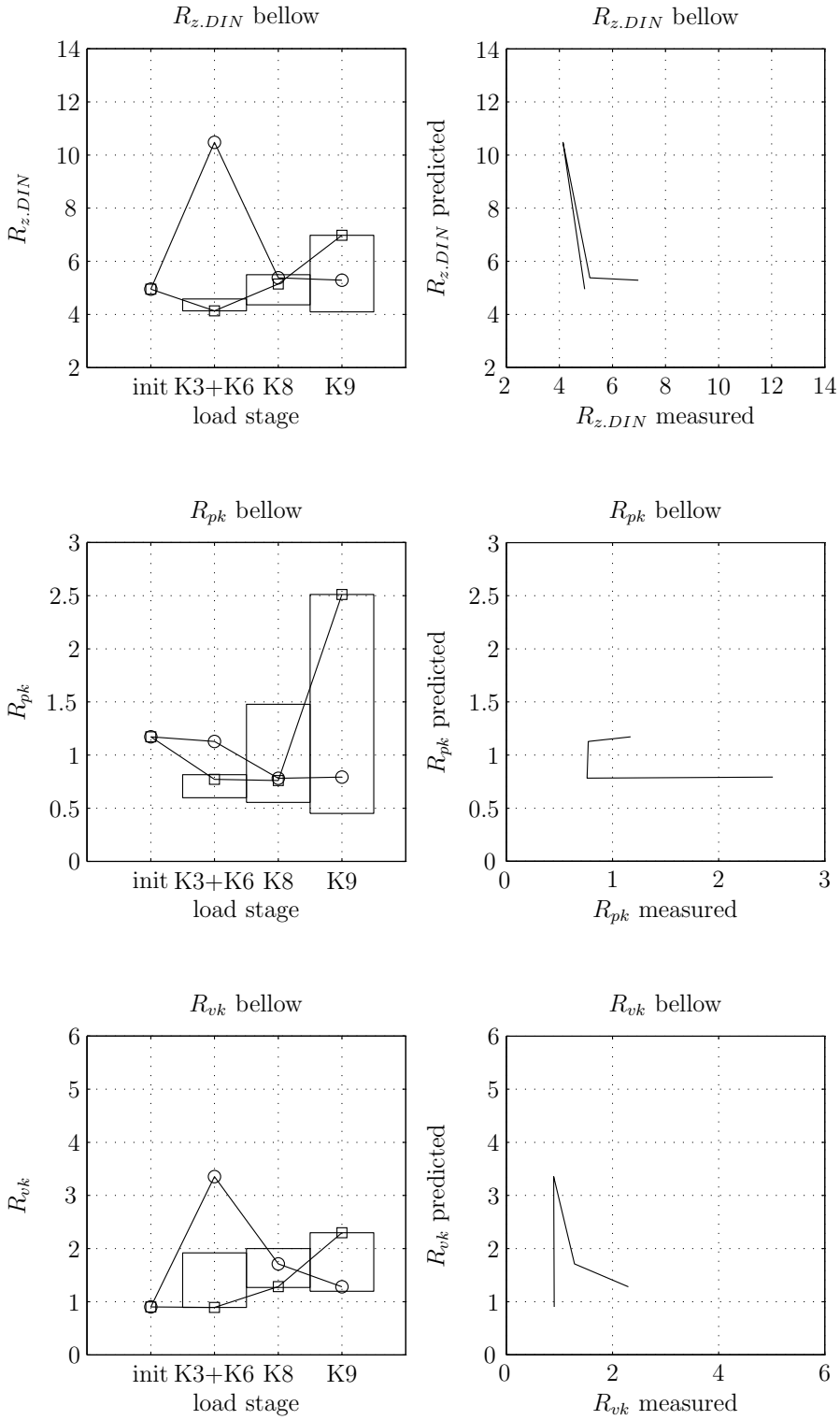


Figure 5.25: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for the simulation case 2 on tooth 2, position 1, and comparison with the measurements by Cardoso.

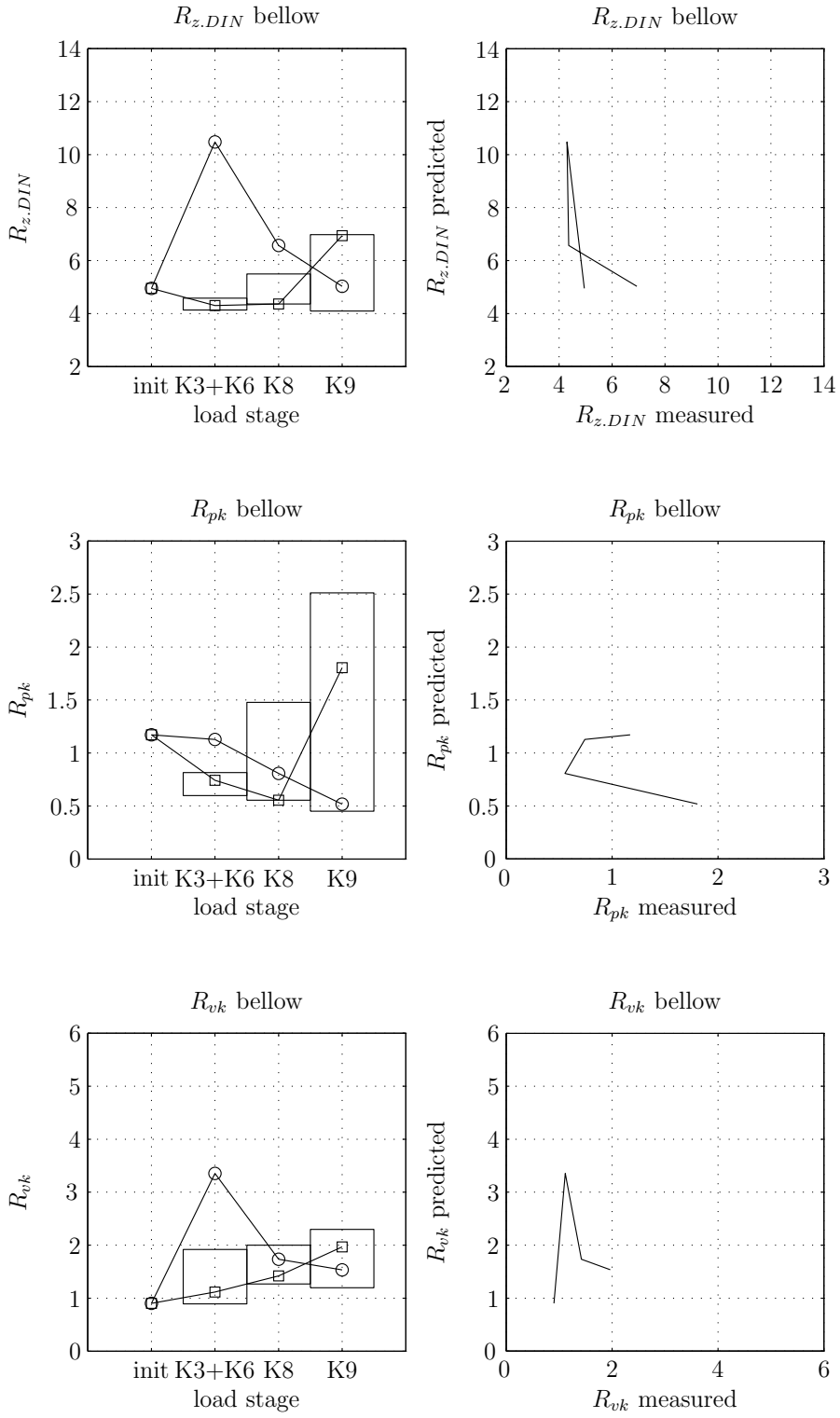


Figure 5.26: Evolution of the roughness parameters $R_{Z.DIN}$, R_{pk} and R_{vk} below the pitch line for the simulation case 2 on tooth 2, position 2, and comparison with the measurements by Cardoso.

once more that this load stage must be excluded from the discussion of the results, since it is manifestly inadequate to simulate load stage K6 with the initial teeth roughness.

With this proviso, the roughness results of simulation case 2 are acceptable.

5.5.3 Discussion of the comparison results

An approximate method for the evaluation of the removal of material from the surfaces was proposed, whereby the patches of fatigue initiation would be removed at the end of each stage. Based on this, a prediction of the mass loss by micropits and of the roughness evolution was compared with the measurements. A comparison of the results of simulation case 1 and simulation case 2 were presented.

It was shown that simulation case 1 agreed perfectly with the measured mass loss, but very poorly with the measured roughness evolution. On the other hand, simulation case 2 does not agree well with the mass loss results, but agrees reasonably well with the roughness measurements.

This shows that the approximation used, where the fatigue initiation patches would be removed, is not an accurate description of the way in which the pits form. Nevertheless, the evolution of the predicted mass loss from load stages K8 to K9 is in all cases parallel to the measured one. This shows that a very strong correlation exists between the total surface material removed by micropitting and mild wear and the total area of fatigue initiation patches.

In view of the results of case 1 one might legitimately build a case in favour of the notion that the pits occur at the sites of the patches of initiation with a similar volume—which would account for the agreement of the mass loss prediction with the measurements—but with a different shape, from which fact the differences in the predicted and measured roughness parameters would ensue. This is highly speculative but worth looking into.

The results of simulation case 2 are a very strong argument in favour of the validity of the simulation method presented here. The predicted mass loss evolution is parallel to the measured one and the roughness evolution is fairly close to the real one. This would not be possible if the model were very far from reality.

6 Conclusion

6.1 Review of the more notable issues in each chapter

In Chapter 1, a general formulation of the geometric, kinematic and Hertzian mechanic components of gear meshing were introduced for use in the model.

In Chapter 2, a modified elastohydrodynamic lubrication model in the mixed film regime was presented that permits a fast solution of the contact problem of the gears. The model is both approximate and exact in the sense that some rigour in the mathematical formulation of the problem—it avoids entirely the direct solution of the Reynolds equations—is sacrificed in return for including a more complex set of physical conditions—non-Newtonian rheology, thermal variations, very low specific film-thickness. This permits the solution of realistic gear meshing contact problems in an acceptable time.

In Chapter 3, the stresses in a tooth were discussed both in the macroscopic and in the mesoscopic sense. The Dang Van criterion was discussed, in particular as to its relation with elastic shakedown as a limit to high-cycle—or near infinite life—fatigue. From there, the discussion was shifted to surface damages and their relation with fatigue initiation. In particular, it was shown that the Dang Van criterion is a good candidate for the prediction of gear micropitting fatigue crack initiation

In Chapter 4, fast and unconditionally convergent algorithms, obtained from the open literature, were presented for the solution of the rough boundary lubrication contact problem, of the elastic stress calculation problem and of the problem of determining the mesoscopic stresses from the macroscopic ones. The last in particular is a marked improvement over the practice that is publicized in the literature on the application of the Dang Van problem.

In Chapter 5, a comparison was made between an actual micropitting test and its simulation using the present model. This simulation underlines the main difficulty in using the model: it requires a wealth of detailed information about the materials, such as the knowledge of the precise non-Newtonian rheological behaviour of the lubricating oil, of the boundary friction coefficient, of the load sharing function for the mixed film lubrication regime, of the elastic and fatigue properties of the outer layer of a tooth—in this case a carburized layer, of the initial stresses in the tooth and of the roughness profile for the whole length of several teeth, as many as must be treated to form a complete cycle—in the present case, the roughness profiles for two driving gear teeth and three driven gear teeth would be required. These data

are extremely difficult to obtain in the literature, therefore, the needed information must be obtained by running experiments before the simulation.

Because it was not possible to do so in the present case, a number of reasonable assumptions had to be made, coupled with the repetition of the simulation varying the material properties, in an attempt to force a pattern to emerge that would allow the validation of the model.

Another difficulty, but in the evaluation of the results, not in the use of the model, is that the initiation predicted by the model is an inherently microscopic phenomenon, and is therefore not in a direct relation with the observed phenomena of a micropitting test, in which the micropits are only observable when the propagation of the fatigue cracks has fully developed. To permit a comparison of the model results with those observed in the test, it was assumed that the fatigue criterion violation patches correspond approximately to the micropits mass loss. This was deemed reasonable, even though no canonical physical evidence can be advanced in the defence of this hypothesis.

6.2 Concluding remarks

For technical reasons already discussed, it was not possible to isolate the running-in loading stage from the subsequent one in the simulations. As a result, the predictions of the model were uniformly bad as regards this bundled load stage (K3+K6). Several explanations can be advanced. The first and more obvious one is that the combination of a relatively high load with the high initial roughness present before running-in is an unrealistic representation of the physical reality of the two load stages. A second possible explanation is that running-in cannot be adequately modelled by the fatigue model here presented because of the rapid rate of change of the surface topography and because running-in is most probably caused by a host of phenomena, beside fatigue. Nevertheless, the mass loss results obtained from the simulation of the K3 load stage by itself do not seem to be very wrong, at least qualitatively. Thus this second explanation may or may not be valid.

More encouraging are the results of the load stages K8 and K9, whose prediction showed a remarkable correlation with the measured mass loss of the driving gear. The roughness evolution results were not so good, but they are more prone to local error and this is compounded by the fact that the roughness measurements could never be guaranteed to be made exactly on the same spot at the end of each load stage. Therefore, the roughness evolution results are not altogether unsatisfactory, since it was shown that in certain cases of material properties it was possible to maintain a high correlation of the predicted mass loss while maintaining the roughness parameters within acceptable bounds. When one reflects that this was achieved without a proper propagation model, it becomes at once clear that these are satisfactory results.

These considerations dealt with the comparison of the model with the test results

and showed that, while not entirely conclusive, the comparison of the simulation with the test goes a long way toward validating the model.

As to the events inside a tooth, the main conclusion to be drawn is that within the framework of the Dang Van fatigue criterion, the most unfavourable mesoscopic load state occurs outside of the contact. This shows that the high hidrostatic compressive stresses inside the loading have no lasting beneficial effect. On the contrary, they derive from a stress state that violently distorts the gear so that the maximum shear strain is increased at the mesoscopic level, with the consequent harmful effect to the fatigue behaviour of the gear. In fact, the only really beneficial hydrostatic compressive stress is that provided by the initial residual stresses. The fact that the most unfavourable mesoscopic stress occurs outside of the contact also demonstrates the enormous importance of the residual stresses, both the initial ones and those resulting from the load cycle, in determining the fatigue damage in the tooth.

As a concluding remark, the author of the present work wishes to reiterate that the results of the model are very encouraging, even if not entirely conclusive.

6.3 Proposals for further work

The most urgent task in furthering the development of the present model would be to conduct a series of tests designed exclusively for the rigorous validation of this model with all the input data exactly determined beforehand. This would involve the realization of a number of experiments prior to the tests proper, in order to determine all the needed parameters already mentioned. Having done this, it would be possible to determine the areas where the model is deficient and thus refine it.

It was seen that micropitting and wear were difficult to isolate. Indeed, the model seems to deal with mild wear as well as micropitting. While this work was mostly concerned with micropitting, it would be interesting to ascertain to which degree wear—and running-in—are modelled.

Less urgent is the combination of the initiation model with a propagation model that would allow us to answer such questions as: which point in the violation patch really does originate a crack? why do the fatigue cracks have always a very definite orientation on the teeth (inwards toward the pitch line in the driving gear teeth, inwards away from the pitch line in the driven gear tooth)?. These are impossible to answer using merely the initiation model presented here.

Of course, these questions are more interesting from a scientific standpoint than from an industrial one. From an industrial standpoint, an initiation model properly validated and refined could lead to a security criterion based, for instance, on the total predicted area of fatigue violation patches. This in itself would be a very desirable outcome for this work.

Bibliography

[1] J. Castro, J. Seabra, Scuffing and lubricant film breakdown in FZG gears part I. Analytical and experimental approach, Wear 215 (1-2) (1998) 104–113.

[2] A. Oila, S. J. Bull, Phase transformations associated with micropitting in rolling/sliding contacts, Journal of Materials Science 40 (18).

[3] G. Peridas, A. M. Korsunsky, D. A. Hills, The relationship between the Dang Van criterion and the traditional bulk fatigue criteria, The Journal of Strain Analysis for Engineering Design 38 (3) (2003) 201–206.

[4] F. Antoine, J.-M. Besson, Simplified modellization of gear micropitting, Journal of Aerospace Engineering 216 (6) (2002) 291–302.

[5] A. Oila, S. J. Bull, Assessment of the factors influencing micropitting in rolling/sliding contacts, Wear 258.

[6] T. Krantz, The influence of roughness on gear surface fatigue, NASA TM-2005-213958, ARL-TR-3134,.

[7] J. Tao, T. G. Hughes, H. P. Evans, R. W. Snidle, N. A. Hopkinson, M. Talks, J. M. Starbuck, Elastohydrodynamic lubrication analysis of gear tooth surfaces from micropitting tests, Journal of Tribology 125 (2) (2003) 267–274.
URL `http://link.aip.org/link/?JTQ/125/267/1`

[8] N. Cardoso, Mechanical performance of biodegradable, low toxicity ester based industrial oils, Master's thesis, Faculdade de Engenharia da Universidade do Porto (2007).

[9] L. Faure, Aspect des dentures d'engrenage après fonctionnement, Centre Technique des Industries Mécaniques, 1990.

[10] L. Mummery, Surface Texture Analysis: The Handbook, Hommelwerke GmbH, 1990.

[11] MAAG Gear Company, MAAG gear book: calculation and practice of gears, gear drives, toothed couplings and synchronous clutch couplings, MAAG Gear Company Ltd., 1990.

[12] J. O. Seabra, Mecânica do contacto Hertziano, Faculdade de Engenharia da Universidade do Porto, 2003.

[13] S. Timoshenko, J. Goodier, Theory of elasticity, McGraw-Hill, 1970.

[14] J. O. Seabra, A. Campos, A. Sottomayor, Lubrificação elastohidrodinâmica, Faculdade de Engenharia da Universidade do Porto, 2002.

[15] L. Houpert, Contribution a l' étude du frottement dans un contact elastohydrodynamique, Ph.D. thesis, I.N.S.A. Lyon (1980).

[16] B. J. Hamrock, Fundamentals of fluid film lubrication, McGraw-Hill, 1994.

[17] A. Campos, Modelização de um contacto elastohidrodinâmico linear considerando um comportamento não Newtoniano do lubrificante e a dissipação de energia térmica no contacto, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto (2004).

[18] A. Sottomayor, Reologia de um lubrificante não Newtoniano no interior de um contacto termoelastohidrodinâmico. Determinação dos parâmetros reológicos de um lubrificante, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto (2002).

[19] D. Dowson, G. R. Higginson, Elastohydrodinamic lubrication, S. I. Editon, Pergamon Press Ltd., 1977.

[20] R. C. Castle, C. H. Bovington, The behaviour of friction modifiers under boundary and mixed EHD conditions, Lubrication Science 15 (3) (2003) 253–263.

[21] P. Vergne, Super low traction under EHD and mixed lubrication regimes, in: A. Erdemir, J.-M. Martin (Eds.), Superlubricity, Elsevier BV, 2007, pp. 429–445.

[22] J. Castro, Gripagem de engrenagens FZG lubrificadas com óleos base. Novos critérios de gripagem globais e locais, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto (2004).

[23] X. Ai, H. Cheng, A fast model for pressure profile in rough EHL line contacts, Transactions of the ASME Journal of Tribology 115 (3) (1993) 460–465.

[24] A. A. C. Batista, Deterioração por fadiga de contacto de engrenagens tratadas superficialmente, Ph.D. thesis, Universidade de Coimbra (1997).

[25] V. Bakolas, A. Mihailidis, Analysis of rough line contacts operating under mixed elastohydrodynamic lubrication conditions, Lubrication Science 16 (2) (2006) 153–168.

[26] A. Constantinescu, K. Dang Van, M. H. Maitournam, A unified approach for high and low cycle fatigue based on shakedown concepts, Fatigue & Fracture of Engineering Materials & Structures 26 (6) (2003) 561–568.

[27] A. R. S. Ponter, H. F. Chen, M. Ciavarella, G. Specchia, Shakedown analyses for rolling and sliding contact problems, International Journal of Solids and Structures 43 (14-15) (2005) 4201–4219.

[28] S. Foletti, H. Desimone, A semi-analytical approach for two-dimensional rolling/sliding contact with applications to shakedown analysis, Wear 262 (7-8) (2007) 850–867.

[29] M. Ciavarella, A. Cirilli, G. P. Demelio, Upperbound per il limite di shakedown nel rolling and sliding contact, in: Atti del XXX Convegno AIAS, 2001, pp. 141–153.

[30] P. D. Chinh, Shakedown theory for elastic-perfectly plastic bodies revisited, Fatigue & Fracture of Engineering Materials & Structures 45 (6-7) (2003) 1011–1027.

[31] H. Föll, Defects in crystals (2007).
URL `http://www.tf.uni-kiel.de/matwis/amat/def_en/index.html`

[32] V. A. Lubarda, Elastoplasticity Theory, CRC press, 2002.

[33] K. Dang Van, B. Griveau, O. Message, On a new multiaxial fatigue limit criterion: theory and application, in: M. Brown, K. Miller (Eds.), Biaxial and Multiaxial Fatigue, EGF 3, Mechanical Engineering Publications, London, 1989, pp. 479–498.

[34] J. A. Williams, The influence of repeated loading, residual stresses and shakedown on the behaviour of tribological contacts, Tribology International 38 (9) (2005) 786–797.

[35] J. W. Ringsberg, M. Loo-Morrey, B. L. Josefson, A. Kapoor, J. H. Beynon, Prediction of fatigue crack initiation for rolling contact fatigue, International Journal of Fatigue 22 (3) (2000) 205–215.

[36] K. Dang Van, M. H. Maitournam, On some recent trends in modelling of contact fatigue and wear in rail, Wear 253 (1-2) (2002) 219–227.

[37] M. Ciavarella, F. Monno, G. Demelio, On the Dang Van fatigue limit in rolling contact fatigue, International Journal of Solids and Structures 28 (8) (2006) 852–863.

[38] H. Desimone, A. Bernasconi, S. Beretta, On the application of the Dang Van criterion to rolling contact fatigue, Wear 262 (4-5) (2006) 567–572.

[39] A. B. Cardis, M. N. Webster, Gear oil micropitting evaluation, Gear Technology 17 (5) (2000) 30–35.

[40] S. M. M. Gonçalves, Modelação do micropitting nos dentes de engrenagens, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto (2006).

[41] J. O. Seabra, D. Berthe, Influence of surface waviness and roughness on the normal pressure distribution in the hertzian contact, Journal of Tribology 109 (1987) 462–470.

[42] I. A. Polonsky, L. M. Keer, A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques, Wear 231 (2) (1999) 206–219.

[43] C. Venner, Multilevel solution of the EHL line and point contact problems, Ph.D. thesis, Universiteit Twente, College van Dekanen (1991).

[44] P. Kumar, J. Mitchell, E. Yidirim, Computing core-sets and approximate smallest enclosing hyperspheres in high dimensions, in: Proceedings of the 5th Workshop on Algorithm Engineering and Experiments (ALENEX), 2003, pp. 45–55.

[45] P. Kumar, J. Mitchell, E. Yidirim, Approximate minimum enclosing balls in high dimensions using core sets (2003).
URL `http://www.compgeom.com/meb/`

[46] B. Gärtner, Fast and robust smallest enclosing balls, in: Proc. 7th annu. European Symposium on Algorithms (ESA), Vol. 1643 of Lecture Notes in Computer Science, Springer-Verlag, 1999, pp. 325–338.

[47] B. Gärtner, Smallest enclosing balls of points - fast and robust in c++. (2007).
URL `http://www.inf.ethz.ch/personal/gaertner/miniball.html`

[48] C. Gounet-Lespinasse, G. Liraut, Etude de l'influence des usinages de rectification et de superfinition sur la durée de vie en pitting des engrenages de BVM, Note de Synthèse n. 60152-998-0512.

[49] S. Bair, The variation of viscosity with temperature and pressure for various real lubricants, Journal of Tribology 123 (2) (2001) 433–436.

[50] S. Bair, J. Jarzynski, W. O. Winer, The temperature, pressure and time dependence of lubricant viscosity, Journal of Tribology 34 (7) (2001) 461–468.

[51] D. Bonte, C. Royer, Méthodologie pour la fatigue des dents d'engrenages, une première approche, Note Technique n. 60152-95-0493.