On tribological design in gear tooth contacts

Ellen Bergseth
On tribological design in gear tooth contacts

Ellen Bergseth

Doctoral thesis

Academic thesis, which with the approval of Kungliga Tekniska Högskolan, will be presented for public review in fulfilment of the requirements for a Doctorate of Engineering in Machine Design. The public review: Kungliga Tekniska Högskolan, Brinellvägen 85, room B319, at 10.00 on October 15, 2012.
Abstract

The correct tribological design will have a considerable effect on a gear’s service life and efficiency. The purpose of this thesis is to clarify the impact of variation in the gear tooth flank tribological system on the gear contact load capacity – to increase the understanding of how surface topography and lubricant interact.

In this thesis the variation in surface topography inherent in the manufacturing method has been shown, by experimental work and computer simulations, to be an important factor for the contact condition in the early life of gears. Surface analysis revealed that the formation and composition of surface boundary layers depends strongly on the chemical composition of the lubricant, but also on pre-existing surface boundary layers. Additionally, surface boundary layers play a major role in frictional behaviour, wear and in allowing the lubricant to react properly with the surfaces.

Paper A presents the current ISO 6336 calculation of surface durability. A robust design approach was used to investigate the extent to which the current standard for calculation of surface durability allows for manufacturing variations and the choice of lubricant.

Paper B investigates the extent to which a logarithmic profile modification can increase gear contact pressure robustness compared to traditional lead profiles for gears.

Paper C compares different gear manufacturing methods and their as-manufactured (fresh unworn) surface topographies, using measured surface topographies as input to a contact simulation program.

Paper D examines surface boundary layer formation and the corresponding wear in relation to different anti-wear additives in an environmentally adapted base oil.

Papers E and F make use of specimens with surface topographies imitating two gear manufacturing methods (grinding and superfinishing) to be used in a twin-disc and barrel-on-disc machine respectively. The contacts are analysed by friction measurements and simulations combined with methods for surface analysis.

Keywords: friction, gear contact, gear manufacturing, surface boundary layers, surface topography, wear
Preface

The research presented in this thesis was carried out between November 2006 and October 2012 at the Department of Machine Design at the Royal Institute of Technology, Stockholm.

First, I would like to express my thanks to my main supervisor Ulf Olofsson and co-supervisor Stefan Björklund for their excellent guidance and for their belief in me. I also would like to thank all involved in the Vinnova-financed KUGG program, with Sören Andersson and Ulf Bjarre as initiators. I would specially like to thank Sören Andersson for his warm welcome to Machine Design.

Special thanks go to Sören Sjöberg for cooperation, engineering know-how, and friendship during this time. I also want to thank Ulf Sellgren for valuable comments on my thesis, Vicki Derbyshire for proof-reading my English writing at all times and Krister Sundvall for a warm welcome and practical help during the early years of my time at the department. I am grateful for the comments made by Thomas Norrby on this thesis and for having trapped me in the world of research.

During this time, many people outside the department have taken the time to help, answer many questions, and discuss this subject. Among these are Marika, Åke, Milica, Pär, Stephen, Roger, Johan, Lisa, Alexander, Karl-Gustav, Hans H, Lars J, Mats B, Mats H, Mats L, Johan, Henrik, Bosse, Julia and Mathias.

Sitting in an office has given me the opportunity to get in contact with many different personalities. Thank you Anders, Björn, Martin G, Jens, Jon, Petter, Rasmus, and all my new colleagues in the A419 for being you. Special thanks to Kenneth for helping me put my warrior face on, Yi for co-authorship, and Minoo and Mario, who in the last year spiced up my interest for gears. Mario also came up with good comments at the very end of writing this thesis - and for that I am very thankful.

All my close friends and family (including the Johansson clan) have supported me and believed in me, and for that I am very grateful. Joje, Carin, Duvnästjeerna, Erik, Elsa, and Johanna – the loveliest fan club that ever existed. I thank my parents Brita and Inge for endless love and support – without you…, my daughter Bodil for making life more adventurous, and Rikard for being the most patient and beautiful person of mankind; I love you!

Stockholm, October 2012

Ellen Bergseth

Death: Are you ready?
A knight: My flesh is afraid, but I am prepared.
The Seventh Seal by Ingmar Bergman
List of appended papers

This thesis consists of a summary and the following six papers:

Paper A

Paper B

Paper C

Paper D

Paper E

Paper F
Ellen Bergseth, Yi Zhu, Ulf Olofsson, “Study of surface roughness and surface orientation on friction in rolling/sliding contacts: barrel-on-disc versus twin-disc”
To be submitted for publication
Division of work between authors

The work presented in this thesis was initiated and supervised by Professor Ulf Olofsson and Associate Professor Stefan Björklund.

Paper A
The work was performed by Bergseth.

Paper B
The work was equally divided between Bergseth and Björklund. Most of the writing was done by Bergseth.

Paper C
The experimental work was equally divided between Bergseth and Sjöberg. Björklund performed the simulation. All authors were equally involved in both writing and editing of the text.

Paper D
The experimental work was performed by Bergseth. Most of the writing was done by Bergseth and Torbacke, but Olofsson contributed. Both Torbacke and Olofsson supervised.

Paper E
Bergseth planned and performed the main work, including writing the paper. Test runs were led by S. Lewis. All authors were involved in editing of the text.

Paper F
Bergseth planned and performed the experimental work. The main part of the calculations was performed by Zhu. Bergseth wrote the main part of the paper. All authors were involved in editing of the text.
Contents
1 Introduction....................................................................................................................................... 1
2 Gear design and manufacturing....................................................................................................... 4
3 Tribology.......................................................................................................................................... 11
4 Methods.......................................................................................................................................... 20
5 Summary of appended papers....................................................................................................... 27
6 Discussion and conclusions............................................................................................................ 29
7 Future work....................................................................................................................................... 34
8 References........................................................................................................................................ 35

Appended papers
A. “Influence of gear surface roughness, lubricant viscosity and quality level on ISO 6336 calculation of surface durability”
B. “Logarithmical crowning for spur gears”
C. “Influence of real surface topography on the contact area ratio in differently manufactured spur gears”
D. “Wear in environmentally adapted lubricant with AW technology”
E. “Effect of gear surface and lubricant interaction on mild wear”
F. “Study of surface roughness and surface orientation on friction in rolling/sliding contacts: barrel-on-disc versus twin-disc”
1 Introduction

In 2005, the Swedish transmission industry had a turnover of 1.4 billion EUR and employed close to 6000 people [1]. Around the same time, the global gear market exceeded 39 billion EUR, three-quarters of which was in the automotive sector [2].

Gears are one of the most important means of mechanical power transmission, partly thanks to their high efficiency, which can reach over 99% for a gear pair [3]. The earth’s limited resources have forced us to use fuel more efficiently, and so there are increasing demands for even higher efficiency, longer service life and more power-density. In addition, customers have begun to make demands relating to environmental aspects such as noise pollution, full recycling and lubricant properties that may affect humans or the environment for a very long time after the product has been taken out of service. To meet these demands, designers and manufacturers have been forced to become aware of phenomena that were previously of less importance. The principles behind these phenomena affect friction, wear and lubrication, which are all covered in the science of tribology [4] - the subject field for this thesis. For example, today it is not just the presence of lubrication in a machine application that is of importance, but also the type of lubricant, where it is intended to act and how to apply it.

This thesis deals with high precision case hardened machined wrought steel gears, which among other applications are used in heavy trucks, but the results can be applied to a wide span of non-conformal hard elastohydrodynamic contacts, e.g. cams and rolling bearings. Figure 1 gives an overview of such gearing systems at different scales: a number of gear wheel combinations assembled in a gearbox, a gear pair in contact, a close-up of two gear surfaces in contact with deviations from a perfect gear profile, and surface active additives adsorbed with the polar end on a metal substrate. This thesis covers the whole spectrum, although with a focus on the small scale of gearbox tribology.

![Scale of details](image)

*Figure 1. Scale of details ranging from a heavy truck gear transmission (Scania GR875 range-change gearbox) [5] to the contacting surfaces on a macro to micro level and finally the contact boundary layers on a nano level.*

When producing high-precision gears, it is necessary to know the effects of manufacturing variations on the gear performance. Manufacturing variations have a number of causes including hardening distortions, individual machine variations, and the choice of manufacturing method; all manufacturing methods leave a unique system
fingerprint. Most research on the influence of gear manufacturing variations on gear performance looks at standard gear parameters related to the shape and position of the teeth, not the surface topography. However, a fresh unworn surface topography from manufacturing strongly affects how friction, temperature and wear in the contact progresses with time. If surface topography is changed by running, and if that change is slow and well controlled, the surfaces will run-in until they have become smooth enough not to wear any more, according to Jacobson [6]. However, he states that if the running conditions change by load, speed or temperature, running-in will start again. Sjöberg et al. [7], adopted a running-in procedure normally used for lifetime testing, found only minor changes in standard gear parameters when comparing three differently produced gears before and after running-in.

Gears rely on a well-designed lubricant (i.e. base oil and additive combination). The current international standard [8] procedure does not wholly reflect how the choice of lubricant influences load carrying capacity. Joachim and Kurz [9] used gear tests to investigate the various tooth flank manufacturing methods and lubricant combinations with regard to their effect on tooth flank service life. They point out that the standard procedures do not take into account either the thickness of the lubricant film or the influence of the surface structure, both of which affect tooth flank service life. Höhn et al. [10] give suggestions for adding an additional factor to the standard rating formulas in order to better correlate the actual lubricant performance and to evaluate the influence of lubricants on the pitting load capacity of case carburized gears.

The work presented in this thesis is part of a Swedish research project, initially known as KUGG and later named FFI Sustainable gear transmission realization, with the overall goal of increasing and strengthening our knowledge of gears in order to secure the future for gear design and manufacturing in Sweden. The project is linked to a range of ongoing projects [11-16]. The specific project goals of this thesis are to investigate the influence of variation in the gear tooth flank tribological system on the gear contact load capacity and to increase the understanding of how surface topography and lubricant interact in the mixed to boundary lubrication regimes. The following are research questions this thesis addresses:

1. To what extent is surface topography handled in today’s gear design tools?
2. What is the impact of surface topography on load capacity?
3. Can the influence of surface topography and lubricant interaction be measured?
4. Can a reduced surface roughness reduce friction and wear, and how will this influence the lubricant design?

This thesis is delimited into tribology testing and numerical computer simulations for studying the gear contact. The meshing of gears, how tooth deflections and friction transmit the torque (or force) and how it is related to shafts, bearings, gearbox casing etc., is not dealt with in this thesis. The road map for the six papers included in this thesis is given below.
Paper A presents the current ISO 6336 calculation of surface durability, and can be seen as a literature survey. The aim was to use a robust design approach to investigate the extent to which the standard treats variations in surface roughness, lubricant viscosity, and gear quality (i.e. accuracy) level. An additional aim was to use the standard without the need for experimental testing, extensive knowledge, or advanced calculation. Thus, the simplest calculation method according to the ISO standard was used.

In paper B, a logarithmic lead profile was compared with traditional lead profile modifications for gears. The profiles were applied on a spur gear pair, and a numerical method for contact analysis was used to calculate the contact pressure distribution. The aim was to find out the extent to which a logarithmic profile can increase the load capacity and robustness in gears.

The aim of paper C was to find a way to compare different gear manufacturing methods and their inherent surface topographies. The 3D surface topography was measured on new (unused) spur gear tooth flanks produced with four different manufacturing methods. The measured topographies were used as input to a contact analysis program, and the contact ratio (ratio between real and nominal contact area) for two different normal loads was calculated at different mesh positions.

The purpose of Paper D was to examine the formation of tribofilm and the corresponding wear occurring in the boundary lubrication regime in environmentally adapted lubricants, when using synthetic ester base fluids with different anti-wear additives. Wear was studied using a pin-on-disc machine. The worn disc surfaces were analysed to reveal the surface reactions formed by the additives.

In paper E, a twin-disc machine was used to simulate a rolling/sliding gear contact for three surface finishes, each run with two types of lubricants. The combinations were compared under set operating conditions to evaluate their effects on friction, wear, and the build-up of the surface boundary layer. The purpose was to find out the extent to which surface topography can change the lubrication regime and its influence on friction and wear. Another aim was to discover the extent to which a pre-formed chemically reacted boundary layer can help to reduce and control friction and wear.

In the final paper F, ground and superfinishing manufacturing methods commonly used for gears were imitated on discs to be used in a barrel-on-disc machine. The purpose was to evaluate if friction and wear can be lowered by significantly reducing the surface roughness. An additional aim was to compare the results with the results from paper E. Thus all test parameters, except the lubricant temperature, were set close to the parameters used in Paper E.
2 Gear design and manufacturing

Gear design and manufacturing involves a range of activities - it is not only about choosing a helical or spur gear pair that can transfer torque between parallel axes. Depending on, for example, types of bearings, preloading, the gearbox casing, gear web, material, heat treatments and the interaction between the meshing gear surfaces, the gear performance will change. This thesis addresses the latter – it involves the tribological design restricted to the gear tooth flank surface. Surface roughness is linked to the manufacturing method itself, for example relevant aspects include the grain size of the grinding tool and any tool tip irregularities. The surface roughness then affects surface pressure and, indirectly, the fatigue limit and wear.

2.1 Gear design

Gear design has a long tradition of empirical testing and standardisation work for dimensioning procedures, beginning in the early twentieth century. Today, the use of computers for simulating gear contact has become almost a natural tool for gear designers. This does not make gear design trivial by any means; rather the opposite. The gear designer has the struggle of constantly weighing the effect of any proposed tolerance on gear performance and cost, and of specifying proper values. A unique tolerance, that can range from just a few to hundreds of micrometres, is set for each gear parameter during gear design in order to adjust for manufacturing variations and to achieve the desired gear performance. The values of the gear parameters and their tolerances will influence gear performance in a number of ways. Tooth thickness, for example, directly relates to backlash, affecting positioning and motion control. The accumulated pitch (tooth-to-tooth spacing) variation is a critical factor for gear ratio accuracy. For noise, the most important aspect is profile (involute) variations [17], due to stiffness variation during the gear mesh. For load capacity, tooth alignment (lead) is probably one of the most critical factors, since edge loading results in a shortened service life.

Gear parameters describing the gear teeth can be divided into those relating to macro geometry and micro geometry respectively. Macro geometry includes aspects such as module, number of teeth, and pressure angle; while micro geometry includes intentional tooth modifications such as lead crowning and gear standard quality levels (e.g. flank deviations). A more general definition of surface topography includes form, surface waviness, and surface roughness. No distinct borders exist between these three, which are separated only on the basis of the objective and the surface wavelengths of interest.

Standard rating formulas for calculating load capacity (surface durability) are based on Hertz contact theory but include numerous factors in order to create realistic evaluations of the stress levels encountered by a gear pair. Preferably, a gear designer should look at the performance requirements of a gear and use tolerance tables such as those included in ISO 1328-1 [18] to select the quality level that will meet these requirements. However, the standard is restricted to using only a few tolerance parameters which are not linked to a specific performance requirement such as noise or strength. Further on, Houser [19] studied the partially equivalent American Gear Manufacturers Association (AGMA)
standard and concluded that several factors likely to be affected by manufacturing are not quantified in terms of quality levels from standard tolerance tables.

2.2 Gear modelling

Gear technology is quite complex and requires a broad understanding of the gear geometry and applied theory [20, 21]. In the absence of complete existing solutions many gear researchers develop their own models. The finite element method (FEM) has become the prevalent technique used for analysis and simulation of gears [22]. The method is used for studying the mechanical behaviour of gears influenced by various factors, such as the tooth profile, the gear fillet, rim thickness, tooth modifications, assembly errors, and so on. Li [23, 24] developed a 3D FEM to calculate surface contact stress and root bending stress for a pair of spur gears with machining errors, assembly errors, and tooth modifications. Li suggests that FEM should be used instead of standard rating procedures for gear pairs exposed to these sorts of errors. Wei et al. [25] presented a similar study on helical gears. Mao [26] applied micro geometry modification to improve fatigue performance on spur gears with axis deviation using FEM. Kahraman et al. [27] predicted the wear of helical gears by taking out the contact pressures from FEM and calculating the sliding distance combined with Archard's wear equation. Combination of FE and standard calculation methods have become available in commercial gear design software packages such as LDP and KissSoft [28, 29], and are usable even when the design chosen does not conform to the design considered in the standard. However, none of these deal with rough surfaces models. An assumption that the surfaces are smooth is often made, since sufficient resolution in the contact area would make the size of the FE model very large. The problem can be addressed by combining a surface integral (or numerical) solution and finite element solutions [30, 31]. However, these methods also contain an assumption about when the two solutions are valid, and so the potential for error is still present. Vijayakar’s [31] solution is made use of in the commercial Helical3D program [32]. An image from Helical 3D is shown in Figure 2.

![Contact pressure distribution](image)

*Figure 2. The coarse FEM grid compared to the small contact area where the contact pressure distributed using a finer contact grid.*
Hertz’s theory remains the basis for the analysis of most contact problems. It is based on the following assumptions: the surfaces in contact are smooth, the contact shape is a point or a line at no load, the contact semi minor axis is generally small compared to the reduced radius of curvature for both surfaces in contact (the half-plane assumption), the strains are sufficiently small for linear elasticity to be valid, and the contact is frictionless so that only normal pressure is transmitted. Today, however, contact problems can be solved without many of these of assumptions.

In the rough case, the load will be carried by several contact spots. Figure 3 shows how the real contact area is formed by the sum of the contact spots. The nominal (or apparent) contact area is formed when the surfaces are perfectly smooth. There are two types of approaches when modelling the contacts between rough surfaces; the numerical and the statistical. Numerical approaches are used for limited contacts where the actual topography of the surfaces is known and where it is possible to calculate the real contact pressure due to the real (or measured) surface topography or a surface created numerically by the user. Björklund and Andersson [33] developed a numerical model for studying micro-slip contact phenomena where the contact patch is subjected to both normal and tangential loading. This model has been verified with a FE-based surface model by Sellgren et al. [30]. Zhu [34] later added the effect of rolling/sliding and fluid to the contact model developed by Björklund and Andersson. Almqvist et al. [35] developed a similar deterministic model to be used for numerical simulation of the contact of linear elastic and perfectly plastic rough surfaces. Statistical approaches, on the other hand, make use of the stochastic nature of rough surfaces and are not concerned with the exact topographies of the surfaces. Greenwood and Williamson [36] made a first approach to this by assuming a convenient shape for the asperities (e.g. spherical) and using statistical parameters to describe their heights and sizes in order to compare different kinds of surface roughness.

![Figure 3. Schematic contact between a rough surface and a smooth sphere [37]. The nominal contact area is marked with a hatched line; contact spots form the real contact area within this.](image)

2.3 Gear manufacturing

Gears are often manufactured in a sequence of machining operations. Since highly loaded gears are hardened, it is convenient to separate machining operations before and after hardening into soft and hard machining operations respectively. The first operation is usually cutting; one common cutting process is hobbing, which is illustrated in Figure...
4. Hobbing is a widely used machining process for generating involute gear profiles. It is based on rack-cutter generation where the surface is obtained as the envelope of a series of cutter surfaces formed during a continuous motion between workpiece and cutter [21]. The rack-cutter profile is a cutting edge whose straight sides and fillets machine the flanks and the gear fillet respectively. Bergseth [38] used a computer-aided design approach to study how the tool wear on a hob reveals its eigencharacteristics on the tooth flanks. However, the scale used for simulating these defects was limited by the computer and program resolution capacity.

![Illustration of a hob and gear wheel setup (a) and a close-up on gear tooth generation (b).](image)

Hobbing today is rarely used as a finishing process. To adjust for the inaccuracies inherent in the cutting process, a further soft operation known as shaving can be used. The shaving tool can be visualised as a helical gear with gashes in the flanks working as cutting edges meshed with the workpiece in a crossed axis relationship. Shaving can be performed using different processes (e.g. parallel, plunge, and diagonal shaving), with feed and sense of rotation being handled in different ways. By removing small amounts of material, shaving can enhance the overall quality of the gear tooth. Shaving can be used as a finishing operation, and so machine settings need to be chosen to make allowance for distortions after hardening. Winkel [39] points out that with new developments in gear hobbing, including new tools and machines, the quality gap between finish hobbing and shaving continues to narrow.

Hardening can be performed in various ways depending on, for example, the selected steel and desired machinability. Gears are usually case hardened; the gear is given a carbon (or nitrogen) rich outer surface which among other things enhances the surface hardness and increases the fatigue life [40]. The machinability of the material is of great interest, since gears are heavily machined to close tolerances. Björkeborn et al. [16] present some suggestions for how the specification of the material should be modified in order to create a more robust and predictable production process.

If increased quality is needed, honing or grinding can be used. These two finishing techniques are both hard machining operations. According to Dugas [41], the process of honing was developed to improve the sound characteristic of hardened gears by a) removing nicks and burrs, b) improving surface finish, and c) making minor corrections
in tooth inaccuracies caused by hardening distortions. Similarly to shaving, honing is characterised by the workpiece (i.e. machined gear) and the honing tool meshing with their axes intersecting. The honing tool, however, is formed close to as an internal helical gear and made of abrasive ceramics. Honing in this study refers to the ‘power-honing’ method which involves both the workpiece and tools being numerically controlled by electric drives during the process. This makes it possible not only to enhance the surface, but also to better adjust the gear tooth shape and tooth-to-tooth deviations [15] and achieve properties comparable to those achieved with grinding (Paper C). Grinding, on the other hand, uses an abrasive-coated wheel as the machining tool. There are essentially two techniques for grinding gears: continuous generating grinding and discontinuous profile grinding. The first is characterised by the arrangement of tool and workpiece as a worm gear, and the second by each tooth gap of the machined piece being processed separately. Grinding produces a series of hills and valleys, resulting in a characteristic surface lay in the face width direction (Figure 5).

Superfinishing of gears has mainly been used in the aviation industry, but has recently also become of interest to the automotive industry, since it can increase the surface fatigue life of gears [42]. Superfinishing is a generic term for techniques that produce very smooth/polished/near mirror-like surfaces, and there are a number of different ways to achieve this. One method is vibratory grinding, where the workpiece is placed in a vibrating container together with abrasive bodies (plastic or ceramics) which remove roughness peaks and round the edges [7]. According to Zhang and Shaw [43], gear superfinishing is an improved traditional vibratory honing process with the addition of mild chemicals designated to achieve a mirror-like tooth surface (with Ra<0.1 µm) without degrading the original gear accuracy. Winkelmann et al. [44] describe a superfinishing process known as isotropic superfinish, which utilizes high-density nonabrasive media leaving a unique surface that has no directionality.

Sometimes designers chose to use manganese phosphate coating on their gears as a final finishing step which is known to reduce the risk of scuffing [45]. Shoot peening may also be used as a final manufacturing step in order to increase the ductile stresses in the root by means of introducing internal compressive stresses. Neither manganese phosphating nor shoot peening are intended to remove material from the surface. However, shoot peening usually increase the surface roughness. Furthermore, manganese phosphate coating damage the surface by creating cracks perpendicular to the surface [46].

2.4 Tooth flank surface topography

Since gear flanks in contact are usually separated by a very thin lubricant film (≤ 1 µm), it is necessary to understand the nature of the surface topography. Surface topography is traditionally measured using stylus instruments; a fine diamond stylus is drawn over a surface, and the vertical movements reveal variations from the theoretical surface. Figure 5 shows the results of using a stylus instrument to measure four gears manufactured by hobbing, plunge green-shaving, power-honing, and generating grinding; these are the same gears that were used in Paper C.
Figure 5. Surface topography of four common manufacturing methods. The 1 x 1 mm areas have had the form removed by a fifth degree 3D polynomial. All topographies make use of the same height scaling: -5 to 5 µm.

Gear flank measurements as shown in Figure 5 are normally not a part of the standard sequence of operations involved in gear manufacturing. Instead, the complex shape of the gear tooth has resulted in unique gear parameters and inspection techniques. Coordinate measurements are commonly used to measure the accuracy of the tooth form, where a tactile probe or stylus can detect the tooth form deviation. Figure 6 provides two illustrations of how the single gear tooth flank form is measured along the involute profile and face width (lead), detecting aspects such as profile form and lead form variations. Tactile methods are robust but time consuming compared to, for example, optical methods. However, the big challenge for optical instruments is to have a large working area on steep flanks to detect form errors and at the same time get high vertical and lateral resolution. This demands enough working distance to register the whole flank from the root to the tip. Beyond the gear parameters used for detecting form deviation, there are a variety of surface evaluating parameters, which must be carefully used.
All machining operations have process marks, at least on some scale. Irregularities with longer wavelengths (known as waviness) can stem from the machine itself - for example, an unbalanced grinding wheel can create vibrations between the workpiece and the grinding wheel. Form errors are irregularities of even longer wavelengths, caused by flexing or bending in non-rigid workpieces [47], or by thermal distortions. Form errors usually produce one or two undulations over the length of the assessed surface. The accuracy of a machining system is mainly determined by static and dynamic characteristics, but according to Archenti [48] stiffness is also an important yet rarely-measured factor. Archenti notes that high static stiffness is needed in order to reduce the deflections as much as possible, since deflections affect the geometry and surface roughness of the machined detail.

The origin of gear noise is the gear mesh. Noise performance remains an important gear issue, since among other reasons it may cause environmental noise pollution. Amini et al. [49] characterized gear tooth surfaces obtained with shaving, honing, and grinding with respect to the surface functional properties; they found that the surface parameters linked to noise activity included, for example, the main direction of the surface texture. Åkerblom [50] performed experimental investigations of the influence of different gear finishing operations and gear manufacturing variations on gearbox noise and vibration. The results provided evidence that, for example, shaved gears do not seem to be noisier than ground gears even if their gear deviations are larger (i.e. have less accuracy). A rougher surface though, increased lead crowning, and helix angle variations seem to increase noise. However, Åkerblom also found strong evidence that gear noise is affected by factors other than the gears themselves, for example assembly-related variations. The static transmission error is sometimes used for simulating form variation impact on mesh quality, since transmission error is considered to be the main excitation mechanism of gear noise. However, Henriksson [51] found that measured (or dynamic) transmission error and noise show weak correlation to calculated static transmission error. A reduction of the static transmission error therefore does not necessarily mean a reduction of gear noise.
3 Tribology

Tribology, derived from the Greek word *tribos* meaning rubbing [52], is the science and technology of interacting surfaces in relative motion and of related subjects and practices. It includes friction, wear, and lubrication. Friction can be directly correlated to gear contact power losses and temperature rise [53]. Gears generally have a slow continuous wear process which eventually causes loss of gear accuracy. Lubrication is the most common way to reduce friction and wear in terms of building up easily-sheared boundary layers; it is also a common way to transfer heat from the contact zone. Choices made during design and manufacturing affect all these aspects. Over the years, tribology has focused towards smaller and smaller scales in the investigation of these phenomena. This progress goes along with the tools of surface science.

3.1 The gear contact conditions

Gear tooth flanks in contact do not geometrically conform to each other; it is a non-conformal contact, with small contact area. In metal gears, the influence of pressure (maximum pressure can reach 1-4 GPa) on the lubricant viscosity plays an important role together with the elastic deformation of the bodies. A lubricant subjected to 1 GPa could undergo a 22000-fold increase in viscosity compared to atmospheric pressure viscosity [54]. In practice, the lubricant becomes glass-like and behaves more like a solid than a liquid [55]. This phenomenon is known as hard (i.e. metallic surfaces) elastohydrodynamic lubrication, where there is no asperity contact at full-film elastohydrodynamic lubrication. Depending on operating conditions, two mating gear flanks will at a certain state be subjected to more or less direct metal to metal contact (and may cycle between different lubrication conditions).

Stribeck [56] studied the general relations between the friction force and the lubricant viscosity, sliding speed, and load on journal bearings; this resulted in the creation of the *Stribeck curve* shown in Figure 7. Depending on the amount of direct contact between the surfaces, the lubrication regime is classified as boundary lubrication, mixed lubrication, or full-film lubrication. At full-film lubrication, the surfaces are totally separated by the lubricant; no metal contact occurs, and the friction is caused by shear forces in the viscous lubricant [57]. Typical of elastohydrodynamic lubrication is that, unlike in hydrodynamic lubrication, lubricant film formation and friction is almost entirely decoupled and friction depends primarily on the molecular structure (molecule linkages) of the base fluid and the contact temperature [54]. As can be seen in Figure 7, the friction remains broadly constant for the elastohydrodynamic condition. In the mixed (or partial) lubrication regime, the lubricant film separating the contacting teeth is thin – usually of the same order of magnitude as the roughness – which may cause partial lubricant film breakdown. Finally, under boundary lubrication, the friction coefficient is typically high, the reduction of friction and wear depends on the lubricant chemistry, and the fluid film has no effect [58]; the load is mainly carried by asperities in contact. The lubricant regime can be estimated by calculating the lambda ratio (or the specific film thickness) $\Lambda$, which is the calculated minimum film thickness for ideal smooth surfaces divided by the...
composite surface roughness. \( \Lambda \leq 1 \) represents boundary lubrication, \( 1 < \Lambda \leq 3 \) represents mixed lubrication, and \( \Lambda > 3 \) represents full-film lubrication. Spikes [59] describes the generation from full-film elastohydrodynamic lubrication to boundary lubrication by describing events as the lambda ratio is progressively lowered from \( \Lambda > 5 \) and that even at \( \Lambda \) near 0.5, the friction coefficient is caused by both elastohydrodynamic and boundary friction.

![Figure 7. The coefficient of friction varies depending on lubricant viscosity (\( \eta \)), sliding speed (\( v \)), and load (\( P \)). BL: boundary lubrication; ML: mixed lubrication; FL: full-film lubrication; HD: hydrodynamic; EHD: elastohydrodynamic.](image)

The meshing of a pair of involute spur gear teeth is illustrated in Figure 8. The line \( I_1I_2 \) is the common tangent to the two base circles from which the involute profiles are generated. This line is known as the line of action. At point \( P \) pure rolling occurs, and at all other points on the line of action a combined rolling and sliding motion takes place. The slide-to-roll ratio, contact conditions, radius of curvature, and entrainment speed for any specific contact position on the involute tooth flank can be estimated using Hertz’s theory from the information in Figure 8 together with fundamental equations; see Johnson [60] for a thorough explanation. Thus the rolling/sliding gear contact can be represented by two cylinders in contact (Paper E).
Figure 8. Schematic view of two involute gear teeth in mesh. Any point on the line of action can be represented by two equivalent cylinders.

The surface topography of interacting surfaces has a strong impact on the discipline of tribology, since it has an effect on the contact behaviour. In the boundary and mixed lubrication regimes, when the average lubricant film is close to or less than the composite surface roughness, the load is wholly or partially supported by the contacting surface asperities. Collisions here may cause local high pressures and temperatures, plastic deformation, and sometimes crack growth in the bulk material. The real contact area ratio, calculated by use of surface measurements combined with numerical contact pressure calculations, can be used to assess the extent to which asperities are carrying the load (paper C).

3.2 Lubrication

Lubricants perform a number of functions: transferring heat, controlling wear, reducing friction, carrying away contaminants and debris, preventing corrosion, and reducing noise and vibrations. In a gearbox, lubrication is usually applied by so-called splash lubrication; the gears dip into a reservoir of oil in the bottom of the gearbox and splash the oil around. The gear pairs in a gearbox rotate even if they are not transmitting any load. These load-independent losses are an important source of the total transmission losses, and are most easily decreased by lowering the viscosity. However, this does not necessarily have to lower the load-dependent contact power loss.

The performance properties, long life properties, and environmental properties are all important when choosing a lubricant. The performance properties of a lubricant (viscosity, thermal properties, high temperature properties, and sensitivity to contaminants) are important from the very instant they are added to the contact [61]. Long life properties will affect surface life, and the environmental properties
(biodegradability, renewability, toxicity and bioaccumulation) of a lubricant will have an influence for a very long time after the product has been taken out of service. Environmental demands and legislation gets tougher every year [61].

Base oils behave differently depending on their chemical structure (i.e. on how they are refined). Gear lubricants usually originate from petroleum (crude oil), and are generally a complex mixture of hydrocarbons – compounds of carbon and hydrogen. The hydrocarbons present in base oils are paraffins, naphthenes, and aromatics, and their different molecular structure will for example influence volatility, solvency, pour point, polarity (or surface activity), oxidation stability, viscosity, and viscosity index (VI). Figure 9 illustrates the origin of different base fluids. In this work synthetic polyalphaolefin (PAO) based and synthetic ester base oils are used. In a synthetic lubricant, the structure can be designed to have specific properties. PAO base fluid originates from crude oil via ethane chemistry, and its characteristics include high VI, low volatility, low polarity, and high thermal stability. Synthetic esters are produced by reacting an alcohol with a fatty acid. The alcohol is usually of petrochemical origin, while the fatty acid is prepared from natural vegetable and animal oils and fats. There is a wide variety of both fatty acids and alcohols. Hence, there is a wide range of esters available (e.g. monoesters, diesters, polyolesters, and complex esters). Esters can be highly polar and may be added to PAO in order to improve the solubility of additives. Synthetic esters have a higher biodegradability and amount of renewable raw material (as the arrow in Figure 9 implies), and are also generally more expensive than PAO base oils [62]. The market is still dominated by mineral oil based lubricants, due to low costs. As a result of this, most of the standards are based on tests using mineral oils. Höhn et al. [63] found that with synthetic oils, oil ageing (caused by high operational temperatures and long exposure times) had only a minor effect on gear life, in comparison to a reference mineral oil.

![Figure 9. The origin of different base fluids (courtesy of Marika Torbacke et al. [61]).](image-url)
3.2.1 Viscosity

The concept of viscosity describes a fluid’s internal resistance to shear, and may be thought of as a measure of fluid friction. Viscosity is extremely sensitive to both pressure and temperature. Lubricants behave almost like a Newtonian fluid, with a linear relationship between shear stress and shear rate, but the influences of pressure and temperature change this linear relationship. The dependence of viscosity on temperature can be predicted from the Roelands equation [57]. However, the viscosity-temperature dependence is generally described by the VI, where a high VI value indicates smaller viscosity changes with temperature. Industrial gear oils are classified according to the ISO viscosity grades (VG) system, based on the kinematic viscosity of the oil in centiStokes at 40 °C and 100 °C. Automotive gear oils, on the other hand, are classified according to the J306 standard published by the Society of Automotive Engineers (SAE). Under this standard, a gear oil classified as 75W/85 has a low temperature viscosity according to the grade 75W (W for winter) and a high temperature viscosity according to grade 85. Lubricants with a little viscosity temperature dependency usually also show less pressure viscosity dependence [55]. The viscosity dependence on pressure (i.e. pressure viscosity coefficient or piezo-viscous coefficient) can be predicted from either the Barus or the Roelands equation [57]. The pressure viscosity coefficient describes the lubricant’s ability to form a film. Another important parameter in the elastohydrodynamic contact is the limiting shear stress, which describes the maximum shear stress a lubricant can withstand at a given pressure and temperature. The limiting shear stress is related to the elastohydrodynamic friction coefficient. In engine applications a high bulk viscosity is not ideal since it impacts the hydrodynamic friction losses. Instead a high pressure viscosity coefficient is desirable. However, there is research showing that a high pressure viscosity coefficient correlates with high friction losses [54], thus the thicker film the higher the friction coefficient. How these two properties correlate has also proved interesting in minimizing the energy losses in transmissions.

3.2.2 Formation of boundary layers by the additives

A lubricant is a carefully-formulated mixture of base oil(s) and additives. These additives must be able to perform their function both in the lubricant bulk and on the gear surface. For example, friction modifiers are highly polar molecules which are often used as additives in lubricants [62]. If then the base fluid itself is highly polar, it generally diminishes the effect of additives. Additives may also enhance or counteract each other’s actions [61]. Mixtures of base oils may help to improve aspects such as the solubility of additives. Surface active additives must have the proper conditions to adsorb to the surface. This action usually depends on the temperature, the load, and the activity of the surface and the additives. The additives that are of interest in this thesis are those that modify the metal surface. This happens in the mixed and boundary lubrication regimes, where temperatures rise in the contact zone due to frictional heat. Friction modifiers (FM) and anti-wear (AW) additives are mainly active in the mixed lubrication regime, whereas extreme pressure (EP) additives are active in the boundary lubrication regimes. In the boundary lubrication regime, the competition of the EP additives with the FM and
AW additives is of minor importance [61]. The FM additives reduce the friction by physical adsorption of polar materials on metal surfaces. FM additives are more easily sheared than AW additives, which form a protective layer by adsorption or through mild chemical reaction with the metal surface [55]. AW and EP additives initially mainly chemisorb to the surface, but if high loads and high shear occur, the hydrocarbon chains of the surface-activated additives will be removed, which will allow a reaction to occur between the polar moiety (e.g. sulphur) and the metal, forming for example iron sulphide [61], which has a hardness four times lower than steel and shear strength six times lower than iron [64]. A schematic illustration of the boundary layers is shown in Figure 10. This surface boundary layer fills the surface valleys and facilitates effective film formation, and thus reduces friction and prevents local welding and wear [65].

![Figure 10. Illustration of the boundary layers or modes of additive molecules (surface interactions) on a metal surface. Partly adapted from [65].](image)

Most AW and EP additives contain sulphur and phosphorus additives. Sulphur additives such as disulphides and polysulphides decompose on the surface above 200 °C to form a protective sulphide layer; the thickness of this layer depends on the quantity and the lability of sulphur in the additive [65]. Phosphite or metal phosphate protective layers form at higher temperatures than sulphur layers. Surface analysis techniques such as glow discharge-optical emission spectroscopy (GD-OES) can be used to study these layers and understand the role played by the additives. Figure 11 (from Paper D) shows two analyses for carburized and ground surfaces – one new as-manufactured surface and one run in a lubricated twin-disc test. The formation of oxide films is also known to play an important role in the reduction of wear and friction in sliding contacts between metal surfaces [66]. The depth at which the amount of oxygen is higher than that of iron is indicated with a line in Figure 11. Zhao et al. [67] noted that iron oxide layers play a role in reducing friction in steels, and their analysis also revealed that the active sulphur and phosphorus additives modified the wear track layer chemically. Shibbe et al. [68] made use of scanning Auger microscope images of the gear tooth surfaces to show that gears tested with sulphur and phosphorus EP additives had reactive surface films that were 20-40 nm thick, similar to the results shown in Figure 11b. If active elements are mainly physisorbing to the surface, this will not be captured by the GD-OES technique. Most additive manufacturers (not always the same as the lubricant supplier) do not provide useful information about the nature of the additive packages; however, the amount of sulphur and phosphorus can be measured.
3.3 Wear

Wear can be defined as the removal or displacement of material from surfaces in contact as a result of mechanical, chemical, or electrical action. Wear can be classified as mild, moderate, or severe [69]. Wear can include several wear mechanisms (e.g. adhesive wear, abrasive wear), and the predominant wear mechanism gives rise to a specific failure mode - for example, scuffing is a sign of severe adhesive wear and is recognised by roughing bands in the involute profile direction (Figure 12a). Generally, the failure mode of scuffing is based on a critical contact temperature at which the lubricant film fails and micro-welding between asperities takes place [16]. Gears mostly have a slow continuous wear process, but gear life is usually deliberately limited by surface fatigue, since a tooth breakage will immediately stop the gear function. Pitting, shown in Figure 12b, occurs when the limits of the surface durability of the meshing flanks are exceeded and particles break out of the flanks leaving pits [8]. The standard calculation procedure for surface durability is for pitting and does not have to be linked to, for example, scuffing. Wear and surface fatigue are often distinguished, since surface fatigue comes from repeated contact stresses. Running-in of the surfaces can be seen as a mild wear condition; it should be noted that mild wear is considered normal in many applications. However, mild wear is also likely to act as a catalyst for fatigue wear, as mentioned by Flodin [70]. Mild surface wear of gear flanks has gained significance in recent years [27] due to advances in modelling capabilities, cleaner gear steel, and advances in engineered surface technologies such as shot peening, chemical polishing, and thin film coatings. Mild wear shows no signs to the naked eye, but the gear flanks may get a mirror-like surface.
Micropitting is one of the most common failure mechanisms of highly stressed case carburized gears [72], and can degrade the gear tooth accuracy by causing non-uniform wear [73]. Micropitting failure was found on case carburized ground surfaces run in a twin-disc machine (Paper E). Micropitting can be regarded as a fatigue failure, with cracks propagating at a shallow angle to the surface and pits that are just a few micrometres deep [74]. Micropitting (or microspalling) is classified as an advanced surface distress, and is common below the pitch line (dedendum) where sliding traction reduces fatigue life [75]; that is, where the sliding and rolling directions are opposite for the driving and the driven gears. There is no full explanation of the main factors contributing to micropitting. Twenty years ago it was not considered a problematic failure mode [76, 77], but the problem later became widespread and its solution now has a high priority [72].

Photos of the run disc surface (paper E) are shown in Figure 13. In Figure 13a three distinct areas are detectable: direct contact close to the contact centre (to the right of the image), a highly polished area, and the part of the surface where no contact has occurred. The polished area is most likely due to particles acting as polishing agents. Figure 13b shows a close up of the wear track. The pits are approximately 30 to 80 µm, and scratches in the rolling direction are visible. Figure 14 shows a radial and an axial cut through the disc material. Micropits are visible in Figure 14a, and cracks with branching are visible in Figure 14b. The entry angle of the crack is 20-30° in the direction of rolling, close to what Alfredsson et al. [78] found when studying the role of a single surface asperity in rolling contact fatigue, having similar contact pressures and asperity heights and width.
Figure 13. Light optical microscope photos of a run test disc. Arrows indicate the rolling direction.

Figure 14. Microphotographs of the discs shown in Figure 13 prepared with 2% nital etch: a) radial cut showing pits in the transverse direction and (b) axial cut showing cracks angled in the rolling direction (indicated by the arrow). Courtesy of A. Drott.
4 Methods

This chapter presents the experimental methods and simulation tools used to efficiently search for answers to the research questions.

4.1 Tribology testing

Gear contact involves a number of time-varying parameters, which makes the analysis very complex. Attempts have been made to measure friction through the gear mesh cycle [79], but in order to reduce the number of parameters influencing the contact conditions, tribotests are commonly used. Figure 15 presents the three different methods for tribology testing (or contact configurations) used in this research.

By measuring and weighing the specimens before and after operation, the wear mechanism and material wear can be estimated. The wear mechanism indicates whether the operating conditions reflect the simulated system, and the less material worn off the material the better the lubricant. The friction or traction coefficient (traction is often used as a substitute for friction in elastohydrodynamic lubrication) is measured during testing and in gears, the lower the friction the better. Furthermore, a change in friction behaviour may indicate a change in contact conditions and wear mechanism.

The specimens used in twin-disc and barrel-on-disc set-ups were manufactured (Papers E and F) by grinding and polishing, referred to by their final machining process step. Grinding was carried out perpendicular to the rolling directions corresponding to the real gear surface. Polishing means that the surfaces are near mirror-like, but the polishing can alternatively also be used as a substitute for superfinishing. In the pin-on-disc test set-up (Figure 15a), the sliding component of the contact is experimentally simulated, giving a test condition that corresponds most closely to, for example, the beginning of the mesh cycle where the sliding velocity between the teeth is high. The barrel-on-disc test setup (Figure 15b) allows for combined sliding and rolling motions, which better corresponds to the real gear contact situation compared to the pin-on-disc set-up. The contact in the barrel-on-disc setup differs from that in the pin-on-disc setup; instead of a point, an elliptical contact is created with its semi-major radius in the rolling/sliding direction. The twin-disc test method (Figure 15c) also permits a combination of sliding and rolling, and is similar to the real spur gear contact in terms of the effective radius and the fact that the semi-major axis is perpendicular to the rolling/sliding direction. The obvious shortcoming of all of these contact configurations is that they model steady conditions rather than the cyclic nature of gear meshing, in which the sliding velocity varies in magnitude and direction, and the load is carried alternately by one or more pairs of teeth.

In this thesis, slip (or slip ratio) is defined as the speed difference divided with the mean entrainment speed expressed in percentage.
4.1.1 Pin-on-disc

The pin-on-disc machine used in Paper D consisted of a horizontal rotating disc and a calibrated dead-weight-loaded pin with a spherical end that slides against the disc. The normal load, temperature, and humidity were kept constant during each test run. The friction force was calculated from the constant normal load and the tangential load, both of which were measured using a load cell. A small brush was used to apply the lubricant continuously with the aid of pressurised air. Due to differences in sliding distance, the wear depth was significantly bigger on the pin than on the disc, and so the wear was measured only on the pin.

4.1.2 Barrel-on-disc

A Mini Traction Machine (MTM) barrel-on-disc test set-up was used in Paper F (Figure 15b). The barrel is loaded at an angle to the disc, and the barrel and disc are driven by two separate motors to create an appropriate slip. The lateral force exerted on the barrel is measured using a force transducer that further provides the friction coefficient. The disc is submerged in a lubricant bath that is temperature controlled during testing. All test parameters such as load, speed, and temperature are controlled by a computer. Stribeck curves (or traction tests) are created by increasing the rolling speed while keeping temperature, slip, and normal load constant. Friction is measured at each step by using the average value of friction forces from two measurements with the same specified slip. This procedure reduces the effect of offset errors in the friction measurements.

4.1.3 Twin-disc

The twin-disc test machine has been widely used to simulate gear contacts [53, 80]. The twin-disc machine, shown in Figure 16 and used in Paper E, is based on a Colchester Mascot 1600 Lathe that provides a drive system for the upper specimen and a solid machine bed on which the rest of the machine is mounted. The discs are hydraulically loaded together and continuously driven at controlled rotational speeds by independent electric motors. The required slip is achieved by adjusting the rotational speeds, and
calculated approximately once per second with the motor control signals being adjusted as required after each calculation. A torque transducer is assembled on the top disc drive shaft to provide friction measurements, and a load cell is mounted in line with the hydraulic loading piston below the pivoted bearing housing and gives a direct measurement of the normal load. Load is controlled manually during the testing on the basis of the readout. All data are acquired by a personal computer. The machine is discussed in greater detail by Fletcher and Beynon [81] and by Lewis and Dwyer-Joyce [82].

Figure 16. Photo of the twin-disc machine.

4.2 Surface topography

In the research reported in this thesis, a stylus device Form Talysurf PGI 800 (Taylor Hobson) was used to measure 2D surface roughness and to make 3D measurements of surfaces. The equipment is traceable to national standards and has a stylus tip of 2 µm. The 2D roughness measurements of the contacting bodies were evaluated in terms of the following parameters, defined in Thomas [82]: mean peak to valley roughness (Rz), average roughness (Ra), root-mean-square roughness (Rq), skewness (Rsk), kurtosis (Rku), root-mean-square slope roughness (Rdq), the autocorrelation length (a texture parameter). The 3D surface measurements using the same stylus device were evaluated by the following parameters: average amplitude (Sa), distance from mean plane and deepest valley (Sv), and the distance from the highest peak to the deepest valley (St). In addition to the Form Talysurf measurements device, an atomic force microscopy (AFM) mounted on a light optical microscope was used to make the 2D measurements in Paper E. AFM has high vertical resolution, but the drawback is the poor ability to measure tall surface structures, since standard scanners generally have less than 5-10 µm of vertical range [61]. Therefore, only smooth surfaces with low surface roughness can be analysed. With AFM the surface is mechanically examined using a very sensitive tip radius of 20...
nm mounted on the end of a flexible cantilever. The tip motion is conducted by a piezoelectric scanner and while scanning over the surface, small forces are recorded between the outermost atoms of the tip and the surface atoms, tip-sample interaction is monitored by reflecting a laser beam off the back of the cantilever [61].

4.3 GD-OES

In order to obtain information about the composition of chemically reacted surface layers, and find out whether wear and friction could be linked to these layers, the specimens used in Papers D and E were analysed with glow discharge-optical emission spectroscopy (GD-OES), a technique recognised within the international standards [84]. GD-OES can detect chemically reacted layers that change the surface composition by combining sputtering and atomic emission techniques for element depth profiling.

GD-OES is based on the use of a glow discharge device as an optical emission source. The glow discharge device consists of a vacuum chamber filled with an inert gas, usually argon. The glowing plasma is generated in the analysis chamber by a voltage of 500–1000 V applied between the anode and cathode, where the solid sample surface forms the cathode. Material is continually removed from the sample area by ionised argon atoms at a rate of 1 μm/min in steel (so-called ion sputtering). The surface atoms ejected into the plasma are then excited in an electrical field. The majority of these excited atoms emit characteristic optical emission upon relaxing into the lower electronic state. The light, which is characteristic to the emitting element, is detected and analysed using a conventional optical emission spectrometer, and the elemental depth profile is obtained from the emission intensities as a function of time. The system is presented in Figure 15. The advantages and limitations of the method are discussed by Dizdar [85].

4.4 Numerical contact modelling

A numerical model for contact analysis was used in three of the appended papers. The assumptions used for a numerical model are basically the same as for analytical solutions, except that the method is based on replacing the continuous pressure distribution with a discrete set of pressure elements (Figure 18). In Paper B, contact computation was
applied to normally loaded contacts along with nominally smooth surfaces. In Paper C, real (or measured) surfaces of differently manufactured tooth flanks were used as input. The output indicates how asperities may induce local high pressure spots that may act as stress raisers, which in turn may influence friction behaviour and early wear. In Paper E, tangential loading and the effect of rolling/sliding were included in the model, but here the surfaces were smooth or created numerically to a simplified directional surface roughness or texture pattern. However, the calculated friction coefficient is most accurate for smooth surfaces since the transient motion is not reflected in the simulation.

![Figure 18. Schematic contact between a rough surface and a smooth sphere [37] with each real contact area divided into a mesh of cells.](image)

The contact program was developed by Björklund and Andersson [33], but for Paper E the program was further developed by Zhu [34] to include the rolling/sliding component. Similar numerical methods are discussed by Johnson [60].

### 4.4.1 Normal loading

Assuming that the contact is frictionless, the tangential traction can be omitted and only traction components in the normal direction are taken into account. Consider a contact that has been divided into a mesh of rectangular cells (Figure 18) where each cell is subjected to a uniform pressure and the magnitudes of these pressures are the unknowns of the problem. There exists a unique set of pressures that exactly matches the boundary conditions at the centre of the cells. The solution is obtained from an equation system that written in matrix form becomes:

\[
C_{zz} \, p = \delta_z - h
\]

where \( C_{zz} \) is the influence coefficient matrix for normal loading, \( p \) the unknown pressures, \( \delta_z \) the applied normal displacement, and \( h \) the gaps between the cells before deformation. The influence coefficients for a uniform pressure on a rectangular cell were found by Love [86]. The equation is solved by iteratively varying \( \delta_z \) in order to find the known applied normal force. As the real contact area is not known in advance, it is necessary to start with estimated contact regions. These estimates, which must contain the true regions, are obtained as the region one would get if the bodies were penetrating each other without any interactions (Figure 19). Solving Eq. (1) will therefore result in some pressures having negative values, indicating that these elements are outside the contact region. These elements are removed, and Eq. (1) is solved again until all pressures become positive.
4.4.2 Rolling/sliding

When a tangential load is applied to a normally loaded contact, sliding will not start simultaneously over the whole contact area. Some contacting regions will stick and some slip before pure sliding occurs. The method for solving tangential loading is in many ways comparable with the method for normal loading. In addition to the unknown pressures and the applied normal displacements, there are now unknown tangential tractions and applied tangential displacements. When adding rolling/sliding to the model, the displacements are no longer constant for the whole contact area, but depend on the relative slip and the contact length in the rolling direction. In modelling a rolling/sliding contact, the contact starts with a sticking area and slip occurs when elastic deformation cannot support the relative motion of the two bodies. With increasing tangential displacement, the slip areas increase and sticking areas decrease, leading to more tangential force being transmitted. The relationship between micro-slip velocity and the strains in the surface has been thoroughly discussed by Johnson [60]. If one observes the contact under the assumption of steady rolling without spin, the integration of an equation found by Johnson expresses the tangential deformation \( u \) as:

\[
 u = \xi x - s + c_0 
\]  

(2)

where \( \xi \) is the relative slip, \( x \) is the distance to the leading edge, \( s \) is the slip distance, and \( c_0 \) is a constant generated from the integration. On the discretised grids, the general equation Eq. (2) is rewritten in matrix form:

\[
 C^a q = \xi x - s + c_0 
\]  

(3)

where \( q \) is the traction and \( C^a \) is the influence coefficient matrix that governs the relationship between the tractions and deformations. The traction on each element \( q_{i,j} \) at input pressure \( p_{i,j} \) is limited by the limiting friction coefficient \( \mu_i \) of that element:

\[
 q_{i,j} \leq \mu_i p_{i,j} 
\]  

(4)
The slip distance $s$ is not actually calculated in the solution, since the solution starts with the assumption that there is no slip in the contact area. Once Eq. (3) is solved, the tangential stresses near the trailing edge will violate the limiting friction condition, Eq. (4). These violating elements are then moved to the right-hand side of the equation, since the tangential traction is known and they still affect the remaining elements. Eq. (4) is solved repeatedly until all remaining elements are sticking. Finally, the equation to calculate the friction coefficient $\mu$ (the ratio of total tangential force and normal load) is given by:

$$
\mu = \frac{F_t}{F_N} = \frac{\sum q_{ij} dA_r}{\sum F_N} = \frac{\sum q_{ij} A_r}{F_N}
$$

where $F_t$ is the traction (i.e. friction) force, $F_N$ is the normal force, and $A_r$ is the area of the contacting cells.
5 Summary of appended papers

This thesis comprises six papers (Appendices A-F) relating to phenomena occurring in the gear contact. A summary of each paper is given below.

Paper A can be seen as a literature survey of the current international standard gear rating procedures. This survey reveals that the standard is not useful in finding a robust gear design. Generally, the ISO standard calculations are time consuming, and it can be difficult and sometimes confusing to get a grip on the factors, assumptions and advice given in footnotes. The results in Paper A prompt the recommendation that these quality levels should mainly be used for guidance and for communication purposes only. Additionally, variations in surface roughness and lubricant design are restricted to standard reference test gears and mineral oils respectively.

Paper B covers potential lead modifications to allow for non-perfect gears in contact. The logarithmic profile was found to be less sensitive to small misalignments, which is of interest in terms of achieving a robust design. The logarithmical profile must be developed further in order to allow competitive manufacturing costs.

In paper C, differently produced gears were compared via surface measurements and contact simulations. The real contact area ratio was successfully used for describing the gear contacts. It was also proposed that the real contact area ratio could be used as an advanced roughness parameter. Honing and green-shaving were found to be preferable to grinding in terms of contact area ratio. The ground surface was reproduced on specimens used for tribotests in Papers E and F.

The results from Paper D showed that the chemically reacted surface boundary layers played an important role in terms of wear. More specifically, the oxide layer thickness had a significant influence on wear. The findings also demonstrate the complexity of lubrication design formulations coupled to these layers. For example, it was found that the pre-existing surface boundary layer (before any lubricant has been added) plays an important role in allowing the lubricant to react properly with the surfaces.

Paper E used twin-disc testing to demonstrate the impact of surface micro geometry and surface finish on friction, wear, and the chemically reacted surface layer. Micropitting was found on all ground discs, for both lubricants. A pre-existing surface boundary layer formed by one sulphur AW additive did not prevent micropitting initiation. The thick sulphur layer formed by the preheating seemed to serve as an efficient sacrificial layer, especially when combined with the commercial lubricant, since by the end of testing no trace of the pre-formed sulphur layer remained. Although both lubricants resulted in micropitting, differences in friction and wear were found. The ester-based environmentally adapted lubricant (EAL) showed thicker reacted surface boundary layers than the PAO-based commercial lubricant. The thicker layer did not reduce friction, suggesting the boundary layer itself needs to be sheared off, but it did seem to confer less wear. This relationship was also reflected in Paper D, where different AW additives were added to a complex ester base fluid. Superfinishing of the discs greatly increased the
lambda ratio, and almost no wear was detected; the discs were most likely separated by an EHD film.

In paper F, a barrel-on-disc machine was used to study the impact of grinding and superfinishing. The choice of manufacturing method significantly influenced the friction coefficient. Measurement results showed that the change of lubricant had an impact on friction in the mixed to boundary lubrication regimes similar to that of the change of main surface orientation. Measurements and simulations showed that the barrel-on-disc and twin-disc setups reflected the same friction trends. However, the friction coefficient using the barrel-on-disc setup was almost twice as large as that found using the twin-disc machine. The difference in contact geometry is believed to be the main reason for the higher friction level in the barrel-on-disc machine.
6 Discussion and conclusions

The main objective of this thesis is to increase the understanding of how surface topography and lubricant interact in the mixed to boundary lubrication regimes. The results in this thesis linked to lubricant properties mainly contribute to the early product development process. Changing a gearbox lubricant requires extensive assessments, since the bearings and synchronizers are also parts that need to be considered. However, surface topography changes on the gear flank are more easily implemented as part of the continuous improvement process for an already mature gear system.

6.1 To what extent is surface topography handled in today’s gear design tools?
Wavelengths detected by standard rating gear parameters do not include the inherent manufacturing fingerprint; that is, the surface roughness and texture orientation of the gear flank surface. It is clear from Paper A that the standard quality levels, which make use of six individual flank deviations, are mainly useful for guidance and for communication purposes, and are not linked to a specific performance requirement. Today, commercial gear design software based on the finite element technique can help in investigating the influence of these parameters on aspects such as contact stress and transmission error, which is useful for comparing and analysing the effect of potential tooth modifications. The introduction of a logarithmical profile suggested in Paper B, and how it interacts with other tooth modifications, should be validated in such way. Nevertheless, commercial gear design tools do not incorporate the effects of rough surfaces. In Paper C, the contact area ratio (ratio between real and nominal contact area) is used to compare four differently manufactured gear surfaces in contact. The findings show that two gears finished by honing or grinding can be classified as equal by standard accuracy levels despite significant differences in surface topography characteristics. As inspection machines improve, a number of gear parameters describing the individual flank deviations can be added, which will drive creativity for more advanced gear design tools.

6.2 What is the impact of surface topography on load capacity?
Hertzian pressure is the basic principle for the assessment of the surface durability of cylindrical gears, and is a significant indicator of the stress generated during tooth flank engagement [8]. However, it is not the only cause of surface degradation. For example, the coefficient of friction, the amount and direction of sliding, and the lubricant also influence pressure distribution. As described in Paper A, these effects are empirically incorporated in the standard calculation procedure by factors and the choice of material property, but the span of these factors is limited to reference gear tests.

One direct way of increasing the load capacity is to make sure that the maximum width of the contact patch is used, based on established peak misalignments. This involves the use of robust lead (or crowning) modifications that prevent excessive pressure peaks from occurring at the ends of the flank widths (i.e. edge loading). In Paper B, traditional
lead profile modifications are compared with a logarithmical lead profile. The logarithmical profile was first developed for roller bearings by Lundberg [87], but there is not much evidence for its use in gears. The results from Paper B are used to conclude that the logarithmical profile is of interest in terms of achieving a robust design and in order to reduce the maximum contact pressure. Both of these can contribute to improvements in gear performance, which today lies on the border of what is possible. Implementing such a lead profile might interfere with the manufacturing and inspection process. Although the manufacturing process is different for gears and bearings, the work of Fujiwara and Kawase [88] gives a thorough insight into how to implement a logarithmic crowning equation for bearings with design freedom and allowance for a flat region from the viewpoint of machining.

The variation in surface topography inherent in the manufacturing method was found to be an important factor in Paper C, at least for early life contact conditions. A low real contact area means that the real contact pressure can be significantly higher than the maximum Hertzian contact pressure. Thus surface stresses can locally be higher than the allowable contact stress.

6.3 Can the influence of surface topography and lubricant interaction be measured?

Three model tests were used in Papers D-F to evaluate the coefficient of friction and wear during mixed and boundary lubrication conditions and hence reveal information about the contact performance in gears. The use of GD-OES to investigate changes in elemental composition (Papers D and E) made it possible to determine the lubricant performance.

The surface analysis in Paper D revealed that the formation and composition of surface boundary layers depends strongly on the chemical composition of the lubricant, but also on contaminations in the as-manufactured stage. With this knowledge and in order to achieve similar positive effects as for manganese-phosphate gears [15], ground specimens for twin-disc tests were pre-treated with a sulphur additive intended to build an easily sheared surface boundary layer. The twin-disc tests setup showed that this pre-formed surface layer equalized the frictional behaviour differences between the two lubricants, but not the amount of wear. Thus this needs further investigations. Moreover, GD-OES analysis was able to distinguish the two lubricants and the two different lubricant regimes that the ground and polished surfaces were operating in. The environmentally adapted lubricant (EAL) in Papers E and F showed an increase in friction compared to the commercial heavy truck lubricant. From the surface analysis it can be concluded that the thicker reacted surface layer for the EAL might be a contributing factor. On the other hand, in Paper D and E, this layer seems to confer less wear. For the polished surfaces, promoting less asperity contact, the higher friction for the EAL is probably linked to its higher viscosity. However, physisorbed or weak chemisorbed layers cannot be detected by use of GD-OES, and GD-OES cannot distinguish whether AW or EP additives are activated (or occupied) on the surface. The ideal situation would be to use a combination of a number of analytical instruments that can help to explain friction and wear by revealing what happens in the surface boundary layers. However, GD-OES combined
with Stribeck (or traction) curves, as obtained with the barrel-on-disc set-up in Paper F, is a powerful tool when testing lubricant performance linked to contact behaviours.

An additional test with both experimental and computational components was added in Paper F to evaluate the effects of a change in main surface texture direction; that is, longitudinal (circumferential ground discs) and transverse (radially ground discs) roughness textures. However, the measurement results showed that the change of lubricant had a similar impact on the friction level as the change of main surface orientation. Ideally, grinding should be performed in the opposite direction to the rolling direction, as in real ground gears, but as long as the tests are performed in the mixed to boundary lubrication regime this was found to be of no essential importance in terms of friction and wear when run in a barrel-on-disc set-up. Likewise, the influence of the slide to roll ratio on friction coefficient is of little importance in mixed and boundary regimes according to Brandão et al. [89], who made use of the same test setup using a standard ball instead of a barrel.

Twin-disc tests using ground disc surfaces all resulted in micropitting that was not visible to the naked eye. The fact that this type of surface damage is generated is a positive feature which speaks in favour of this test setup, since it is a prevalent surface failure and is not fully understood. Furthermore, the mean friction coefficient in the twin-disc contact area was comparable to other studies [53, 80]. The mean friction coefficient using the barrel-on-disc set-up was almost twice as large as that of the twin-disc set-up. The simulation results in Paper F show that the difference in contact geometry is the main reason for this behaviour. However, the friction trends in twin-disc and barrel-on-disc set-ups were in good agreement. Due to the small wear track on the disc in the barrel-on-disc setup, wear studies using this setup will require more advanced surface analysis instruments, such as SEM or optical measurement devices.

Experimental testing should preferably be used in combination with computer simulations, since it can enhance and contribute to the understanding of the problem. The contact area ratio parameter, which for input data requires only the surface topography, the load, and the elastic properties of the materials, is suitable for practical analysis of surface features in industrial applications. It can serve as a complement to the existing roughness parameters in predicting the risk for severe local asperity collisions that give rise to high local tensile stresses leading to surface-initiated cracks (micropitting). The surface-initiated pits found on the discs used in the twin-disc setup were most likely caused by the high local roughness peaks of the surface, rather than the lubricant itself.

6.4 Can a reduced surface roughness reduce friction and wear, and how will this influence the lubricant design?

The experiments in Papers E and F showed that superfinishing of surfaces reduced the friction forces by approximately 30% compared to the ground surfaces studied. A smoother surface may not only help lower the contact load dependent friction losses, but may also reduce the risk for early fatigue wear or early wear-in damages that most likely reduce gear life. In Paper E, early micropitting was discovered on the ground disc run in
a twin-disc setup. Micropitting tends to occur in rolling/sliding case hardened contacts, particularly when ground gear teeth are used [90]. The ground disc specimens for twin-disc testing were manufactured to imitate the ground gears used in Paper C, which resulted in the lowest contact area ratio compared to the honed and green-shaved surface. While Paper C deals with the initial contact conditions, Sjöberg et al. [7] continued the work by studying the changes during running-in. They concluded that ground gears showed a surprisingly small change in contact area ratio, hence a lower running-in effect. This leads to the assumption that running-in for a short period of time might not have prevented micropitting initiation. The ground discs in Paper F suffered from characteristic surface-initiated micropits with few mild scratches combined with a smoothing of the surfaces. Thus, a micropitted surface might indicate a higher contact area ratio than as-manufactured, and in terms of predicting surface degradation the suggested contact area ratio parameter is probably most suitable for as-manufactured surfaces. Cracks initiated on the surface can be explained by the fact that the surface asperities act as stress raisers and the contact pressure at these local spots is beyond the allowable stress, which the contact area ratio gives a strong indication of. A higher contact area ratio from start (as-manufactured) would most likely have prevented this type of surface failure. A polished disc run in the same device, though exposed to a lower contact pressure, had only a negligible amount of wear. The well-established work of Krantz et al. [42] shows that there is strong evidence that superfinishing significantly improves the surface fatigue life of case carburized, ground, aerospace-quality AISI 9310 gears.

Brechot et al. [76] have shown that additives can promote micropitting; however, EP additives are needed for protection from scuffing and most likely AW additives are needed for protection from mild wear. According to Paper E, the additives seem to suppress the gradual smoothing or running-in on the ground specimens – in other words, the additives have done their job. One interesting question is if a combination of FM (which mainly helps to reduce friction and is not primarily for wear protection) and EP additives may facilitate the running-in. Most desirable of all would be the ability to control when different additives are activated on the surface. The existence of micropitting, which can be described as a form of running-in, is difficult to prove, since gears are seldom disassembled when they are performing properly. Snidle et al. [72] simulated the micro EHD lubrication regime to show that when surface roughness is present, the maximum sub-surface shear stress field occurs much closer to the surface. Thus the micropitting can be suggested to be a balance of two failure modes – mild wear and surface initiated fatigue. The micropit size did not differ between the two lubricants when run with the non-preheated ground surfaces. Thus the pit size cannot be linked to either the difference in lubricant chemistry or the bulk properties of the two lubricants.

Today, we are faced with a broad range of environmentally adapted lubricants based on both natural and synthetic esters. In order to optimise the benefits of an EAL, a better understanding of the additive and base fluid interaction is desirable. In Paper D, the ester base was found to be highly polar, thus outcompeting the traditional additive’s ability to promote a boundary layer. In Papers E and F, the EAL was formulated differently, as a fully saturated mixture of esters, giving the additives a chance to react with the surface.
Designing a gearbox lubricant is a complex task. In the early design phase, it is necessary to review which additives are needed, when they are needed, and why they are needed. Knowledge of how lubricants interact with the contacting surfaces will hopefully lead to more detailed requirement specifications.
7 Future work

Future work should involve investigating how much impact the existence of micropitting has on the service life and gear efficiency, since this is still not fully understood. Further on, how surface initiated fatigue wear depends on mild wear also needs more investigations. These two mechanisms seem to coincide from the very instant the surfaces come into contact. Twin-disc testing with test specimens as close to real gears before using appropriate transmission tests, for example pitting tests in a FZG test rig, would be of great use. In particular, if Striebeck (or traction) curves could be generated from the twin-disc machine in order to set operating conditions to run in the mixed to full-film elastohydrodynamic lubrication regime.

One challenge would be to include the manufacturing fingerprint, standard form deviations and surface intended modifications in a model, effectively combining boundary element and finite element solutions to determine their consequence on gear performance (e.g. transmission error and maximum contact pressure). This could contribute to the understanding of whether superfinishing and coating of gears needs to be followed by a lower ISO standard quality level in order to have a significant effect in increasing the load capacity.

Running-in of gears and its impact on gear life needs more investigations. One interesting question is if running-in can make gears less sensitive to manufacturing variations linked to standard rating accuracy levels, which may be more cost effective than increasing the gear flank accuracy. Running-in might have to be performed with a specially formulated lubricant, since some additives may suppress the gradual smoothing or running-in of the surfaces.

The lubricant properties linked to friction and film build up in an elastohydrodynamic lubricated contact are of interest in minimizing the energy losses in transmissions. The importance of surface boundary layers must be investigated further, as these layers affect both wear and friction. Surface analysis for studying the outermost atom layer corresponding to a physisorbed layer may better describe more aspects on the formations and failure of surface boundary layers.
8 References


2. Gear Industry Vision: A vision of gear industry in 2025, developed by the gear community, Gear Industry Vision Workshop, 10 March 2004, Detroit, Michigan; Prepared by Energetics, Inc., 2004


23. Li S., Finite element analysis for contact strength and bending strength of a pair of spur gears with machining errors, assembly errors and tooth modifications, Mechanism and Machine Theory, 42 (2007) 88–114

24. Li S., Effects of machining errors, assembly errors and tooth modifications on loading capacity, load-sharing ratio and transmission error of a pair of spur gears, Mechanism and Machine Theory, 42 (2007) 698–726


47. Xiao L., Gear tribology: friction and surface topography, Doctoral Thesis, Chalmers University of Technology, Göteborg, Sweden, 2005


55. Van Beek A., Advanced engineering design, lifetime performance and reliability, Delft University of Technology, Delft, 2009

56. Stribeck R., Die wesentlichen eigenschaften der gleit- und rollenlager, Zeitschrift des Vereins Deutscher Ingenieure, 46 (1902) 1341-1348 1432-1438 1463-1470


64. Godfrey D., In: P.M. Ku (ed.) Interdisciplinary approach to friction and wear, Southwest Institute, Washington DC, 1968


73. Wind turbine micropitting workshop: A recap, Sheng S. (Editor), Technical report NREL/TP-500-46572, 2010

74. Martins R., Locatelli C., Seabra J., Evolution of tooth flank roughness during gear micropitting tests, Industrial Lubrication and Tribology, 63 (2011) 34–45


86. Love A.E.H., Stress produced in a semi-infinite solid by pressure on part of the boundary, Philos Trans R Soc London A., 228 (1929) 377–420

87. Lundberg G., Elastic contact between two semi-infinite bodies, Forschung auf dem Gebiete des Ingenieurwesens, 10 (1939) 201–211

