Politecnico di Milano
School of Industrial and Information Engineering
Master of Science in Materials Engineering & Nanotechnology and
Mechanical Engineering

POLITECNICO
MILANO 1863

A study on crack initiation for rolling contact fatigue in planet gears

Advisor: Prof. Andrea MANES

M.Sc.Thesis by:
Paolo CATTANEO
ID number 872285

Academic Year 2019–2020

Acknowledgments

Voglio ringraziare il professor Andrea Manes per avermi dato la possibilità di lavorare a questo progetto di tesi e per essersi dimostrato sempre una guida professionale. Una speciale menzione va anche a Massimo Fossati e Mohammad Razasefat Balasbaneh, i quali con pazienza, disponibilità e competenza mi hanno seguito e consigliato durante tutta la durata di questo lavoro. È d'obbligo ringraziare i miei genitori e la mia famiglia, che mi hanno sempre supportato e che hanno creduto in me per tutti questi anni. Ringrazio anche gli amici di sempre, le persone con cui ho condiviso quest'esperienza e tutti quelli che mi sono rimasti vicini.

I want to thank Professor Andrea Manes allowing me to work on this thesis project and for demonstrating always to be a professional guide. A special thought goes to Massimo Fossati and Mohammad Razasefat Balasbaneh, who with patience, willingness and expertise, followed and advised me during all the duration of this work. Obviously, I must thank my parents and my family, which had always supported and believed in me during all these years. I shall thank also my dear friends, the people with whom I shared this journey and everyone who cared about me.

Contents

List of Figures

List of Tables

Abstract

Rolling elements such as bearing gears are widely used in mechanical applications in industrial, automotive, aerospace and many other fields. Research and design development are focusing on increasing the components reliability and the materials' properties to guarantee a long service life and the performance stability. Despite all the efforts, bearings are still failing, therefore is interesting to understand the mechanisms governing this phenomenon, called Rolling Contact Fatigue (RCF). Numerical and FEM models have been developed. which can simulate the fatigue behavior of those materials starting from theoretical considerations, validated with experimental evidences. A general overview on the state-of-the-art studies on RCF has been illustrated in this thesis by showing the factors that take part in this process. Plus,mechanisms driving crack formation and growth have been analyzed and compared in order to focus on the subsurface crack initiation and propagation in case of lubricated rolling contact conditions, which are proper of rolling elements and gears. Moreover, an analytical-numerical model simulating the damage evolution inside the material is proposed, by utilizing the Continuum Damage Mechanics (CDM) method to predict the fatigue initiation life of a rolling element. Some widely used bearing gears steels, such as AISI 52100 and SAE 4340, have been taken as references for this simulation. In the end, this model is joined by a Finite Elements Model (FEM) which simulates the mechanical behavior of an epicyclic gearbox system, with a specific focus on the section that includes the planet bearing gear subjected to the RCF phenomenon. The stress state resulting from torque and motion transmission has then been compared to literature references and to the analytical-numerical results, to give a first approximation of the component's general fatigue initiation behavior.

Keywords : Rolling contact fatigue, continuum damage mechanics, bearing gears, FEM, analytical-numerical analysis

Estratto in lingua italiana

I componenti meccanici che sfruttano la presenza di elementi volventi, come gli ingranaggi a cuscinetti, trovano un'ampio utilizzo nel settore industriale, automobilistico, aerospaziale ed in molti altri campi. La ricerca e lo sviluppo di sono quindi focalizzati nel migliorare l'affidabilità e le proprietà dei materiali utilizzati, per garantire un lunga durata in servizio ed una stabilità di rendimento di questi componenti. Nonostante l'ampio sforzo, si verificano tutt'oggi dei casi di fallimento principalmente causati dalla fatica per contatto da rotolamento (RCF). Modelli analitico-numerici ed agli elementi finiti (FEM) sono stati sviluppati per simulare il comportamento a fatice dei materiali utilizzati, partendo da considerazioni teoriche, validate successivamente da prove sperimentali. In questo lavoro di tesi viene data una panoramica generale sullo stato dell'arte riguardante lo studio della fatica da rotolamento, illustrando i fattori che prendono parte in questo complesso processo. I principali meccanismi responsabili della formazione e propagazione delle cricche sono stati analizzati e comparati per fornire una descrizione specifica della formazione di cricche al di sotto della superficie in condizioni di contatto da rotolamento in presenza di lubrificanti. Un modello analitico-numerico, in grado di simulare l'evoluzione dello stato di dannno all'interno del materiale, sarà proposto in questo lavoro servendosi della teoria della meccanica di danno continuo (CDM), per predirre l'iniziazione del danno da fatica da rotolamento. I materiali studiati saranno delle leghe largammente utilizzate per la realizzazione di ingranaggi e cuscinetti (AISI 52100 e SAE 4340), spesso indicati come riferimento in letteratura. Nella seconda parte di questo lavoro infine, sarà introdotto un modello agli elementi finiti in cui è stato studiato il comportamento della sezione di una trasmissione planetaria, con interesse particolare sulla sezione di contatto da rotolamento tra cuscinetti ed ingranaggio planetario. Lo stato di sforzo risultante dalla applicazione di coppia e rotazione trasmesse attraverso il sistema, sarà paragonato con la letteratura ed i parametri di sforzo ottenuti dal modello analitico-numerico, per fornire una prima approssimazione del comportamento a fatica dell'ingranaggio planetario sotto esame.

Parole chiave : Fatica da rotolamento, meccanica del danno, ingranaggi a cuscinetto, analisi agli elementi finiti, analisi analitico-numerica

1 Introduction

Fatigue takes relevant technological importance as a failure phenomenon of many engineering applications. It is characterized by the process of damage accumulation inside the material structure of mechanical components subjected to cyclic loading.

This repeated contact produces the formation of micro-cracks or voids in the area of maximum stress concentration. The inner crystalline structure or the presence of micro-defects, as inclusions or imperfections, creating a discontinuity in the lattice, can act as a stress concentration enhancer. In moving and rotating pieces of machinery such as cam-followers, transmissions, rail-wheel, gears, and many others; Rolling Element Bearings (REBs) have a fundamental role in transmitting rotary motion while undergoing relevant amounts of load, reducing friction during motion. Loading conditions are generated by the relative motion and contact between rolling parts (rollers) and the raceway of the gear, thus producing alternating stress inside the material in the region underlying the contact.

A correct design of REBs (Fig. 1.1) implies proper loading distribution and suitable lubricating conditions that will lead to the reduction of wear, leaving fatigue as the main cause of working-life critical limit [1], known as rolling contact fatigue (RCF).

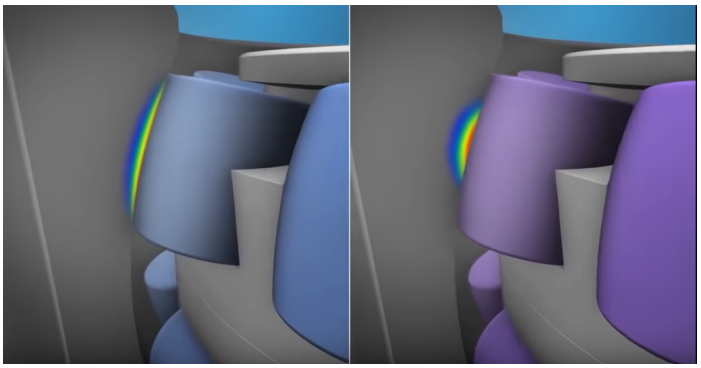


Figure 1.1: *Representation of rolling contact inside a bearing gear [2]*

RCF may lead to the failure of rolling elements in different modes, differentiating by initiation or by the dimension of inherent defect generated.

The two most dominant mechanisms are: pitting generated by surface cracks and spalling initiated by subsurface cracks. These two failure modes differentiate from one another not only by dimension, which is greater for spalling, as it can be seen in Fig. 1.2, but mainly by the initiation mechanisms and spot. Pitting usually occurs at the surface in correspondence of defects as dents or asperities, whereas spalling is generated by micro-cracks formed beneath the surface at inhomogeneities such as inclusions or any other stress raiser. In both cases the cracks initially grow inside the material and then undergo branching, propagating toward the surface and forming the defect with consequent material loss. Factors both from loading and operational conditions, favoring a mechanism or another, will be discussed later but it should be stated that in case of smooth and clean lubricated contact, it is spalling to be the dominant mechanism [1][3].

In this case, initial damage occurs in the subsurface region where the maximum shear is located. Hertzian contact loading history [5] demonstrates that in this zone the

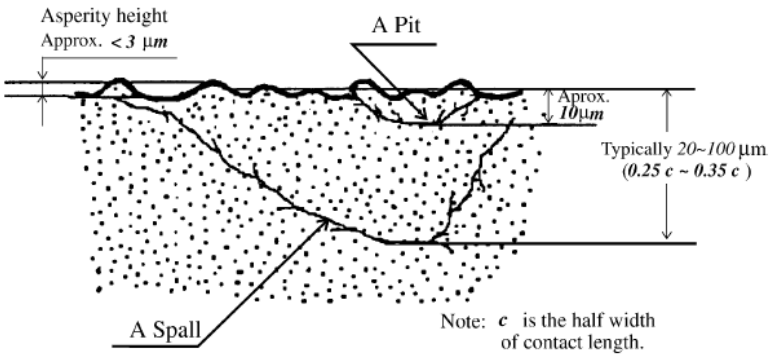


Figure 1.2: *Schematic representation of spalling and pitting [4]*

shear stress undergoes a complete reversal during each rolling contact pass, indicating that spalling should be a shear driven phenomenon.

Investigation of Hertzian load conditions in RCF will require consideration of many factors regarding material properties and treatments, surface quality, metallurgical composition, and numerous working conditions such as lubrication. The combination of all these different variables results in a complex multi-axial stress distribution in the contact region [5][6][7]. We will see in the state of the art, different studies proposing theories on the different stresses that can act as a dominant mechanism for fatigue crack initiation and propagation. Taking reference from this theoretical background, many models have been proposed to predict and simulate the RCF behavior of material and components under different conditions. The most famous are the models proposed by Lundberg and Palmgren [8] and the one by Ioannides and Harris [9], which have been used as the basis for following developments. Initially, models using Weibull probability distribution function [10][11], tried to describe fatigue life of rolling contact components in an empirical way but a precise understanding of the inner physical mechanisms, driving the process, was still missing. This deficiency was then faced using numerical models [12], which can describe with a good degree of reliability initiation and propagation stages of fatigue, but inherently lack the stochastic nature proper of fatigue.

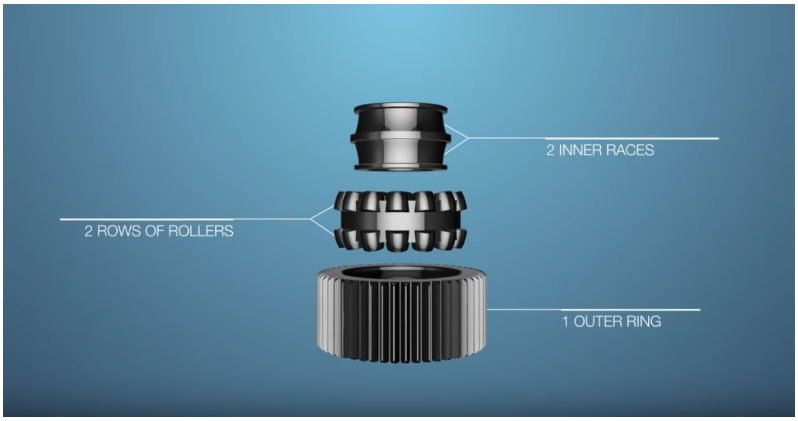
Recently a new simulation method was introduced to overcome this limitation of numerical simulation methods by incorporating microstructural features in the model. For example, implementation of Voronoi Tessellation [13], which divides the simulated domain into randomly generated grains. Following this approach it is possible to replicate the metallurgical structure of the material and then induce randomness in the analysis, resulting in the characteristic scatter of fatigue. This method requires a higher model's complexity and more heavy calculations.

Moreover, when investigating the initiation and following stages of fatigue, cumulative analysis of damage evolution can be very interesting in the prediction of fatigue life but also to describe how the properties of the materials and growth rate evolve during the phenomenon [14][15]. This can be simulated by considering CDM, which is capable to evaluate the degree of damage inside the material at each position at any given time (cycle). Once given the loading conditions, and then considering its effect on the material throughout the simulation, it is possible to overcome the limitations

imposed by linear rules proposed in theory such as Miner's rule.

This study has taken its cue from two similar incidents that occurred on the same helicopter model, apparently for the same fatigue causes.

(a)



(b)

Figure 1.3: *3D representation of (a) an epicyclic gearbox and (b) the bearing gear under study [2]*

On 29 April 2016, a Eurocopter (the European Airbus helicopter branch) EC225 Super Puma helicopter operated by CHC Helikopter Service, carrying oil workers from the Gullfaks platform in the North Sea, crashed near Turøy, a Norwegian coastal island 36 kilometers from the city of Bergen. Several witnesses reported initially seeing nothing untoward with the helicopter flying steadily before the sound suddenly changed and it started to sway. A moment later its main rotor assembly detached from the aircraft and the fuselage plummeted to the ground and exploded into flames. All 13 crew and passengers died in the crash. It required over 2000 hours of testing, forensic investigation, and numerical modeling to simulate and analyze the failure.

Also, the comparison with data collected in 2009 in another similar accident has lead to the identification of the sequence of events that caused the incident. The results indicated that the second stage of the gearbox is the critical point of failure

that led the rotor to separate from the cabin. These events suggested the necessity to improve the understanding of the RCF behavior in this component.

In the Airbus case, it is shown how the main gearbox transmits energy reducing rotational speed to the second stage. This is an epicyclic gear train composed of: a sun gear, eight planet gears, a planet carrier, and a fixed ring. Each Airbus' planet gear (Fig. 1.3b) comprises an inner race on which two rows of rollers act.

It is on this race that wear has begun, initially by a subsurface shallow crack. Then propagation occurred, associated with the classical spalling phenomenon (Fig. 1.4a). Differently to usual wear, it unexpectedly gave way to a core-crack which then grew and propagated through all the component splitting it in half (Fig. 1.4b-c). The displacement due to the crack opening caused a collision between the teeth of the components and the teeth of the sun gear, leading the planet gear to break instantaneously (Fig. 1.4d). Pulled by its inertia, the main rotor sheared off the top section of the gearbox, separating from the structure and causing the helicopter to fall.

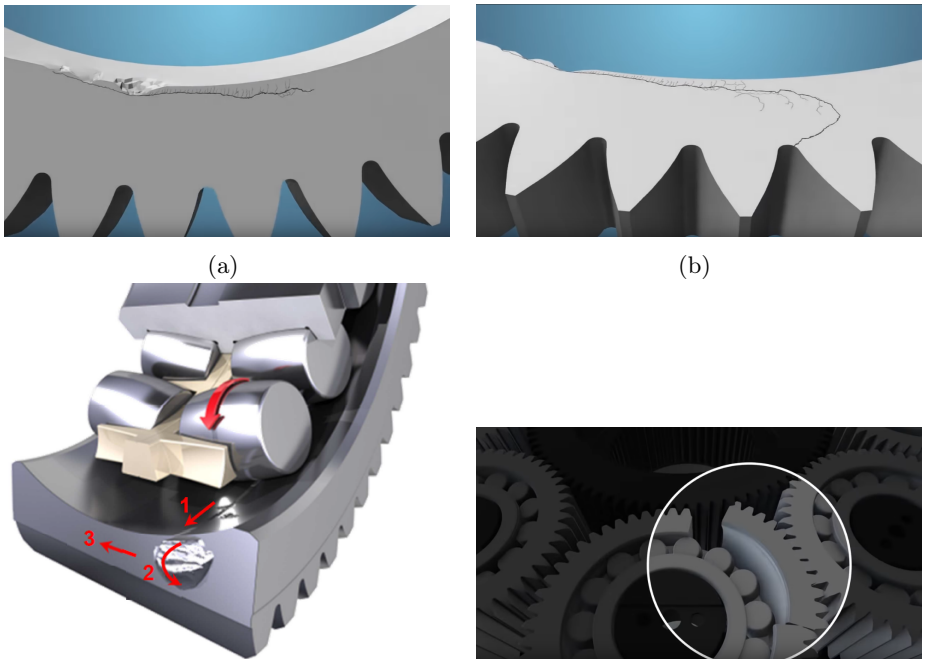


Figure 1.4: *3D representation of: (a) initial crack propagation and spalling, (b) crack penetration through all section leading to failure (c) section-view of crack propagation, (d) resulting failure on gear [2][7]*

As a consequence of the studies performed by the manufacturer, improvements on safety measures to prevent the gearbox from shocks during transport and assembly, and a more accurate micro-spalling detection system has been implemented. But it is remarkable to notice that, the maximum operating life of the planet gear component, previously defined in 2004 following the certification standards, it has been reduced by a factor of 4 after the inquiry in 2017 (Fig. 1.5). This gives a clear indication of the fact that failure should have been caused by a fatigue problem [2].

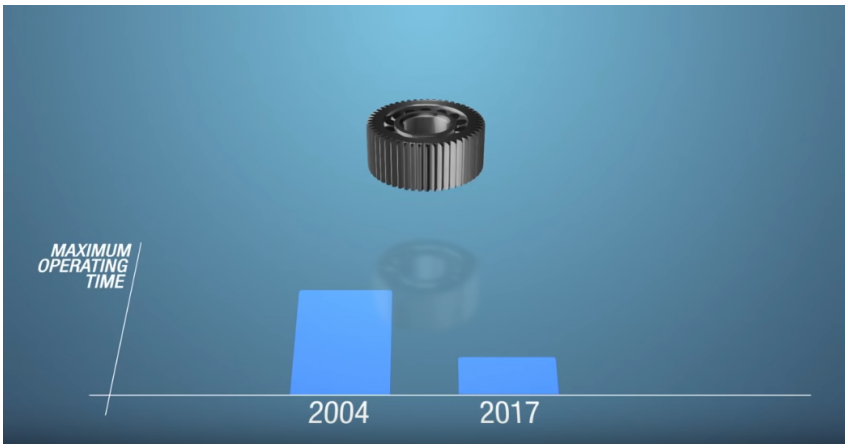


Figure 1.5: *Fatigue life evaluation of bearing gears from the epicyclic stage of the EC225 Super Puma helicopter, before and after the study by Airbus after incidents [2]*

To study the behavior of bearing gears under RCF conditions is therefore of capital importance. It will be given a general definition of all the factors taking part in the stress state. The stages and the drivers of the failure process, from crack formation to propagation, will be identified in order to have a clear idea of the mechanisms taking part in the phenomenon. Finally, a model able to replicate the RCF behavior will be developed. Which can be reliable and based on valid theoretical assumptions coming from the observations on the aforementioned case.

The current work aims to define a methodological design approach, able to consider all the mechanical and material's parameters contributing to the fatigue behavior, in order to avoid fatal incidents in the future.

In this thesis work, two models have been developed to identify the stresses acting in this specific bearing gears system and study the fatigue initiation life of the planet gear component.

Starting from the definition of the contact parameters, such as the maximum pressure and friction, the multi-axial stress tensor has been evaluated by following the generic integral equations given by the Hertzian theory [5].

Once the principal stress components were evaluated, the maximum shear stress has been derived. As explained in the state of the art, the main mechanism for crack formation in lubricated bearing gears is the shear driven Mode II. For this reason, in this model all the considerations on crack generation had assumed the maximum shear values as critical stress for crack initiation.

Crack initiation, on the other hand, has been predicted by applying the principles of continuum damage mechanics. Introducing a material parameter to account for the void density in the material and the correlated damage evolution, leading to crack formation. Considering the hysteresis loop formed by the stress-strain curve and the relative plastic deformations in the material, the amount of damage formation at each cycle will be calculated.

The final aim was to obtain a first approximation of initiation and fatigue life, relying only on experimental data, from simple tension tests on the material, and the load conditions [16]. Thus, the final results had given an indication, starting from a specific set of data on the material and load conditions, on the number of cycles

needed for crack generation and its depth.

A finite element analysis has been also developed to have a direct reference to the stresses generated in the specific system. This analysis has been conducted on a bearing gearbox model based on an actual helicopter's transmission, which is similar to the model that had catastrophic incidents caused by fatigue failure of the previously mentioned bearing gear (Fig. 1.3).

A section of an epicyclic transmission has been modeled in its sun, planet, and outer ring parts. The torque coming from the engine has been evaluated from the specifications on the real engines mounted on the actual reference helicopters. The load has been associated with an output rotational speed, and this combination has been applied to the system, which transmits the solicitation through the gears and to the output shaft by rolling contact with the rolling bearings present in the component.

This interaction is the one generating the critical phenomenon introduced previously, responsible for fatigue. In the FEM model, the rolling contact has been replied and the analysis is focused on the stresses generated in this specific section of the planet gear.

The stress values extracted from the FEM analysis has then been compared with the analytical-numerical model, to check the correct design of the system under specific load conditions.

In Fig. 1.6 is shown a general flowchart of the thesis work.

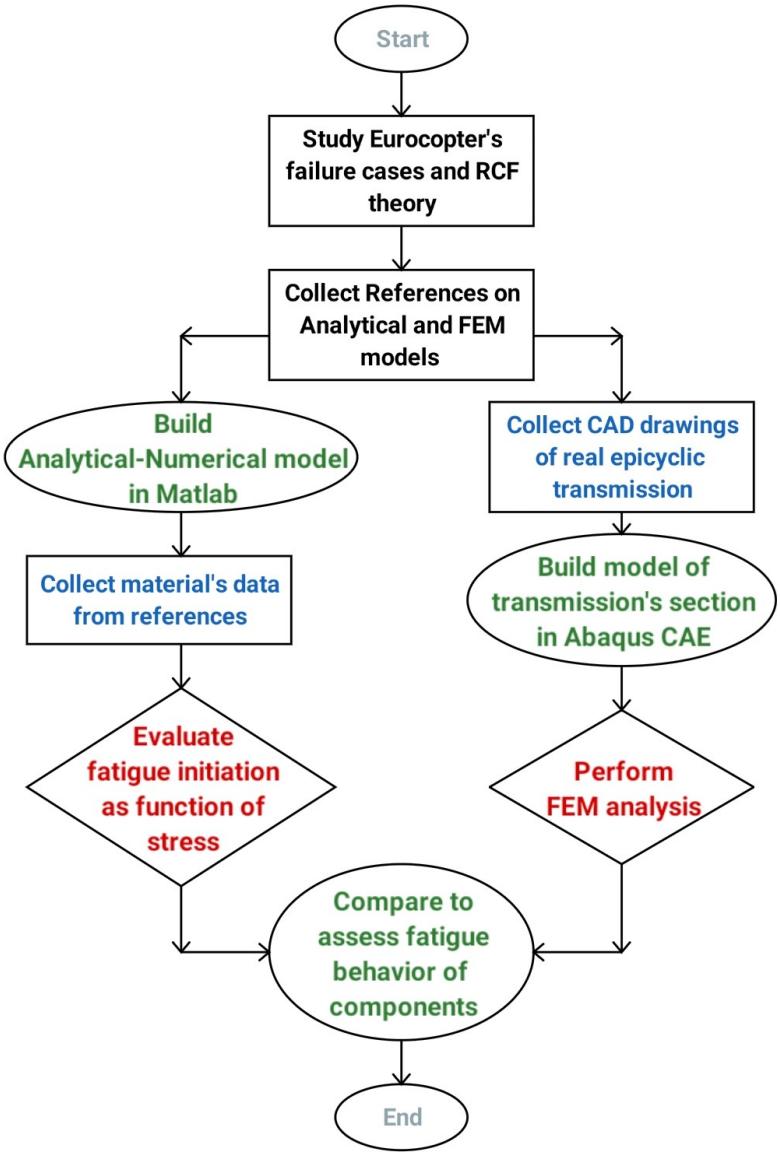


Figure 1.6: *Flowchart of the general procedure of the current thesis work*

2 State of the art

2.1 Bearings

The defining aim of mechanics is the development of moving systems. From the beginning, it has been clear that relative motion between machinery elements is not a smooth operation. Friction plays an important role not only in generating grip and allowing the transmission of forces but has also the downside of producing wear. Bearings were invented to permit the relative motion between components and the transmission of loads, minimizing the loss of power. The same Oxford English Dictionary refers to bearings "as a part of a machine that allows one part to rotate or move in contact with another part with as little friction as possible" [6].

Figure 2.1: *Different types of bearings [17]*

Although these concepts may sound modern, the principle of bearing function has been known for more than four thousand years. In ancient Egypt solutions to reduce friction and allow easier movement of heavy loads was used in the building of pyramids. More recently, the Italian genius of Leonardo da Vinci proposed some models for machinery that comprehended parts remarkably similar to the modern bearings. Finally, the last push to the development of these fundamental elements was given by the Industrial Revolution in the 18th century and the refinement that brought to the actual configurations during the World Wars in the XX century.

Bearings are now used in all kinds of mechanical systems from transportation machines such as trains, airplanes, and cars and from heavy industrial machinery to any kind of domestic appliances as washing machines, refrigerators, vacuum cleaners and computers. It is a simple consequence of this wide field of application, that bearings are divided into many different types. Lelikov [18] reports that only in cars there are 30 different types of bearings, in trucks more than 120 and in airplanes more than 1000. Depending on the nature of the loads involved we can distinguish [19]:

- **Radial bearings:** for loads acting perpendicular to the shaft's axis
- **Thrust Bearings:** for loads acting parallel ot the shaft's axis

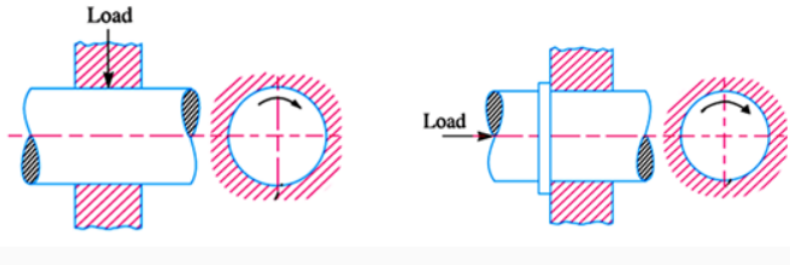


Figure 2.2: *Radial bearing and thrust bearing [19]*

Depending instead on the nature of the contact involved [19]:

- **Sliding contact bearings:** along contact surfaces sliding occurs between a moving and a fixed element. Usually a lubricating film is used to separate the surfaces. Included in this category are also plain bearings, journal bearings and sleeve bearings.
- **Rolling contact bearings:** rolling elements are introduced between the fixed and the moving part. This changes the friction from sliding to rolling type.

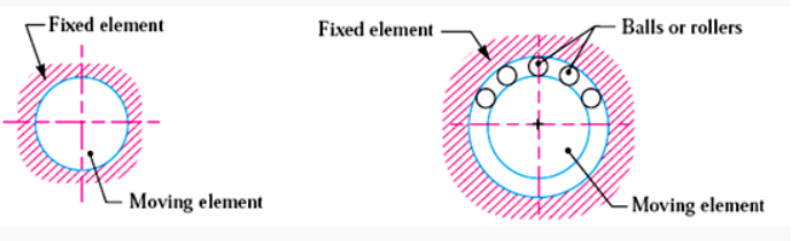


Figure 2.3: *Sliding contact bearing and Rolling contact bearing [19]*

Comparing the friction modes which distinguish these two types of bearings we have different conditions [19]. In sliding, the initial friction is very high due to metal contact and then decreases only when the movement speed increases and the lubricating oil film can be established. In rolling conditions instead a state called "pure rolling" can be achieved, in which friction is zero. From this effect, rolling bearings are also known as anti-friction bearings [18][19]. In reality, the deformation induced by contact leads to a small but positive value of friction.

Sliding contact bearings have the capacity of high load absorption and low sensitivity to impacts, due to simple geometry and large lubrication. But the high amount of initial torque and lubricant consumption limit their applications. On the other hand, rolling contact bearings are more complex and subjected to impacts effects, but offer much less torque resistance and maintenance need, making them perfect for driving unit applications. In this thesis work, the focus will be on a transmission element, and therefore on rolling contact bearings.

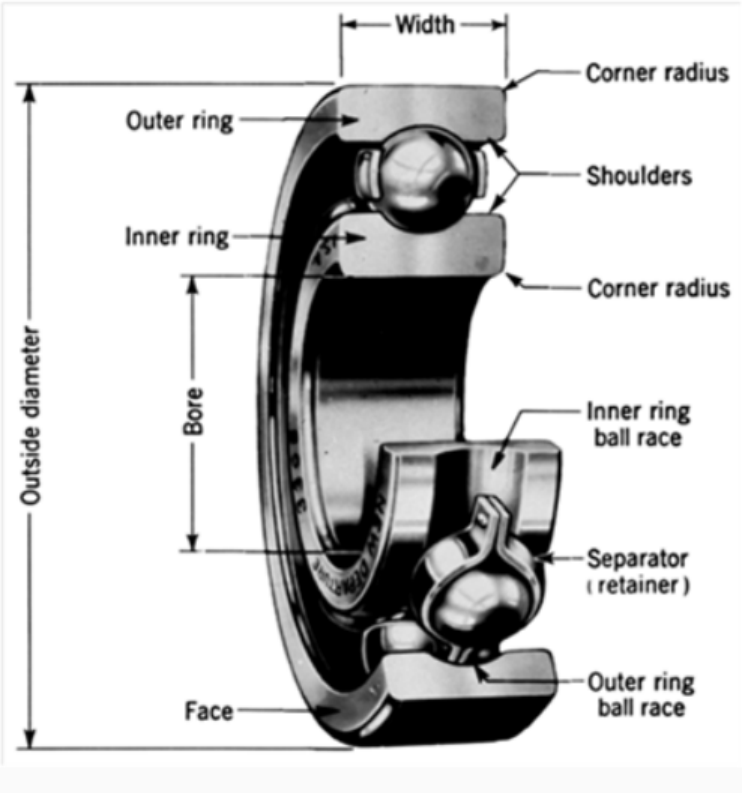


Figure 2.4: *Rolling contact bearing section [19]*

2.1.1 Rolling Contact Bearings

In general, contact bearings have the same basic structure, composed by some specific elements:

- Outer race
- Inner race
- Balls/Rollers (rolling elements)
- Retainers

Usually, the moving element, at which the bearing is joined, is connected to the inner race. Instead, the outer race is usually fixed but in some cases, such as in bearing gears, is joined to a rotating or moving object.

In this case, the outer race moves rigidly as the external component and the inner race rotates inside. Between these two surfaces, there are the rolling elements, which can be spherical, cylindrical or of many other shapes. The geometry is chosen depending on the application and the direction of the loads to be transmitted. Finally, the so-called retainers are components such as cages, which keep the rollers separated and

in position. Other kinds of elements that can be listed as retainers are the shields positioned around the cage, which protect the rollers from external elements and retain lubricants inside.

The wide range of applications requires a large variety of configurations for rolling bearings. There are many different possible geometries and configurations in rolling bearings and any manufacturer has its specifications. But the necessity of interchangeability and uniformity lead to standardization of construction parameters. Even if grouped, also in the ISO standards [20] are reported numerous different categories of rolling bearings:

- Ball bearings
- Needle roller bearings
- Cylindrical roller bearings
- Spherical roller bearings
- Inert bearings
- Spherical plain bearings
- Tapered roller bearings

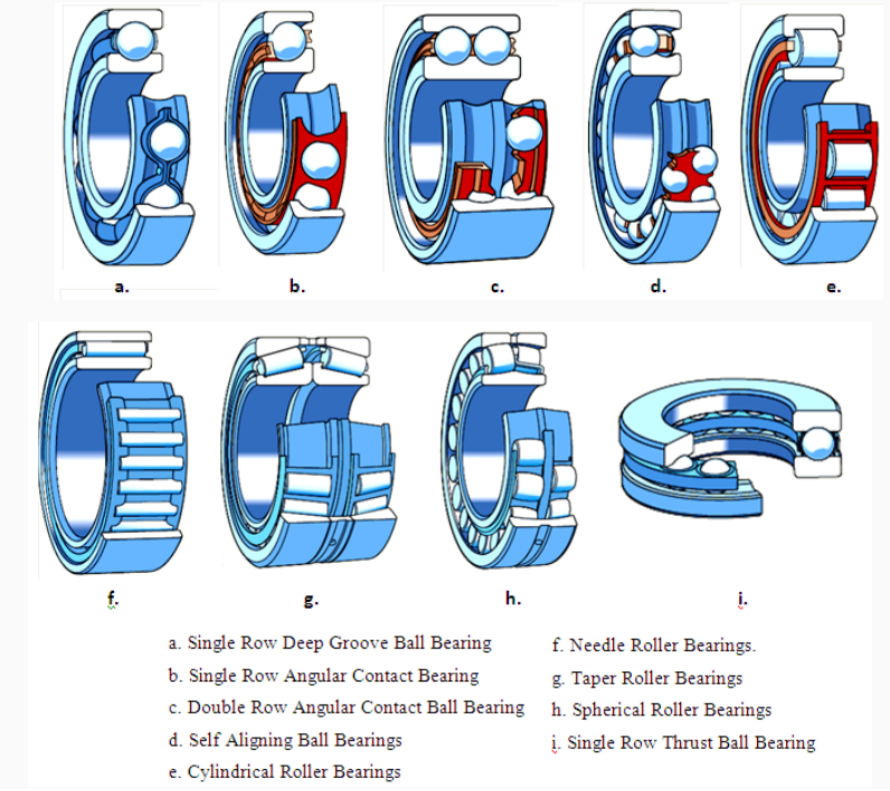


Figure 2.5: *Types of rolling contact bearings [19]*

The most common type is without any doubt the single row ball bearing (Fig. 2.5(a)), and can transmit both axial and radial loads. In the case of an axial load anyway, the rings get slightly displaced and the balls may have contact with the sidewall of the races. This configuration is usually applied as support for shafts. Needle roller bearings are a particular configuration in which the rollers have a very high length to radius diameter (from 3 to 8)[19], and can work under intense radial loads but are highly sensitive to axial loads and mistaken alignment, which limits their application in vibrating systems.

Taper rollers "have a shape of a frustum of a cone" [19] and are suitable for higher load both on axial and radial directions. The specific orientation of the rollers, converging at the axis of the bearing, makes it possible to achieve a condition of almost pure rolling. The double set of rollers makes it possible to withstand axial load in both directions.

Cylindrical roller bearings exploit the linear shape of the contact surface to support larger radial loads with respect to ball bearings. On the other hand axial loads may be critical and also an accurate alignment is required. This problem may be mitigated by adopting the curved "barrel" shape typical of spherical sets, which increases axial resistance and alignment capability, maintaining an elongated print on the surfaces. In this way, the configuration adopted will be similar to a single-row spherical rolling bearing. This configuration is used in motors, train axles and gearboxes.

2.2 Materials for bearings

The job of bearings is to endure severe loads, static and cyclic, while operating in difficult conditions and environments [6]. Machinability and mechanical properties make steel the most suited material for the manufacturing of these components. As said before, refinement in actual bearing technology comes from more than a century of work and it includes the study on specific alloys.

A report on the advances in metallurgy and processing technologies in this field has been given by Zaretsky [21][22]. Bhadesia [6] has reported an extended review on the metallurgical side of bearing production and design, from which will be reported some principal concepts. The initial production process of bearing steels, as for many other applications, implies an intense plastic deformation of the cast metal, formed into square billets [6]. Further machining operations such as rolling, drawing and heat-treatments help to reduce the dimensions of the billets and release the accumulated detrimental stresses in this stage. After the material has been cut into pieces of dimensions suitable for the manufacturing of balls, the final stage is performed by quenching and tempering to induce the hardness required by the specific application.

Rings instead, may be produced from seamless tubes [6] by a process of hot rolling and hardening. All the parts obviously undergo machining, grinding and polishing to achieve the tolerances on shape and surface finish. Materials selection for such applications as bearings should then also take into account the complex manufacturing processes, required to realize the components. This is why machinability is an important quality to be considered, together with the mechanical properties, necessary for the final application.

From this assumption is easy to understand how the dominating materials have al-

ways been the steels originally designated for machining tools. This category is usually characterized by a carbon content in the range of 0.8-1-1 wt% and a total substitutional content under 3 wt% [23][24]. The relatively high carbon content can speed up the production of cementite and allow it to perform soft annealing on the components. These steels are also prone to hardening, which is of capital importance in wear applications. Their structure can be transformed into martensite or cementite, depending on the speed of the process, by quenching from a temperature at which the material is mostly austenitic. These two crystalline structures are much harder than austenite and well suited to load-bearing and wear resistance applications. Not only the chemical composition of carbides and alloying products is favorable but in particular the grain structure.

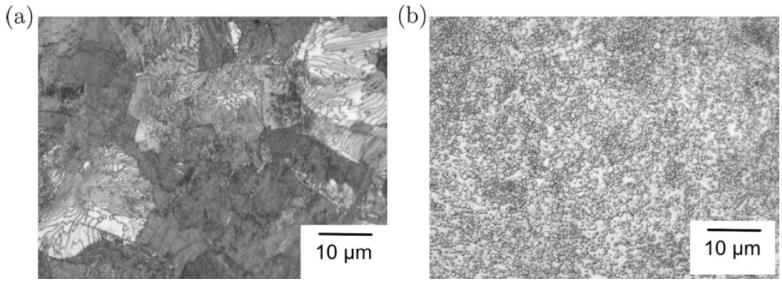


Figure 2.6: *(a) Microstructure of hot-rolled 52100 steel as supplied by the manufacturer. (b) Microstructure after spheroidisation. Courtesy of W. Trojahn [6]*

Grains are smaller and homogeneous, then able to behave better under loading and offering smaller weak planes for crack formation (Fig. 2.6) Subsequent tempering to relax internal unwanted tensions is usually performed. In the case of large dimensions, however, this process becomes more difficult. Carburizing may, then, become necessary to produce harder structures inside the material. Also increasing the concentration of alloying elements the same effect will be obtained, but with consequences on the mechanical properties.

This process is aimed at increasing material's hardness and thus called hardening. As stated before, it can affect all the components or only the surface region, depending on the specific dimensions and requirements. It can be the be distinguished:

- **Throughout hardened steels:** in which all the section is transformed in a martensitic or bainitic condition
- **Case or induction hardened steels:** having a softer core and hardened surface layers

Different procedures can be adopted and many variables as time, temperature and heating ratio can influence the final results. To determine the optimal sequence of operations, a refined analysis and design of the component should be made. When case hardening is desired, for example, carburizing and nitriding are the most popular processes.

In this case, the main advantage throughout hardening is to have a hard surface maintaining a tougher core. This feature is of particular interest in rolling contact elements. because the softer core allows the material to absorb better the stresses at which it is subject, thus maintaining high wear resistance at the surface.

Also residual stresses are induced with this kind of treatment, which improves fatigue life. Usual hardening depth is around 1 mm but can higher or lower depending on the machining needed. One other factor to consider is the cost of the treatment, which is obviously higher for through-hardened components. For 52100 steel family, hardness of 1000 HV can be achieved at the surface with laser treatment.

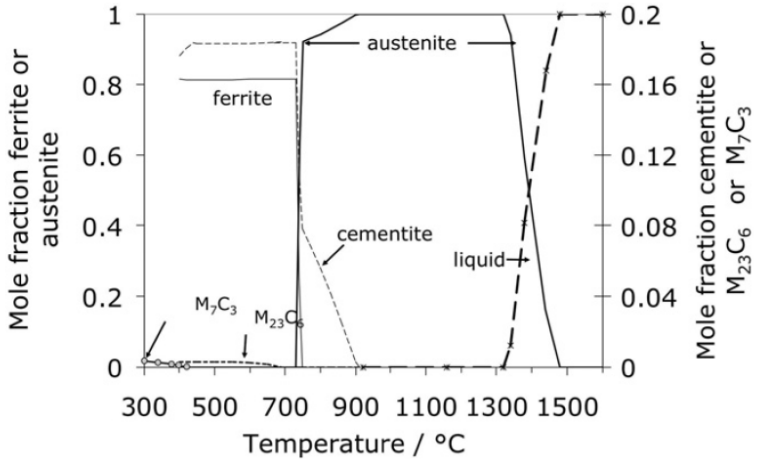


Figure 2.7: *Phase fraction transformation in steel 1C-1.5Cr [25] as function of temperature [6].*

A table of the most popular alloys in this category has been reported in Fig. 2.8.

The most employed quality is the 52100 series, containing 1.0-1-5 wt% Cr and covers an important part of the total production in the world, which amounts to about six million tonnes per annum.

Apparently, the study on these steels can be traced back as far as 1901 [26], when Stribek performed some tests to assess the suitability of the material for ball bearings application. Then in 1905 the producer Fichtel and Sachs of Schweinfurt adopted the material in the manufacture of bearings, continuing to use it up to this day [24][27].

Of course, during the years some improvements have been done, in particular in the cleanliness of the metal structure and inclusions. It is known that all oxygen present in the alloy tends to form oxides [28][29]. These inclusions, usually segregate at the grain boundaries which are already weak spots for crack formation. The initiation of damage is then highly probable in the vicinity of this kind of defect, leading to a limit in the oxygen concentration in modern bearing steels. This value is reported to be less than 10 ppmw [30][31].

2.3 Hertzian Theory

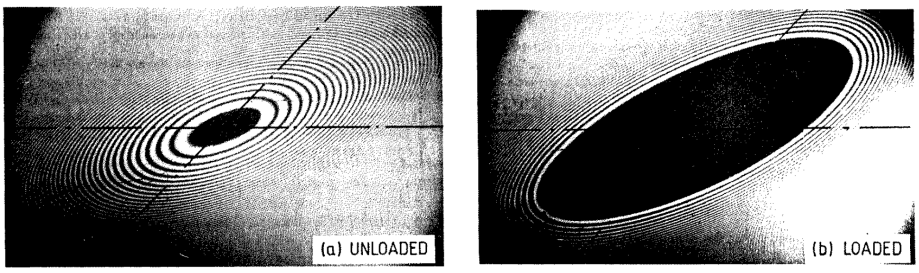
K.L.Johnson is the author of a comprehensive book on contact mechanics [5], from which reference will be taken for this paragraph. Contact between elastic solids is a particular kind of interaction, which was first described by Hertz in 1882 [5]. The first

study was focused on the influence of elastic deformation on materials by investigating Newton's optical interference fringes in the gap between two lenses (Fig. 2.9).

The effect of contact pressure between the bodies was so first observed during the Christmas period of 1880 when Hertz was twenty-three years old. From this experiment, a publication followed and aroused great attention, standing at the tests ever since. The theory taking his name describes the stresses and deformation generated by solid bodies, and specifically in the case of non-conforming contacts.

Grade	C	Mn	Si	Cr	Ni	Mo	Cu	S	P	Others
AISI 1070	0.71	0.76	0.20	0.09	0.08	0.02	0.07	0.012	0.006	
En31	0.90–1.20	0.30–0.75	0.10–0.35	1.00–1.60				0.05	0.05	
AISI 52100	0.95–1.10	0.20–0.50	⩽0.35	1.30–1.60	–	–	⩽0.025	⩽0.025	–	
SAE 52100	0.98	0.38	0.16	1.39	0.07	0.02	0.12	0.06	0.12	
SAE 52100	0.97	0.31	0.32	1.43				0.017	0.019	
"1C–1.5Cr"	0.98–1.10	0.15–0.35	0.25–0.45	1.30–1.60	⩽0.25	⩽0.10	⩽0.35	⩽0.025	⩽0.025	
ShKh4	0.98–1.03	0.18–0.29	0.17–0.28	0.38–0.47						
ShKh15	1.05	0.28	0.28	1.50	0.11		0.06	0.015	0.013	
ShKh15G	0.95–1.05	0.95–1.17	0.45–0.61	1.35–1.60						
SUJ-2	1.03	0.37	0.23	1.35	0.51	–	0.15	0.023	0.018	
SUJ-2	1.04	0.39	0.25	1.47	0.07	–	0.19	0.005	0.015	
SUJ-2	1.01	0.36	0.23	1.45	0.04	0.02	0.06	0.007	0.012	O 0.0008
SUJ-2	0.98	0.37	0.18	1.42	0.07	0.03	0.12	0.003	0.016	O 0.0006
SUJ-2	0.97	0.38	0.20	1.35	0.08	0.03	0.11	0.005	0.016	O 0.0005 Al 0.009 Ti 0.0023
MMM (SKF 3M)	Mo modified 52100					0.25			⩽0.015	
"1C–1.5Cr–Mo"	0.98	0.45	0.97	1.98		0.42		⩽0.002	⩽0.011	
"Si–Mo"	0.96–1.12	0.56–0.66	0.49–0.70			0.23–0.33				
100CrMo7–3	0.97	0.66	0.27	1.79	0.11	0.26	0.15	0.007	0.009	O 0.0005 Al 0.034
52CB	0.85	0.35	0.85	0.90		0.60				
Microalloyed	0.44	0.99	0.43					0.004	0.009	V 0.10
4320	0.21	0.62	0.20	0.49	1.73	0.20	0.16	0.018	0.010	
1070M	0.68	0.95	0.17	0.13	0.11	0.05	0.19	0.022	0.009	
S53C	0.53	0.74	0.19					0.020	0.015	O 0.008
SMn60	0.60	1.22	0.24					0.011	0.007	O 0.008
SMn65	0.66	1.19	0.24					0.009	0.006	O 0.009
SAE 1072	0.74	1.18	0.23					0.009	0.008	O 0.007
SAE 1072Cr	0.72	1.18	0.24	0.48				0.009	0.006	O 0.006
80CrMn4	0.78	0.78	0.24	0.82						
100Cr2 (W1)	0.9–1.05	0.25–0.45	0.15–0.35	0.4–0.6	⩽0.30		⩽0.30	⩽0.025	0.03	
100Cr4 (W2)	1.0–1.1	0.25–0.45	0.15–0.35	0.90–1.15				⩽0.025	0.03	
SAE 51100	0.97	0.39	0.25	1.04				0.013	0.020	
100Cr6 (W3)	0.90–1.05	0.25–0.45	0.15–0.35	1.40–1.65	⩽0.30		⩽0.30	⩽0.025	0.03	
100CrMn6 (W4)	0.90–1.05	1.00–1.20	0.50–0.70	1.40–1.65	⩽0.30		⩽0.30	⩽0.025	0.03	
100CrMnMo8	0.90–1.05	0.80–1.10	0.40–0.60	1.80–2.05	⩽0.30	0.50–0.60	⩽0.30	⩽0.025	0.03	
GCrSiWV	0.98	0.49	0.75	1.50				0.004	0.012	W 1.21, V 0.29

Figure 2.8: *Composition ranges (wt%) of common bearing steels [6].*



(a) (b)

Figure 2.9: *Interference fringes at the contact between two lenses inclined by a 45° angle. (a) Unloaded, (b) loaded [5]*

The distinction between conforming and non-conforming type of contact depends on the geometry of the bodies. In the first case, the surfaces 'fit' exactly [5] or closely, without deformation. An example of conforming contacts is journal bearings. In the case of different surface profiles instead, the definition of non-conforming contact is used.

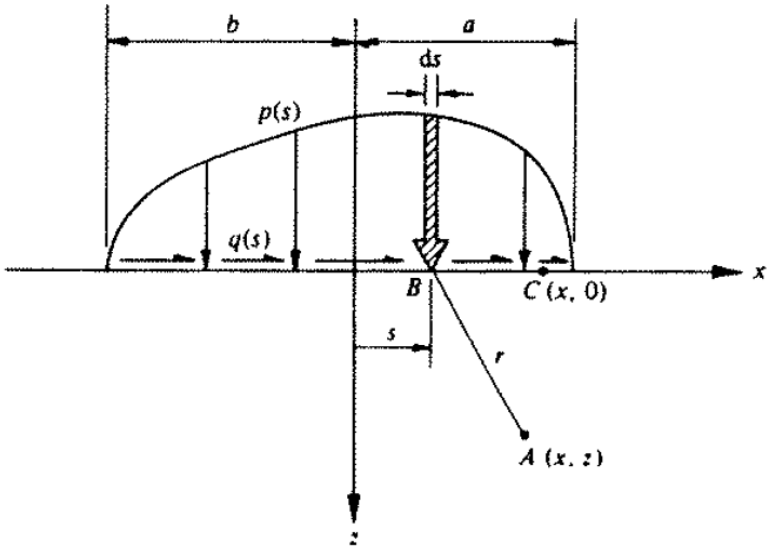


Figure 2.10: *Hertzian load profile [5]*

Roller bearings are the main example. When the roller makes contact with the race without deformation, the first interaction occurs along a line or a point, depending on its geometry. The case of line contact is a mix between the former two described. To generate a line contact, the two surfaces should be conforming in only one of the two directions in the plane of contact.

The non-conforming contact area is smaller with respect to the bodies involved. In the region close to the contact surface the stresses should be highly concentrated. From the initial Hertz's observations on interference fringes, the contact area was defined as generally elliptical, as shown in Fig. 2.9 [5]. Considering the symmetry of the shape and the bodies' geometries, Hertz introduced a simplification to calculate local stresses and deformation. Each body is then taken as an elastic half-space over which an elliptical loading region is applied. The validity of the operation depends on the satisfaction of two conditions:

- $a << R$: the contact are should be smaller than the relative radius of curvature
- $a << R_{1,2}, a << l$: the contact area should be smaller than the significant radii of the bodies and their lateral and depth dimensions

Identifying the major dimension of the contact area as a, and $R, R1, R2$ as the relative and bodies' radii.

The imposition of a small area with respect to the components' dimension aims at eliminating any interaction between boundaries and the region of interest. The

curvature condition instead is necessary to allow to approximate the material to a flat half-space and assure that strains can be evaluated within the linear elastic limits.

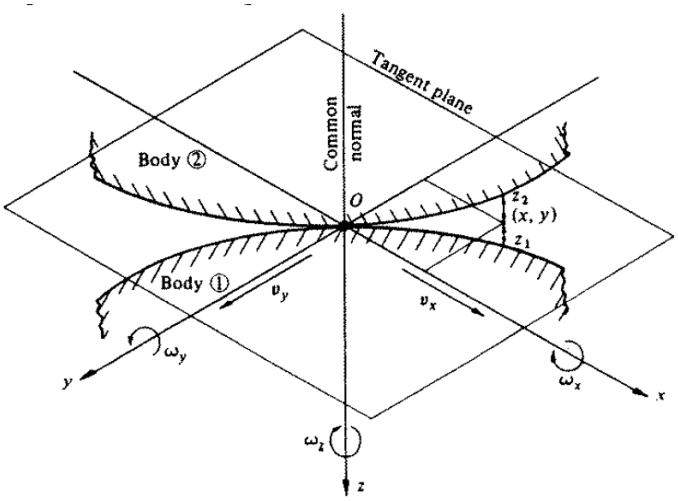


Figure 2.11: *Non conforming contact [5]*

This last consideration is of great importance when studying metallic materials. These assumptions are made to decouple the stresses generated purely by contact from those due to support configuration and shapes. Moreover, the adopted particular geometry is well known in the boundary conditions solution.

Considering these assumptions, the integral form equation for contact stresses can be formulated, for generic load distribution. It should be taken into account that contact surfaces transmit also tangential traction due to the presence of friction and not only normal pressure. This generates a multi-axial state of stress.

In Fig. 2.10 it is shown a generic load distribution in the region $-b < x < a$, over which are acting the normal pressure $p(x)$ and the tangential traction $q(x)$. The stress components at any point A in the elastic half-space can be found by this equations:

$$\sigma_x = -\frac{2z}{\pi}\int_{-b}^{a}\frac{p(s)(x-s)^2 ds}{[(x-s^2)+z^2]^2} - \frac{2}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)^3 ds}{[(x-s)^2+z^2]^2} \tag{2.1}$$

$$\sigma_z = -\frac{2z^3}{\pi}\int_{-b}^{a}\frac{p(s) ds}{[(x-s^2)+z^2]^2} - \frac{2z^2}{\pi}\int_{-b}^{a}\frac{q(s)(x-s) ds}{[(x-s)^2+z^2]^2} \tag{2.2}$$

$$\tau_{xz} = -\frac{2z^2}{\pi}\int_{-b}^{a}\frac{p(s)(x-s) ds}{[(x-s^2)+z^2]^2} - \frac{2z}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)^2 ds}{[(x-s)^2+z^2]^2} \tag{2.3}$$

The distributed forces acting on infinitesimal area elements are considered as concentrated infinitesimal forces $p(s)ds$ and $q(s)ds$. The variable s is the dummy variable and its interval covers the region of load application. Integration in closed form is difficult, but it can be evaluated if the force distribution is known.

Non-conforming elastic solids must be considered in this thesis work.

The general expression for elastic displacements is:

$$\vec{u}_{z1} + \vec{u}_z 2 < \delta - Ax^2 - By^2 \tag{2.4}$$

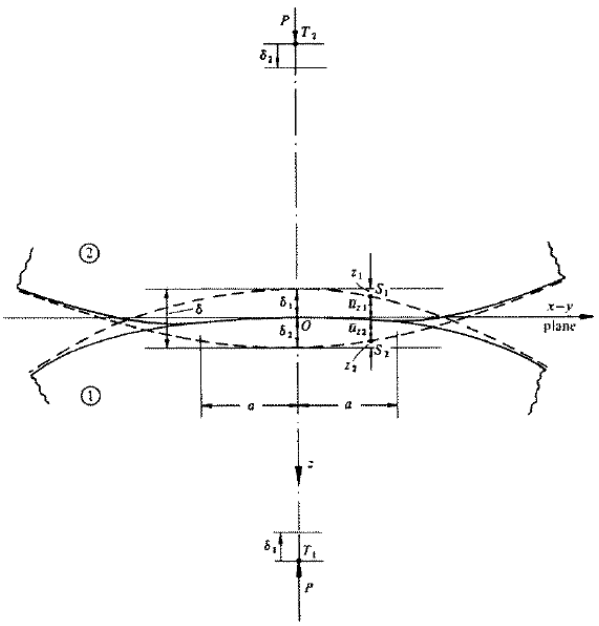


Figure 2.12: *Non conforming contact and relative deformation [5]*

From the configuration in Fig. 2.11, a geometry of simple solids of revolution is imposed.

$$\begin{aligned} R_1^{'} &= R_1^{''} = R_1 \\ R_2^{'} &= R_2^{''} = R_2 \end{aligned} \tag{2.5}$$

The generated contact area will have a circular shape with radius a. Imposing the circular condition on the area it is found that:

$$\frac{1}{R} = (\frac{1}{R_1} + \frac{1}{R_2}) = A = B \tag{2.6}$$

Thus, the relative boundary condition on the displacements becomes:

$$\vec{u}_{z1} + \vec{u}_{z2} = \delta - (1/2R)r^2 \tag{2.7}$$

The distribution of pressure should then satisfy this equation and it has been found by Hertz:

$$p = p_0[1 - (r/a)^2]^{1/2} \tag{2.8}$$

Considering the pressure acting on the two bodies as equal, the dimension of the contact is given by Eq. 2.9 or 2.10, if the total load is specified.

$$a = \pi p_0 R/2E^* \tag{2.9}$$

$$a = (\frac{3PR}{4E^*})^{1/3} \tag{2.10}$$

By dimensional reasoning and substitution in general equations, it can be verified that this formulation doesn't give rise to stresses outside the contact area. Moreover, the uniqueness of the solution is verified by imposing a pressure distribution of a slight different formulation:

$$P = \int_0^a p(r) 2\pi r dr = \frac{2}{3} p_0 \pi a^2 \tag{2.11}$$

This pressure distribution will produce a uniform normal displacement inside the loading region. When subtracting or adding this pressure to the Hertz pressure obtained before, the condition on displacements in Eq. 2.7 is still satisfied. If added, a punch-like cylindrical deformation will occur, generating an infinite gradient of the surface.

This will not be possible for elastic smooth solids without considering the formation of interference. If instead subtracted from the Hertzian pressure, tensile normal traction of infinite magnitude will be generated. But this condition is also impossible due to the lack of adhesion between the surfaces. It is then concluded that the Hertz pressure formulation is the unique solution to the problem.

In Fig. 2.13 are shown the stresses distribution at the surface and along the axis of symmetry. A comparison between a uniform pressure and the Hertzian distribution is shown. It can be seen how all the stresses on the surface are compressive except the radial stress at the edge of the contact. This is responsible for "ring cracks" formed at contact in brittle materials such as glass.

From integration along z axis of the expressions, can be found that the principal shear stress has the formulation:

$$\tau_1 = \frac{1}{2} |\sigma_z - \sigma_r| \tag{2.12}$$

The maximum value is reached at the depth of $0.48a$ (with $v = 0.3$, Poisson ratio), with a maximum value of $0.31p_0$ [5]. At the origin the shear stress is lower, around $0.10p_0$, and also at the edge it reaches the value of $0.13p_0$. It can be concluded then that the elastic maximum yielding will initiate beneath the surface. However, not only shear is present but also normal stresses in radial and tangential direction (considering the circular symmetry). As shown in Fig. 2.13. This makes the state of stress multiaxial. Even if it is shear that is said to generate the most yield, the other stress contributions may be triggered by operational conditions and interactions between the surfaces and external factors. This will influence the mechanisms dominating fatigue crack growth.

2.4 Rolling Contact Fatigue

If properly designed, rolling elements are loaded, lubricated and positioned to withstand the work conditions without any critical failure. In this case, fatigue becomes the dominant failure mechanism. Moreover, the stresses at the contact between rollers and gears are significantly higher compared with structural rotating components, as described by Hertzian theory. Rolling contact fatigue is the result of these loading cycles.

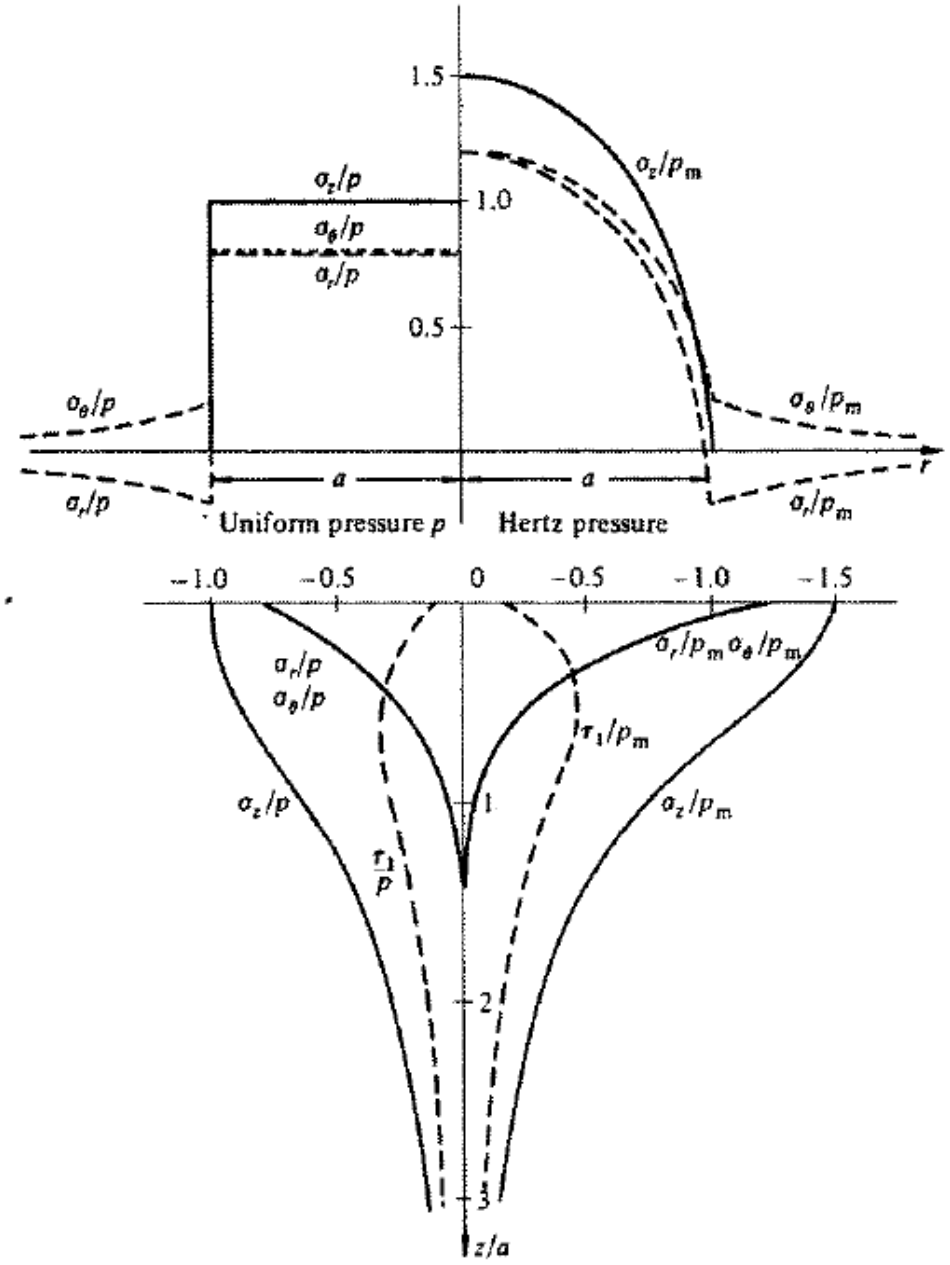


Figure 2.13: *Hertzian stress distribution at the contact surface and along the axis of symmetry. Comparison between (left) uniform load pressure and (right) Hertzian pressure circular profile [5]*

The usual outcome of RCF is the flaking of metallic particles or in extreme cases, the complete failure of the component [7]. The differences between classical fatigue and RCF are many other, enough to require separate theories for the study of rolling elements. The main differences, which make it impossible to use classical approaches onto RCF are [32]:

- State of stress is multiaxial [5].
- Loading history is non-proportional; i.e. each stress component has a different phase (peaks do not coincide in time) and different ratio (τ_{xz} undergoes a complete reversal while σ_x, σ_z are only compressive).
- Critical planes and principal axes constantly change in direction [32]. It is difficult to identify the plane of maximum damage.
- The contacts are highly localized, involving small volumes. Usual contact dimensions are about $200\mu m$.
- RCF life can be divided in initiation and propagation phases (plus failure, which is unstable and is usually not studied). These two steps are different by driving mechanisms and phenomenology.

In previous sections (Chap. 2.3), the nature of the stress tensor and its characteristic configuration have been analyzed. Now the focus will be on fatigue life phases and their description, based on dominant stresses and consequences on the components.

Fatigue life (N) is commonly divided in two phases: initiation (n_i) and propagation (n_p) (Eq. 2.13).

$$N = n_i + n_p \tag{2.13}$$

Before 1980 the majority of the studies have been based on crack initiation. Lundberg and Palmgren [8] estimated that for bearings fatigue life, the propagation phase might be very short. This theory was verified by experimental tests by Yoshioka and Fujiwara [33]. Many modifications were made to this model during the years, in order to adapt it to particular operational configurations. Some notable examples are Tallian [11], Ioannides-Harris [9] and Zaretsky [34].

These studies assumed that crack initiation should be the dominant process for rolling contact conditions in bearing steels.

Otherwise, many studies as the Keer-Bryant model [35] or more recently by Benedetti et al. [36], show that the presence of micro-cracks arises in the very early stages of fatigue life, giving more relevance to propagation. Both theories have validity because are based on different assumptions and observations on the configuration of the component, loading conditions and operational parameters.

2.4.1 Modes of Crack Propagation in RCF

Every stress applied in a different direction causes the crack to open in a different mode. Each categorized by fracture mechanics:

- **Mode I**: application of a tensile stress normal to the plane of the crack, which causes its opening

- **Mode II**: application of shear stress parallel to the crack's plane and perpendicular to crack front, resulting in a sliding mode
- **Mode III**: also application of shear stress parallel to the plane of the crack, but also parallel to the crack front. The result is a tearing mode

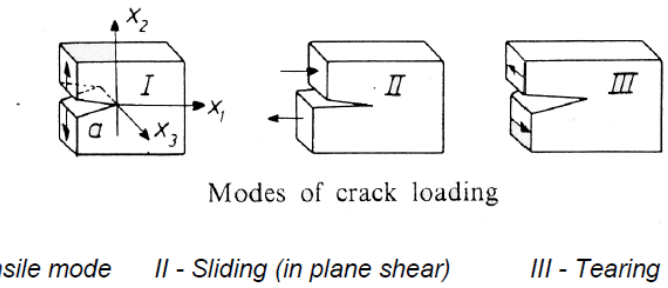


Figure 2.14: *Growth mechanisms for fatigue cracks [37]*

To understand the specific propagation behavior of fatigue cracks, Fleming and Suh [38] introduced the use of fracture mechanism in order to evaluate the stress intensity factors. Deriving from Griffith's crack theory [39] for the failure of brittle materials, Irwin's [40] assumptions on energy contribution in cracks were developed by the theory introduced by Paris [41]. The rate of crack growth was then related to the stress intensity factor, coming from the Paris-Erdogan law [42]:

$$\frac{da}{dN} = C(\Delta K)^m \tag{2.14}$$

Where da/dN is the crack growth on the single cycle and C and m are experimental material parameters. The stress intensity factor range ΔK is the difference between the maximum and minimum values in a fatigue cycle:

$$\Delta K = K_{max} - K_{min} \tag{2.15}$$

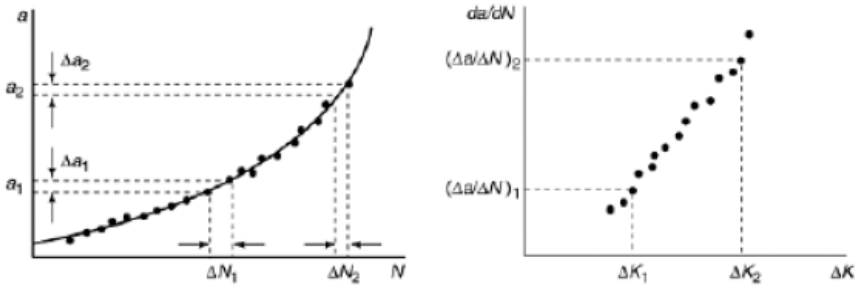
The Paris-Erdogan power law can be plotted on a double logarithmic scale to obtain a linear relationship between the range of the SIF, on the x axis, and the effective crack growth rate on the y axis. When referring to complex loading states with different stresses acting on a crack, each component of the stress tensor will have its relative stress intensity factor range. Usually, the higher range value identifies the dominant stress component for crack growth in that condition.

On this theoretical basis, it will be shown how different conditions in rolling contact fatigue can favor crack initiation and propagation in the different modes. The competitive mechanism between the stresses of the multi-axial tensor of the Hertzian load will be broken up and analyzed in its different components.

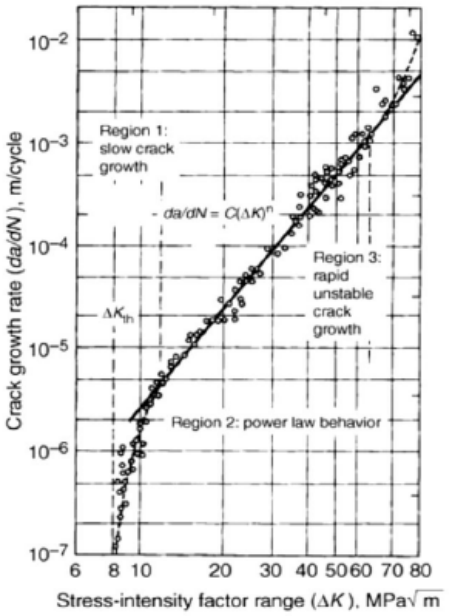
2.5 Fatigue Cracks in Rolling bearings

The possible applications and working conditions available for REBs are widely different. Numerous are also damage mechanisms that can be activated, as stated by

Roylance [43]. A first distinction between the main causes leading to failure in rolling contact conditions can be made in wear, as adhesion, corrosion and abrasion, or contact fatigue.



The ideal sequence for the determination of da/dN vs. Δk behaviour of materials



The resulting fatigue crack growth curve for an engineering structural material

Figure 2.15: *Determination and plot of Paris' law [44]*

As explained before, wear phenomenons are highly targeted during the steps of material selection, surface finishing and hardening processes, in order to be prevented. This makes them almost ineffective and leaves fatigue as the most important source of failure. Fatigue can manifest itself in different ways, depending on its source.

It should be clear that all mechanisms can be activated at any time and also simultaneously, but they will have different rates. In many cases, the activation of one mechanism can also hinder the activation of another one. This competitive situation depends on the loading intensity, materials characteristics and working conditions, which may favor one mechanism in particular combinations. As stated by Hyde [45], there are three principal effects of contact fatigue, being:

- **Case crushing**: sub-case phenomenon, typical of hardened materials

- **Spalling**: subsurface initiated
- **Pitting**: micro or macro phenomenon, surface originated

Pitting and spalling have a similar nature and shape but differ from one another by depth. Pitting has usually a depth around $10\mu m$ and it is strongly dependent on surface interaction parameters such as friction, roughness and lubrication [46]. Spalling instead is characterized to be deeper, from around 20 to $100\mu m$. Identified usually with a ratio to the contact width of $0.25 - 0.35$ [4]. Tallian [47] addresses spalling as a macro-scale phenomenon, caused by fatigue crack propagation where instead pitting is limited to surface damage with different causes.

This introduces the difference between surface and subsurface fatigue. As reported by Santus et al. [46], surface pitting is supported by the interaction of lubricant fluid with the crack, which can act as an enhancer or limiter to propagation mechanisms. Micro-pitting has, by definition, typical low depth but the effect of fluid pressurization can elongate the cracks, depending on several factors. This phenomenon will be analyzed in detail in a later section.

Experimental evidence [46] shows how the first RCF surface micro-cracks are generated at shallow angles $(20 - 30°)$.

These cracks will usually generate micro-pitting but can continue to propagate in the same orientation evolving in deeper damage. According to Glodez et al. [48] and Fajdiga and Sraml [49] the following growth will maintain the same morphology of the initial micro-crack and a similar orientation.

Subsurface spalling is instead expected in high precision components such as bearings. In the case of good lubrication and high material quality, surface effects will be reduced and subsurface stresses will become dominant. However, it should be reported that a clear distinction between pitting and spalling and their interaction, has not been generally established due to the complexity and the number of factors involved.

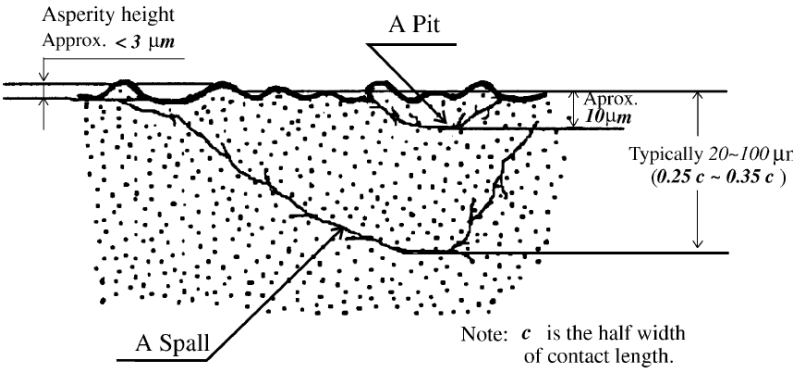


Figure 2.16: *General scheme of pitting and spalling phenomena [4]*

As reported by Bormetti et al. [50], in the specific case of bearings subjected to pure rolling, it is usually observed an initial micro-pitting generated at the surface. Thee mechanisms then often stops and gives way to more critical subsurface initiated spalling, which may cause branching and failure.

Anyway, the failure of the component is always determined by the mechanism with the fastest growth rate in that specific set of conditions, which may vary during the

service life or along a single loading cycle. When analyzing the behavior of a component during its design it is then important to consider all the damage mechanisms that can be potentially activated and evaluate the relative rates and interaction. This task is however almost impossible quantitatively, due to the influence of several parameters such as load, rolling speed, lubrication conditions, surface roughness and material properties.

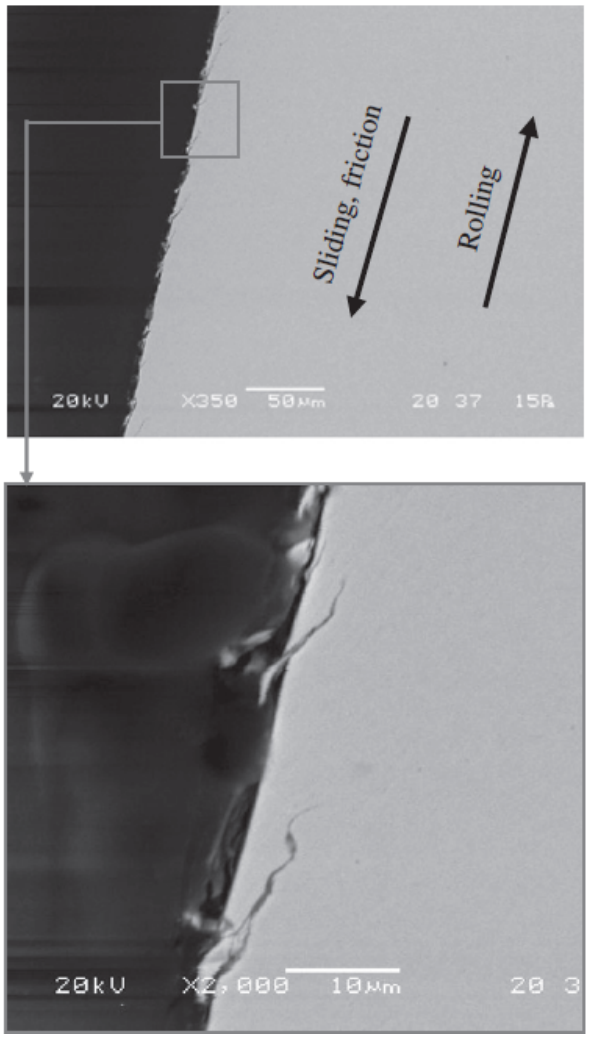


Figure 2.17: *Experimental evidence of micro-cracks. Typical formation of shallow angle cracks in gears [46]*

2.6 Surface cracks

2.6.1 Initiation

Generated at the contact interface, surface cracks are largely influenced by working conditions and material properties. Important characteristics of the material influencing formation of surface cracks are: hardness, roughness, and presence of residual stresses. It is widely known that the same rolling contact condition can generate a progressive micro-hardening in the surface region. The most relevant operating conditions instead are lubrication, contact load and presence of traction. Some authors as Moyer [51] consider lubrication and surface roughness in a unique factor as a ratio between the lubricant film thickness to composite roughness. If the lubrication parameter is high enough, it reduces the stress intensification brought by asperities. For this to happen, the lubricant film should remain continuous or the surface should be very smooth. When these conditions are respected a resulting delay in surface crack formation is obtained.

It has been demonstrated by Nelias et al. [52] that, on the contrary case of dry contact conditions, micro-cracks appear in very early stages on the surface. Tests showed how the preferred spot for crack formation is in correspondence with defects acting as stress raisers, such as inclusions, grinding marks, asperities or any source of roughness. Several authors relate surface micro-pitting to the micro-Hertzian stress field which is generated in proximity of the aforementioned asperities. The localized defects that arise have the dimensions of a few microns. Sliding and friction lead to a cumulative plastic strain in the outer layer of the material where the presence of asperities and inhomogeneity enhances the grip between the objects in contact. This mechanism is driven by traction force [52], producing plastic strain on the surface and also promoting further growth at steeper angles.

Anyway, surface crack nucleation is favored if the Hertzian stresses exceed a threshold value given by the specific material. This value is addressed as shakedown value [50] and in this case, the stress will generate shear bands in the surface layer, also in pure rolling condition.

In case of severe wear, as for railway wheels, the surface material consumption can eliminate the cracks and pits generated on the surface, re-establishing the integrity. Also, plastic strain induced by Hertzian contact can reduce stress intensification, generating a progressive hardening in the near-surface region, thus reducing the risk of crack formation and growth.

2.6.2 Propagation

Cracks nucleated at the surface have small dimensions, in the order of a few microns, and usually give way to the phenomenon of pitting. But this is not the only consequence possible for surface cracks growth. They can continue to grow, generating deeper damage such as spalling or even stop after short distances (few tenths of microns).

Way's theory [53] has been the dominating opinion in the literature regarding pitting formation [4]. It asserted that surface-initiated, crack growth was driven by the hydraulic pressure exerted by the lubricant onto the crack surfaces. The fluid

would fill the open crack and remain sealed into it when the contact rolled over the crack mouth. The pressure exerted will result in a tensile Mode I propagation.

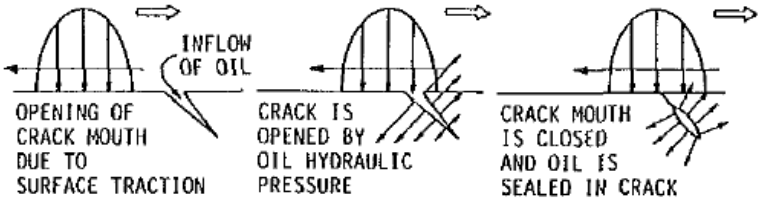


Figure 2.18: *Stages of oil seepage [54]*

Half a century later, Keer and Bryant [35], challenged this theory and proved through experimental evidence that surface cracks are dominated by shear propagation mode. Many studies have focused on this complex interaction between Mode I (tensile) and Mode II (shear) propagation in rolling contact fatigue cracks.

Bormetti et al.[50], state that in the first growth period, only the Mode II propagation mechanism is activated due to the small dimensions of the crack. At this stage lubricant interaction is hindered by the compressive field produced by the multi-axial state of stress and by the hardening treatments.

$$I = \sqrt{(\frac{\Delta K_I}{\Delta K_{Ith}})^2 + (\frac{\Delta K_{II}}{\Delta K_{IIth}})^2} \tag{2.16}$$

$$I = (\frac{\Delta K_{II}}{\Delta K_{IIth}})^2 \tag{2.17}$$

In Eq. 2.16 and 2.17, I is a failure index used for fatigue life estimation. In the general form (Eq. 2.16) both contributions from Mode I and Mode II are considered. After considerations made above, only the shear mode is considered, reducing the formulation to the form in Eq. 2.17

2.6.3 Effect of lubricant on crack growth mode

Mode II propagation role becomes less dominant once the crack reaches longer dimensions. When a crack reaches a critical length, lubricant can penetrate into the cavity and generate pressurization into the crack volume. This pressure induces tensile stress associated with Mode I crack growth, which couples with the Hertzian shear stress and pushes the crack tip to further propagation. Dallago et al. [55] assert that the action of lubricant increases sensibly the Mode I SIF, inducing the crack to curve towards the material surface.

Evidence shows that fluid seepage effects become evident when the crack reaches a dimension of 1/15 of the contact width (a). Many authors, among whose there is also Bower [56], stated that during fluid pressurization, the crack tip undergoes a complex non-proportional cycle of Mode I and II. It is evident that a general interpretation and a precise model is thus difficult to achieve. However, some studies will be proposed in order to show some of the approaches adopted to evaluate the phenomenon.

Numerical calculations [50] shows that the crack can open and allows the fluid to penetrate between the faces. Kaneta et al. [57] developed a three-dimensional

numerical model to inspect the behavior of semicircular cracks under the effect of fluid pressure. This approach confirmed the theory by which tensile growth in Mode I gains a dominant role in these conditions. Moreover, the presence of lubricant is stated as "necessary for the development of surface pitting" [57]. The the same idea is shared also by Jin [58], which confirms that internal fluid pressure increases when the load approaches the crack mouth. The cyclic iteration of this phenomenon contributes to tensile fracture mechanisms between the faces of the crack. Where $x_c^{'}$ is the crack's tip position.

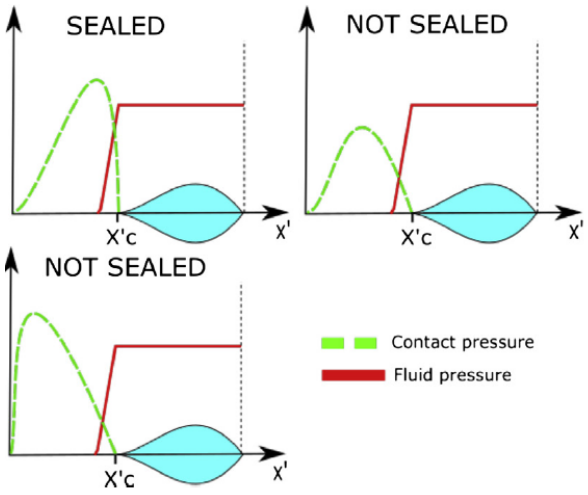


Figure 2.19: *Fluid pressure and contact pressure profiles in different cases of crack opening [55]*

Keer et al. [35] used the distributed dislocation method to calculate the stress intensity factors at the tip of a pre-modeled two-dimensional crack subject to contact loading cycles. The analysis gave the result of crack propagation occurring in Mode II. Also, Bower [56] studied the phenomenon of surface rolling contact fatigue cracks by implementing a 2D model, taking into account the interaction between crack faces and lubricant. It also concluded that the mechanisms leading to possible crack propagation were different: Mode II usual growth, fluid enhanced growth by lubrication of crack faces and lubricant seepage. Based on this study, Kaneta et al. [59] used the body force method proposed by Murakami and Nemat-Nasser [60] to confirm the analytical predictions. It was concluded that a surface-initiated crack in rolling contact conditions is prone to extend along its original plane, driven by shear Mode II growth.

However, after some degree of propagation, it switches toward tensile Mode I, being likely to branch and generate pitting. The presence of lubricant oil is the source of this transition. The fluid may exert two different types of hydraulic pressure, being a contact pressure onto the crack faces while the crack mouth is open or, when the crack closes, a pressure given by entrapment phenomenon. When this closure occurs together with the application of compressive stresses onto the crack region, the relative motion of the crack faces results resisted by a frictional force between them, in some cases producing interlocking. This can be smoothed by the presence of lubricant penetration inside the crack. Kaneta et al. [54] stated that in the condition of fluid penetration inside the crack, the lubricant holds the crack faces apart by the pressure

exerted.

Also in this case the possible propagation modes are still the tensile growth produced by the fluid pressure or the shear mode favored by the effective lubrication operated on the crack faces.

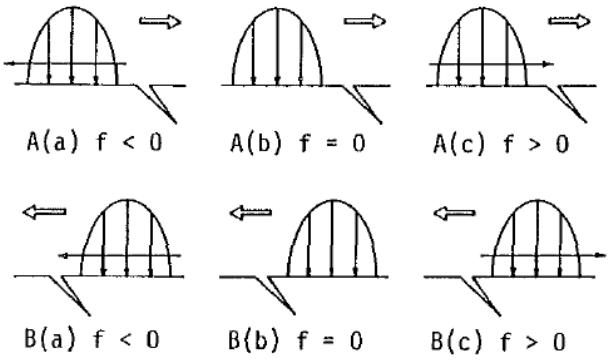


Figure 2.20: *Relation between relation of motion and fluid interaction with crack growth [54]*

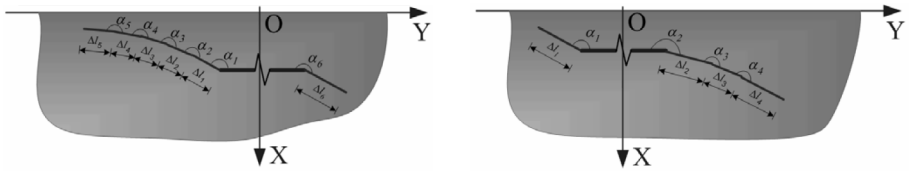
In Fig. 2.20 it is shown the relationship between the relative crack position, inclination and load movement [54]. The major possibility of oil seepage occurs in case $A(a)$, when crack axis and load motion are in the same direction, followed by the case depicted in $A(b)$ of pure rolling in the same configuration. When surface traction becomes low and crack opening increases ($A(c)$) the possibility of penetration is still present but depends on the contact distribution shape and crack geometry.

In these conditions, the Mode II stress intensity factor was evaluated by Kaneta et al. [54]. If no action of the fluid is assumed on the faces, the SIF associated with shear growth takes the maximum value at the apex point of the crack tip, at maximum depth. So, it can be concluded that shear growth occurs at the deepest point of the crack, generating an almost coplanar propagation along the original crack plane or even directed toward the core of the material. When the crack size becomes important instead, the possibility of tensile growth gains importance, always at the deepest crack tip. In this case, the propagation angle is much greater than 0 obtained for shear coplanar mode. The result, generated by seepage and the effect of hydraulic pressure, is the transition from inward Mode II growth to Mode I induced pitting.

Must be noted anyway, that Keer and Bryant [35] and Bower [56] stated that considering a model in which growth is due directly to the hydraulic pressure, while the crack mouth is open, is somehow unrealistic. The cause of such an assertion has to be found in the high intensity of the stresses in play. However, the analysis for all the models described was conducted on two-dimensional domains and it seems difficult to predict correctly the three-dimensional interactions at the surface and the crack tip, considering the presence of a fluid.

While the principles of lubricant interaction in crack growth may seem clear, the solution and simulation of this problem are very complex and articulated. To run a faithful simulation of the fluid pressurization effect, the real load profile generated by lubricant should be calculated. Actually, this depends on many factors as the rheological properties of the fluid itself and the modality of entrapment [61]. The viscosity of the fluid can therefore influence its penetration into the crack. The more a

fluid is viscous, the more difficult it will be penetrating, meaning that shorter distances in the crack length will be pressurized. In the case of incomplete pressurization, the "dry" component of Mode II will remain dominant [50].



(a) (b)

Figure 2.21: *Crack branching simulation under (a) unidirectional and (b) reciprocal loading [58]*

Based on these assumptions, Datsyshyn and coworkers [62][63] studied the possible trajectories of propagation of the edges of a crack, subjected to what is called 'boundary lubrication' conditions. A step-wise reconstruction was performed on the crack path by a curvilinear analysis. The same kind of analysis was conducted by Jin [58], whose numerical results show that the Mode II stress intensity factor diminishes and crack is driven in different directions. By Jin's analysis [58], the crack first propagates at the edge facing toward the rolling direction of the load, branching toward the surface. In that specific case of a one-direction passing load, 5 out of 6 newly formed branches were facing toward the surface, thus likely to produce pitting. But there was also the last generated branch which faced toward the core. Then a different loading path was implemented to verify its influence on the crack growth direction. This time the load went from the center of the crack toward the left tip and then back to the right. The results obtained were completely opposite, with the most branches facing toward the core.

This showed that the propagation pattern at each tip is different and controlled by the loading path. This is evidence of the competitive mechanism between external shear stress and an "internal" tensile stress generated by the presence of lubricating fluid.

Anyway, Murakami [64] first suggested that the presence of positive friction between the surfaces will lead to a closure of the crack mouth. This condition will stop lubricant penetration. Because of this, in many lubricant interaction models [50], the friction is defined as negative. Some attempts for the solution using FEM analysis have been carried out for example by Benuzzi [65], but a more complex and precise model is necessary to have a complete and simple understanding of the phenomenon.

Also, other authors such as Dubourg [66] take as acceptable the hypothesis of neglecting any trapped fluid effect when modeling a crack at certain lengths or depths. Which will mean the loss of the "surface effect".

2.6.4 Quiescent zone

The hypothesis of arrested surface cracks has been proposed by Tallian [67] and verified by Miller et al. [68], which introduces the presence of a "quiescent zone" between the surface layer and the subsurface region.

In the most external part of the material acts a local stress concentration field, near asperities, as it has been presented for the crack initiation phase. This field is dominant only for a few microns around the stress concentrating agent and loses effect when the crack branches toward the center of the material. For some load conditions, the surface-initiated cracks could not be able to propagate enough to reach the Hertzian dominated zone. We know that in the subsurface region the Hertzian contribution becomes dominant and acts as a driving force for crack initiation and propagation. The quiescent zone, in which micro-cracks are stopped, lies between these two stress dominated ones. Its extension depends on the amount and properties of asperities, on the contact condition, on the Hertzian stress profile and the shear fatigue threshold of the material (δK_{IIth}).

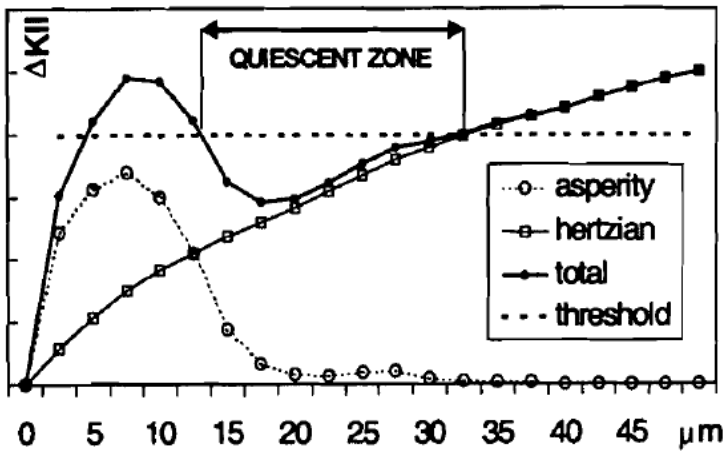


Figure 2.22: *Quiescent zone represented on the plot of ΔK_{II} as function of depth [50]*

If surface cracks are initially driven by surface friction and local stresses, they may reach the stable zone where they remain quiescent, without further propagation.

Experimental evidence showed that typically micro-cracks stop after $10 - 20\mu m$ in many different cases. This distance may vary depending on the conditions of contact, as displayed in numerous references [69][70][71]. Should be noted how this quantity is comparable to the usual characteristic microstructural dimension of hardened steels employed in REBs.

Further propagation it is possible when the Hertzian stress field is high enough to cover the quiescent zone and reach the maximum propagation depth of surface micro-cracks. In this case, a macro-surface fatigue phenomenon will occur. Usually, surface macro-cracks propagate in a coplanar way, under a mixed-mode state of stress.

2.6.5 Surface Cracks: Discussion

From experimental evidence [50] it can be stated that surface cracks can grow for tens of microns, in an almost coplanar way. This can be attributed to the presence of the lubricant effect enhancing the macro Hertzian stress action. This combination generates a mixed-mode propagation, in which Mode I and II will take relative importance depending on the dimension of the crack and the amount of fluid pressurization. For

shorter cracks, Mode II will remain prevalent due to a small amount of seepage. While for longer propagation both mechanisms are active.

Evidence from studies by Cheng [72] showed how surface fatigue phenomena are limited in pure rolling conditions, due to the lubricant effect. This has been confirmed also by studies conducted at the University of Brescia by Adamini et al. [73]. Tests on carburized and nitrided steels under Hertzian pressure were performed in lubricated and pure rolling conditions. Results showed that micro-pitting was formed in the first contact stages, but then was stopped or disappeared without macro-damage formation. Subsurface damage was instead formed later in the test.

It can be concluded, as stated by Ding [4], that despite the concurrent presence of Mode I and II mechanisms, the direction of propagation of the crack will follow the orientation of the maximum shear stress at the tip o curve toward the surface. Evaluating the initial inclination of a surface crack, the propagation angle could then be estimated. Always in the review by Ding [4] it is indicated that in the case of an initial angle smaller than 45° if Mode I is the dominant mechanism, the crack will tend to have a coplanar propagation or branch toward the surface. In any case, the result is likely to be the formation of a shallow pit. In the case of Mode II dominance instead, the growth direction will be parallel to the crack plane. Analytical and experimental works are reported to confirm that a surface crack alone will not be able to grow toward the core of the material [57][74][75]. In order to achieve deeper propagation, subsurface cracks' behavior has to be taken into account.

2.7 Subsurface Cracks

2.7.1 Subsurface Initiation

Subsurface effects of rolling contact fatigue can be different, depending on the properties of the material studied. In particular, hardness is a relevant parameter for the crack initiation mechanism selection. In softer materials, the plastic strain becomes the dominating mechanism, by accumulation in the zone of maximum stress. This leads to the progressive loss of resistance in the material and consequent crack initiation. In case of the presence of inherent stress raisers, as defects or inclusions, the over-stress is re-distributed by the plastic flow. In this way, the ductile materials can "absorb" more energy, but has much less resistance to constant high loads.

Hard materials, as used in rolling bearing elements instead, show better resistance to high loads and no plasticity. But hardness leads to an increased sensitivity to over-stresses and effects of inhomogeneities in the structure. The localized stress raising, coupled with the incapability of ductile behavior enhances crack initiation near inclusions. This is why in literature, subsurface cracks are usually initiated in stress concentration zones such as hard and brittle oxides or voids.

To describe this phenomenon, a diffused experimental procedure relies on drilling some micro-holes in the material, before undergoing a fatigue simulation. Bormetti et al. [50] used holes of a size around one hundred microns just beneath the surface of contact. To evaluate the proper position at which the holes have to be drilled, the driving loading condition, being the maximum Hertzian shear contribution, has been calibrated in depth. The holes then acted as the interface between the inclusion and the

matrix of the material, which acts as a weak spot and local stress intensifier. Results showed that hard materials are more sensitive to the presence of inherent defects, and sudden formation of cracks around the hole was verified at sufficiently high contact pressures (around 2000 MPa). In Fig. 2.23 a carburized steel with a surface hardness of 700 HV shown several cracks, initiated under pure rolling conditions.

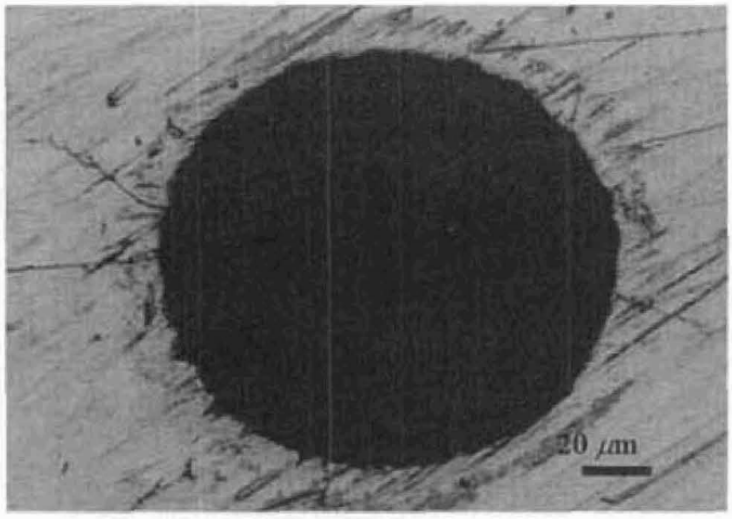


Figure 2.23: *Fatigue crack nucleation at the surface of drilled holes [50]*

In cases of multiple cracks in a limited region, the relative influence on neighboring ones can be taken into account. It was pointed out by Dubourg [66] that, in the case of a network of cracks, a mutual protection effect [76][77] takes place. The phenomenon is characterized by load sharing between the defects. So it can be argued that multiple cracks will undergo a reduction of growth rates due to a decrease in the effective stress intensity factor ranges that may lead to a general self-arrest. Thus single cracks and multiple systems behave differently in propagation. Usually, the failure phenomenons are caused by only one rare crack which can meet all the required critical conditions. The majority of the studies thus, focus on the behavior of only one crack.

2.7.2 Subsurface Propagation

It is well known that subsurface cracks propagation mainly occurs by shear. Maximum Hertzian shear stress is then in competition with hardness profile and compressive residual stresses to decide crack propagation.

As reported by Santus et al. [46], subsurface cracks are prevented to experience tensile Mode I SIF during contact. Due to the position and the nature of the interaction, the stress field will be purely compressive. So, the only propagation mechanisms remaining effective will be the shear Mode II (sometimes combined with Mode III [46]).

Evidence shows that cracks have a stable coplanar growth at shallow angles to the surface [78], same as if surface-initiated. Also, subsurface propagation usually tends to branch, after the crack has reached a critical length. The question of the predominant

growth mechanism has lifted some contrast in the literature. Studies by Fleming and Suh [38], Keer and Bryant [35], O'regan et al. [79], Kaneta [54], Blake and Cheng [80] and recently by Choi and Liu [81][82] supported the Mode II propagation theory through all the process. Ding et al. [4][83]–[85], instead proposed a distinction between two phases in the subsurface crack life.

Nucleation should occur at the presence of a point defect inside the material by shear stress. Further propagation instead is stated as no more strictly dominated by Mode II. It has been proposed a theory in which the growth is supported by a mechanism of linking between multiple cracks, followed by the collapse of ligaments.

This approach seems to be in contrast with the study conducted by Chen et al. [86] which confirms propagation and branching in Mode II. Different perspectives are present on the matter of RCF subsurface crack behavior. The multi-axial nature of the state of stress and the wide range of variables in play are an indicator of the complexity of the phenomenon. Santus et al. [46] state that, also in this case, a competitive situation has to be taken into account. All the mechanisms will be present at the same time, with different rates and relevance, depending on the specific set of conditions. The three main resulting effects on subsurface cracks will be [46]: multiple-cracks linking, Mode II propagation and branching. Anyway, to adopt a fail-safe approach, the most relevant stage will be the initial one, usually dominated by shear on the majority of the models. In the following sections, examples of different approaches and models to evaluate the SIF and interpret the growth behavior will be proposed.

Since the fatigue problem considered is a crack growth problem that can be associated with tribology theory, the use of fracture mechanics can be a valid method to analyze and evaluate the driving forces for failure. Many studies have assessed the possibility of crack propagation comparing the threshold stress intensity factors with the applied stresses. In the case of first propagation, the dimension of inclusions, from which cracks take part, are small. Then, the short-crack theory can be used to assess the phenomenon. According to this approach, defects can be treated as equivalent cracks. Murakami [87] developed a law to evaluate the Mode I fatigue threshold as a function of the crack dimension.

$$KI_{th} = 3.3 * 10^{-3}(HV + 120)(sqrtA)^{1/3} \tag{2.18}$$

where the Vickers hardness is indicated by HV and A [μm^2] is the projection of the defect area onto the plane orthogonal to the dominant stress for crack opening.

This dependence takes into account the difference between the opening and closure effect.

Since, for components subjected to rolling contact fatigue such as rails and bearings, Mode II crack growth is the most observed mechanism, the validity of the law has been checked.

In shear propagation the presence of friction can be hypothesized, thus making possible the "short crack" effect. Following this hypothesis, Donzella [88] proposed a model to estimate the Mode II stress intensity factor in the short-crack range, using the Murakami law. This consisted of imposing two bounding values to the estimation, the upper based on the assumption of equivalence between long and short crack behavior; the lower based on the dependence with crack size, as for Mode I.

The fitting obtained in this study is in agreement with measurements performed by many other studies, on hardened materials by Lunden [89], Hellier [90] and Sakae [91].

Influence of different parameters on the stress intensity factors has been analyzed using this model, such as fluid properties and rolling conditions. The study concluded that the Mode II fatigue threshold value, for crack propagation, is not significantly affected by the presence of friction on the contact surface, and though by the lubrication regime. This result has been confirmed by experimental evidence by Lunden [89].

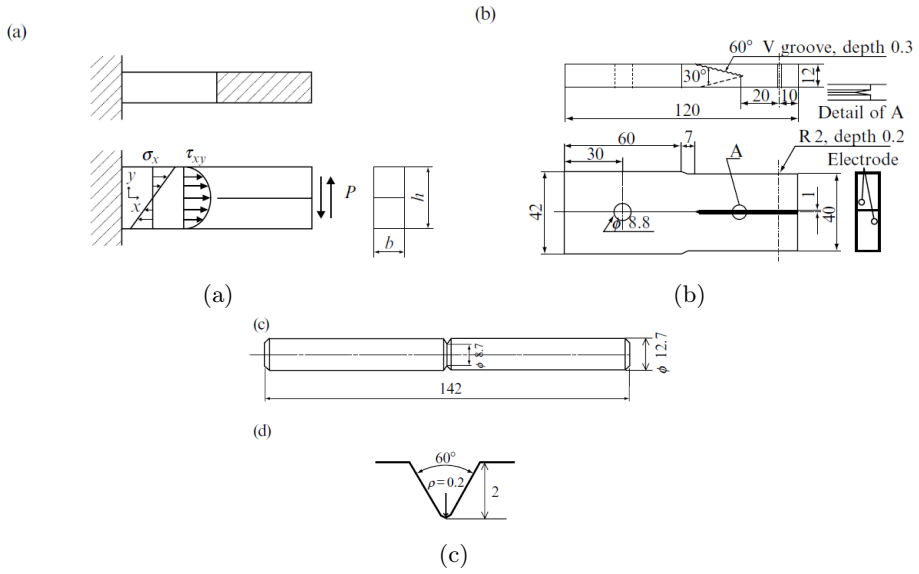


Figure 2.24: *Specimens for experimental tests on Mode II and II fatigue crack growth. Reported dimensions are in mm. (a) Basic principle for Mode II test. (b) Specimen for Mode II test. (c) Specimen for mode III test. (d) Notch detail [92]*

Once theoretical formulation was validated, the relative dominance of different growth mechanisms has been investigated to determine the main drivers. Even if laboratory studies on Mode II crack growth are more difficult than for usual Mode I fatigue experiments because in shear experiments the cracks tend to switch to Mode I during the test, Murakami et al. [93] conducted a test to determine threshold stress intensity factors. They concluded that for steels the stress intensity factor threshold of Mode II is higher than for Mode I. Murakami and Hamada [94] realized a specific specimen design to perform these fatigue tests. They designed a double cantilever (DC) specimen in which the shear stress will reach the maximum value in the neutral axis, where instead the direct stress will be zero. Fatigue crack growth was measured by the AC potential drop method.

Tested specimens were fractured using liquid hydrogen to perform the investigation on the crack surface. In Fig. 2.25 can be seen the Mode II fatigue crack growth morphology, with clear shear abrasion signs.

It is evident how the crack starts growing in Mode II but, as long as propagation continues, the stress intensification range decreases [94]. This is the cause of branching in Mode I that follows the first growth stage. Usually, branching occurs at steep angles, due to the different critical planes for shear and tensile stresses in rolling contact loads. Fig. 2.26 shows this clearly.

Branching angles are on average around $\pm$ 70.5°, the direction of maximum tan-

gential stress at the crack tip. It can also be seen how Mode I growth occurs when previous Mode II has arrested, and when the values of stress intensity factor ranges of the two mechanisms are similar and slightly inverted toward mode I dominance. ($dK_I = 1.15dK_{II}$). Also, Keer and Bryant [35] indicated that the presence of tensile stress may lead to branching and relative curvature of cracks, to form a cavity.

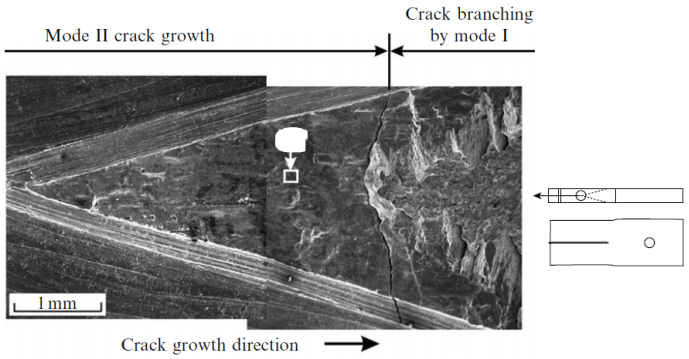


Figure 2.25: *Crack surface morphology [92]*

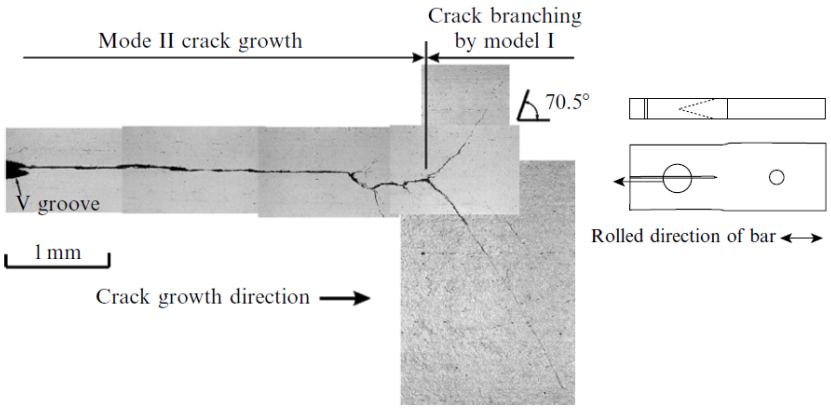


Figure 2.26: *Crack propagation and branching [92]*

Using fracture mechanics, Rosenfield [95] and Sin and Suh [96] among others, showed how subsurface propagation will occur parallel to the contact surface, mainly in Mode II. The rate will also increase with the size of the crack. Kaneta [54], modeled a pre-existing subsurface crack and confirmed that fatigue growth is driven by shear and can occur at both tips with different directions, as shown in Fig. 2.27.

This study proposed that in the case of pure rolling the crack tends to grow along its original plane at both tips. Instead, in the case of an inclined crack (about 5° [54]) the general tendency to parallel growth is conserved. The results mentioned above find validation in the experimental work of Yoshimura [97]. Dubourg et al. [66] conducted a study on steels under RCF conditions. Although it being more focused on railway steels, it also concludes that the major crack growth mode is shear dominated Mode II.

Moreover, also for Dubourg, this results in a coplanar extension along the initial

crack direction. In this specific case, branching is probable in Mode I conditions, at an inclination of 75°. Considerations upon lubrication conditions relatively to growth direction were made [66]. It is found that propagation toward the bulk of the material and across long distances is more likely to occur in mixed conditions. Not referred to the stress field this time but to lubrication. In particular, Dubuorg [66] states that a "compromise" between oil lubrication and dry conditions, which can be as water lubrication, may give way to long inward cracks.

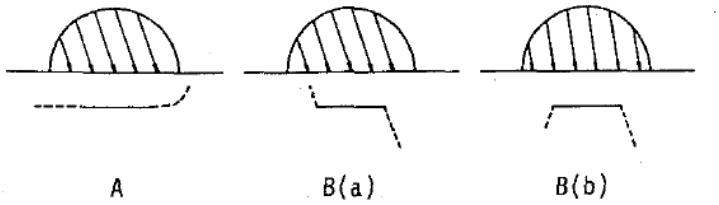


Figure 2.27: *Fatigue crack growth direction for (A) shear and (B) tensile Mode. (a) High traction. (b) Low traction. [98]*

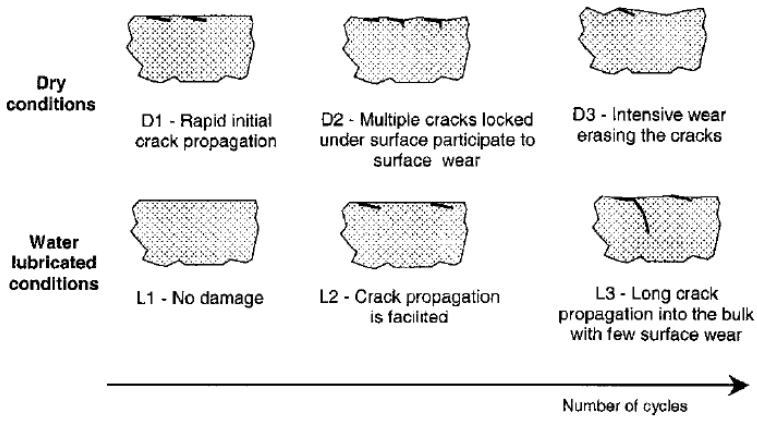


Fig. 20 Schematic representation of damage under dry conditions and water lubrication conditions under RCF

Figure 2.28: *Crack growth behavior under different lubrication conditions [66]*

This spurious condition may be caused also by an initial phenomenon of pitting or spalling. Although not being critical for the component's failure, it can increase friction between rolling surfaces and promote subsurface crack growth.

Recently Doquet and Bertolino [99] presented a study of crack growth orientation, based on the critical plane analysis. They presented the presence of a threshold value for Mode II SIF, proper of the material, over which bifurcation of cracks is prevented in favor of coplanar growth.

In this case, the propagation will occur at a length proportional to the stress intensity factor range of Mode II. Making use of kinetics theory is otherwise possible to have good evaluations even at high values of SIF. They concluded that when a mixed-mode loading condition is present, and Mode I and II are applied sequentially, coplanar crack growth results to be dominant. This occurs if the ratio dK_I/dK_{II}

is greater than a threshold value, identified at 0.5 by Wong et al. [100] or at 0.25 by Doquet and Pommier [101]. In the case of overlap of the mode cycles instead, bifurcation of the crack will be favored, as showed before, enhancing the dominance of Mode I growth.

2.8 Mode III and RCF for Rail Components

Also, studies on multi-axial stresses considering mode III contribution has been carried on by some studies. Since mode III cracks can be generated in components that transmit torque such as shafts. This mechanism is therefore investigated by torsion fatigue tests on notched components with circumferential geometry [102]–[104]. It has been found that under high enough torsion, the fatigue surface results being flat and consequence of a shear mechanism. Otherwise, if stress is low, Mode I results favored and produces a 'factory-roof' morphology. As stated by Murakami et al. [92], the transition process between a shear crack propagation and a factory-roof is not clear, but it has been reported that for both modes II and III, the growth induces a micro-structural modification on the crack surface.

Ritchie et al. [105] proposed a micro-mechanical model for fatigue crack growth driven by Mode III. The crack propagation was considered to occur by coalescence of micro-cracks in Mode II, ahead of the main crack front at the presence of inclusions. The same team [106], found that there was no direct correlation between the crack growth rate and the stress intensity factor of Mode III, except in the case of superimposition of a static Mode I loading. Also Houlier and Pineau [107], showed that this superposition can cause a modification in growth rate and also in the crack path during propagation. Akama et al. [78] concluded that Mode II and Mode III shear growth mechanisms are "essentially the same". From experimental analysis on mixed-mode crack propagation, fracture surfaces obtained under I/III loading and I/II modes are very similar [108]. In both cases, the shear stress intensity range resulted as the main driving force, without a great difference between the in-plane or out-of-plane shear mechanism. This view is confirmed also by Murakami et al. [92], who states that if the applied stress intensity factor range is high enough, then the crack growth should occur in shear mode both for torsion fatigue tests and in Mode II growth tests.

All these models referred to the rail-wheel contact in case of train application, involving high friction and traction. Anyway is interesting to notice the contribution and effect of mode III superimposition in the case of rolling contacts, which can occur in a relative measure also in bearings.

2.9 Residual stresses

As we have seen in the previous chapters, to improve the fatigue life of metals, hardening is performed. This process involved the production of residual stresses inside the material, which interact with the state of stress, limiting the damaging effect. Different sources of residual stress induction are available: heat treatment, machining (in different manners as rolling, shot peening, grinding, etc.) and carburizing or nitriding.

From a rough and general point of view, the presence of a residual stress field is stated to increase or decrease the maximum shearing stress, according to a simple law [109]:

$$(\tau_{max})_r = \tau_{max} - \frac{1}{2}(\pm S_r) \tag{2.19}$$

Where S_r is the residual stress.

Should be noted the presence of the $\pm$ sign. In case of compressive residual stress (negative) the resulting shearing stress will be reduced (sing of τ_{max} is negative). Also, Dubourg [66] indicates that positive values of residual stress, meaning a tensile field, increase the crack growth rate. The cracks will grow faster and are more prone to branching due to the opening of the crack. This opening is maximized at a perpendicular inclination to the surface, thus the driving direction for Mode I branching. Otherwise, a compressive field, identified with a negative sign, reduces the opening of the crack and the sliding between the faces, reducing the growth rates [66]. The fatigue life is proportional to the critical shearing stress:

$$Life \propto [\frac{1}{(\tau_{max})_r}]^9 \tag{2.20}$$

Anyway, more recently a different general idea of elastic superimposition between the stress tensor and the residual stresses has gained validity [16]. It derives also from the assumption that the most critical shear contribution to fatigue is the maximum shearing stress, which results from the evaluation of the principal stresses. This approach will be taken into account in the numerical model section.

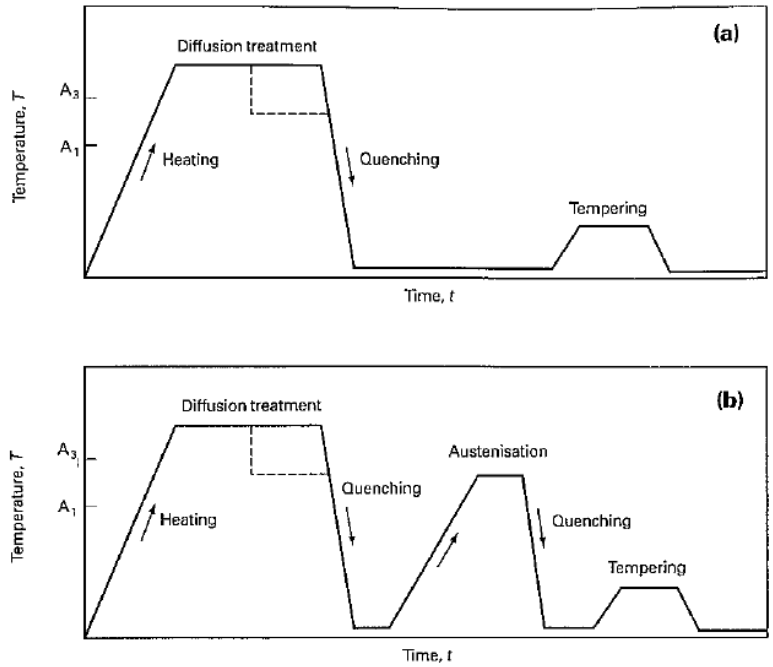


Figure 2.29: *Heat treatment cycles for hardening carburised components: (a) direct hardening, (b) single hardening [110]*

However, the specific profile of the residual stress induced by general processes, are difficult to predict correctly. This is caused by the self-equilibrating property of residual stresses, which can modify the distance and intensity reached. Multiple factors have an influential role in this phenomenon. In this work, only a generic distribution attributed to carburizing and nitriding processes will be described.

Usually, mechanical processes such as machining, coating, shot peening, etc. [6] induce short-range residual stresses. Otherwise, thermo-chemical processes such as carburizing and nitriding can produce long-range effects. Carburizing, for example, generates a compressive state of stress in the treated material. This effect is due to the carbon atoms which, taking place inside the metal lattice, generate and expansion of the neighboring cells, inducing compression.

During carburizing the alloy is exposed to high temperatures, in the presence of carbon-bearing material or a carbon-rich atmosphere (Fig. 2.29). The material will then absorb part of the carbon while heated, which will induce the formation of an austenitic hard structure in the carburized layer during the quenching phase. The following process of austenitization and tempering can be performed, depending on the desired results [110]. Tempering may be required for mechanical requirements, but has a negative effect on the level of residual stresses, which are lowered [111].

To counteract crack formation and the subsequent propagation, a compressive stress field is desirable in the most critical region. Unlike the stresses generated by rolling contact, the residual stress field is a static state of stress, function of the depth. The presence of a compressive stress field is not directly responsible for crack initiation, which instead is used to hinder, but can influence growth orientation and behavior.

The near-surface region, called the carburized layer, shows a higher amount of compressive stress to counteract the critical stresses occurring there. The profile presented (Fig. 3.37) is obtained by in-depth X-ray diffraction experimental analysis on the outer race of the bearing gear [7]. Also in the model from Depouhon, an approximation of the material to a semi-infinite plate. The profile is then matched with an eigenstrain, purely hydrostatic and localized state, shown with a red line in Fig. 3.37. It is also shown that the core of the gear is subjected to a tensile stress state, responsible for a mode I crack propagation in the critical failure stage (Fig. 2.30). This double nature is due to the failed penetration of carbon in the inner regions of the component. The shrinkage and phase transformations occurring during quenching are then responsible for the double nature of the stress field across the section.

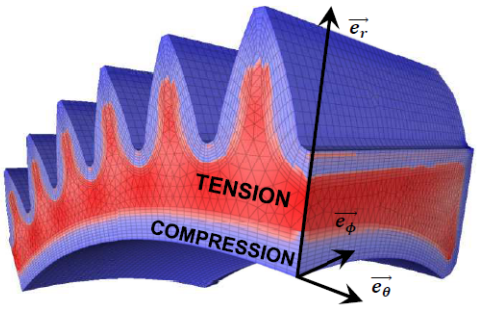


Figure 2.30: *Rendering of the residual stresses field across the whole component [7]*

A similar distribution was already shown from Bhadeshia [6], in the state of the art, where the depth of the treatment was indicated to be around 1 mm [6][109]. This

will be the total affected zone or carburized layer, that exceeds by almost an order of magnitude the dimensions found for the critical stresses. Moreover, the exact profile and entity of the residual stress field are difficult to identify, because are a specific characteristic of the component, the material and the application.

Should be reminded that the formation of carbides may act as a stress raiser during the cyclic loading. Thus being the generating point for crack initiation. It is clear then that this process should be precisely designed to achieve the desired mechanical properties, without compromising the metallurgical quality of the material. It has to be mentioned that, in a new bearing, a compressive stress field is present close to the surface [6]. This is due to the finishing of the component. A generic residual stress distribution has been reported by Bhadeshia. (Fig. 2.31)

In the previous paragraph about the Hertzian stresses (Chap. 2.3), it is shown that the critical shear stress is not located near the surface. In Fig. 2.31 is also evident how the peak of the residual stresses is located beneath the surface and not immediately close to it. From the fatigue studies reported before, we've seen that the shear stress is the most critical factor for crack propagation. It should not be strange that a similar depth profile results to be the most suited for fatigue prevention. The sum of the contributions from finishing and thermo-chemical treatment will give a distribution with an initial high value at the surface, a leveled out zone beneath it and then a gradual decrease toward the core [109]. The compressive surface layer can exert the effect of crack closer [50], preventing lubricant penetration. This phenomenon limits Mode I importance onto the crack growth, favoring shear dominance. The consequence is a more stable coplanar growth possibility, driven by shear Mode II.

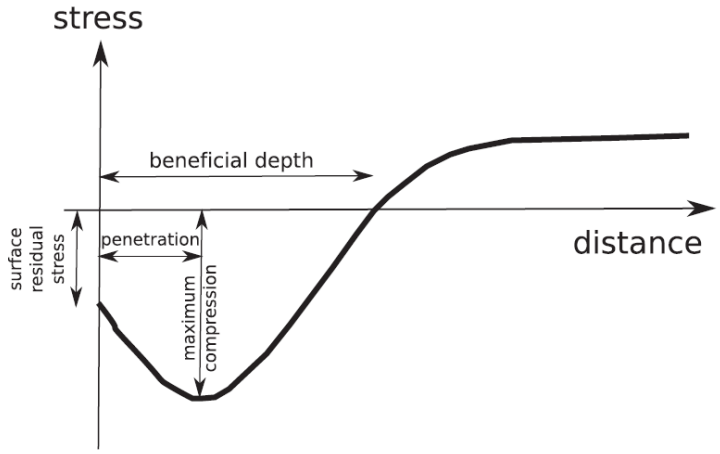


Figure 2.31: *Generic residual stress depth distribution [6]*

2.10 Simulation Models

Rolling contact fatigue has been studied scientifically since the 1940s and over the years many approaches were adopted [32]. Different models have been proposed to give a realistic prediction of the fatigue life of bearing components and Sadeghi et al.

[32] have proposed a review of the most influential ones.

A general distinction is made between probabilistic engineering models and deterministic research models. The first group identifies the empirical ones, concerning extensive testing to derive the needed variables. By this approach, the detailed constitutive behavior of the materials or the influence of specific parameters is not analyzed deeply. But the experimental result is s direct consequence of the chosen set. On the other hand, the deterministic models are based on theoretical assumptions and rely directly on the constitutive relations of the materials. This basis is used together with failure models, to model the fatigue behavior.

Before highlighting the most influential models and being more specific, some considerations must be made. The majority of the theoretical research models proposed in literature give a homogeneous representation of the materials considered.

In this way, the contributions of defects such as inclusions or the structural behavior, are neglected. This can lead to mistakes in life evaluation because of the important role assumed by these factors as stress raisers and crack initiation spots. On the other hand, a micro-mechanical approach is rarely considered due to the complexity of the task and the intrinsic variability of metallurgical composition and modification during production processes, as previously reported. Although the cracks seems to start often at weak spots, deterministic models can't by their nature consider each specific case. Experimental evidence shows clearly that material properties under RCF follow a Weibull distribution. This means a non-uniform behavior, intrinsic also in classical fatigue, that should always be considered.

Otherwise, experimental models can reproduce the scattered results of fatigue. But here a great conflict is based on the critical stress that has to be considered as dominant. Different depths have been found as initiation spots from evidence [86]. Different studies proposed the orthogonal shear stress as fatigue criteria, as Lundberg and Palmgren [8], but then also the maximum shear stress [112], the von Mises equivalent stress [9], and the octahedral shear stress [113] has been considered. Each of them reaches its peak at different depths, as observed. Thus it depends on the case under study and set of parameters which one will be dominant in the specific experiment. An exact dominant stress definition for crack initiation has not been stated yet [32]. It must also be mentioned that the empirical approach for calibration of engineering fatigue life is becoming more and more relevant by means of time needed and costs.

Although having a more analytical nature, this kind of model is usually limited to a specific aspect of the general phenomenon. This is way in the review on crack behavior, a wide range of studies authors will be referenced for different particular cases of the general topic. Sometimes also in contradiction with one another due to the influence of the many operational factors involved in RCF.

Table 2.1: Chronological list of probabilistic bearing life prediction models [32]

Year	Researchers	Model Description
1945	Palmgren	Empirical [114].
1947	Lundberg and Palmgren	First probabilistic bearing life model, termed the LP model, based on probability of crack initiation at a subsurface depth (z_0) where orthogonal shear stress (τ_0) is maximum in stressed volume (V) in the contact, expresses the probability of survival (S) after N RCF cycles as in Eq. 2.21 [8].
1952	Lundberg and Palmgren	The L_{10} life is expressed as in Eq. 2.22 [115].
1971	Chiu et al.	Probabilistic model that attributes spalling to intrinsic material defects and includes influence of material elastic and plastic properties, defect type, concentration, and geometry on the contact stress field (Eq. 2.23) [10].
1985	Ioannides and Harris	Improves on the LP model by: (i) assuming discrete material volumes, each with its own probability of survival, and overall risk obtained by integration; and (ii) introducing a stress threshold (σ_u) below which no failure is possible (Eq.2.24) [9].
1986	Schlicht et al.	Model assumes that fatigue failures originate at the surface due to von Mises stress, and plastic flow and residual stress development due to RCF influence bearing life. The life model is given by $L = a_1 a_{23} f_t (C/P)^p$ where a_1 and a_{23} are life modifying factors based on reliability, material, and operating conditions [116].
1987	Shao et al.	The development of cracks due to RCF in case-hardened bearing steels was described by a three-stage process: incubation, stable propagation, and branching/propagation [117].
1988	Leng at al.	The three stages described by Shao et al. [46 sad] were allocated 13%, 56% and 31% contributions to life [118].
1989	Current ISO Standard	The life relation is given by $L = a_1 a_2 a_3 (C/P)^p$, where a_i are life modifying factors [119].

Table 2.1 – continued from previous page

1992	Tallian	Probabilistic life model uses orthogonal shear stress field solution in conjunction with a Paris-law exponent (ψ) and growth parameters (n_0), material fatigue susceptibility (ϕ_0), and defect (ϕ_2) parameters [11].
1994	Zaretsky	Life equation is $ln(1/S) = N^e \tau^{ce} V$ and is similar to LP model, with two modifications: (i) dependence of the stress-life relation on the Weibull slope e was eliminated, and (ii) dependence on the depth term was eliminated. τ^{ce} is the maximum shear stress and not orthogonal shear stress [120].
1996	Harris and McCool	Statistical comparison of actual and computed bearing lives from LP and Ioannides and Harris's [9] models showed wide dispersion [121].
1996	Otsuka et al.	RCF testing showed that subsurface crack growth behavior was controlled by mode II stress intensity range, ΔK_{II} [122].
1999	Ioannides et al.	Modification of load-life relation from the 1985 model [9] to give $L_{10} = (A/[1 - (\frac{P_u}{P})^w]^{c/e})(\frac{C}{P})^P$, $P > P_u$, where P_u is the load corresponding to σ_u [123].
2000	Kudish and Burris	Model improved by Tallian [11] by including the effects of contact pressure and size, friction coefficient, residual stresses, initial defect, size, location and orientation distributions, material fracture toughness, material hardness variation with depth, and material fatigue parameters as function of hardness [124].
2002	Shimizu	A three-parameter Weibull life distribution function was proposed, after showing that bearing steels subject to RCF do not exhibit a fatigue endurance limit, given by: $L - \gamma = (C/F)^p (lnR/ln0.9)^{1/m}$, where γ is the minimum life prior to failure [125].
2003	Miyashita et al.	The location of $\Delta\tau_{xy}$ estimated by FEA was shown to coincide with the depth of crack initiation observed in RCF experiments using sintered alloys [126].

Table 2.1 – continued from previous page

2005	Kotzalas	A three-parameter Weibull distribution was shown to be able to predict finite life in the high reliability regime, based on statistical distribution of tapered roller bearings [127].

The initial studies had to deal with the inability to describe RCF with the classical fatigue theories. Accordingly, early works are based on empirical observations and the resulting mathematical interpretation [114]. The first probabilistic model was proposed by Lundberg and Palmgren in 1947 [8][115]. They assumed that cracks were initiated at a subsurface weak spot and driven by orthogonal shear stress. The weak spots, usually inclusions, were modeled by implementing a stochastic distribution with the Weibull statistical strength theory. By this model, under a pure Hertzian stress field, a probabilistic evaluation of fatigue life was calculated. The initiation step was taken as dominant.

$$ln(1/S) = A(N^e \tau_0^c V/z_0^h) \tag{2.21}$$

where A, c, and h are experimentally determined material parameters and e is the Weibull slope for the experimental life data S is the probability of survival of the component, after N cycles. z_0 is the depth at which τ_0 , orthogonal shear stress, occurs. V is the stressed volume and the other variables are empirically derived parameters.

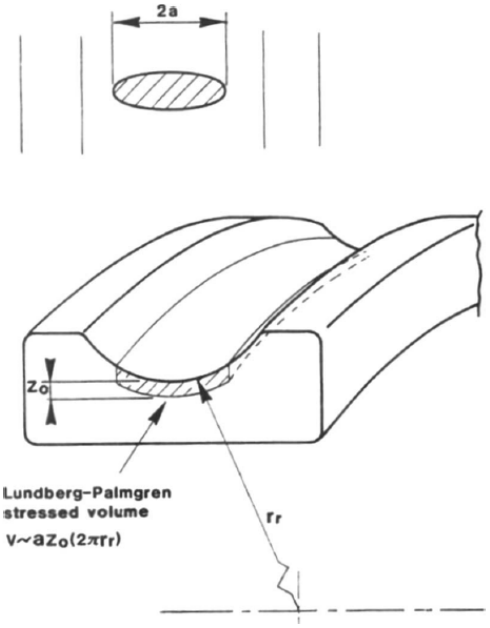


Figure 2.32: *Stressed volume configuration in Lundberg-Palmgren model* $V \sim az_0(2\pi f_r)$ *[32]*

From this equation, a load-life proportionality was derived:

$$L_{10} = (\frac{C}{P})^P \tag{2.22}$$

where C is the bearing basic dynamic load rating, P is the equivalent load acting on the bearing, and exponent p depends on contact shape. The factor L_{10} identifies a 10% probability of failure in the given conditions. The exponent varies depending on the geometry of the contact and components, and it is derived empirically. This equation is also the basis for the standards used today in the industrial production of bearings (ISO 281 [119]). It is clear how the initial approach by Lundberg and Palmgren has been used and survived for many years, but it suffers from many limitations if analyzed in detail.

For example, lubrication interaction is completely neglected and also traction on the surface is not considered in the pure Hertzian load. Some amount of sliding and friction is always present in rolling contacts, and it affects the resulting stresses, moving the peak toward the surface. Also, the presence of any sort of roughness deviates from the ideal smooth assumption. Cheng et al. [10] proposed a statistical model based on crack propagation, taking into account the elastoplastic properties of the material's matrix and the physical nature of the defects. The resulting law was an evolution of the previous survival equation:

$$ln\frac{1}{S} = (\sum_{i=0}^{4} \phi_i) N^{\beta} \tag{2.23}$$

In this formulation, the empirical parameter is ϕ_i, which depends on the configuration (material, geometry, load, defects). Subsequently, in 1985, Ioannides and Harris [9] proposed the presence of a threshold in the stress, similar to the classical fatigue limit, below which the material was to consider safe from any failure. They implemented a discrete model, in which the component was divided in small volumes, each with a different survival probability. An integration over all the discrete elements was then performed to evaluate the overall behavior.

$$ln\frac{1}{S} = AN^e \int_V \frac{(\sigma - \sigma_u)^C}{z^h} dV, \sigma > \sigma_u \tag{2.24}$$

where σ_u is the stress threshold and A is an empirical constant. Another novelty introduced by Ioannides and Harris regarded the critical stress component considered. They didn't limit the analysis to the orthogonal shear stress but considered also the equivalent von Mises and maximum shear stresses as potential candidates. The resulting life-prediction equation from this approach will be:

$$L_{10} = \frac{A}{[1 - (\frac{P_u}{P})^w]} (\frac{C}{P})^P, P > P_u \tag{2.25}$$

where P_u is corresponding to the previous threshold stress. This equation was proposed by the same Ioannides in 1999 [123]. A comparison between this and the Lundberg-Palmgren law is shown in Fig. 2.33. The former prediction is clearly more conservative and behaves linearly, because of the less volume considered. Instead, the Ioannides-Harris curve has an asymptotic, similar to classical fatigue theory.

Many developments have been made on the basic laws that have just been introduced. But the original idea was conserved. The proof is that the ISO standard

[119] to evaluate the life of rolling bearings is a modification of the original Lundberg-Palmgren equation:

$$L = a_1 a_2 a_3 (\frac{C}{P})^P \tag{2.26}$$

where the a_i constants are modifying factors depending on the set considered of material, component and operating conditions.

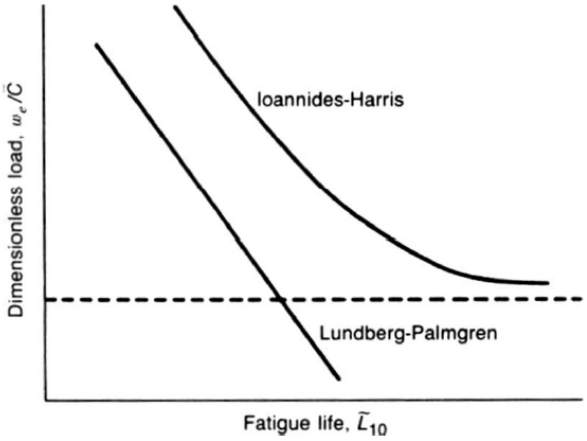


Fig. 5 Comparison between the Lundberg–Palmgren [33] and Ioannides–Harris theories [37]

Figure 2.33: *Plot comparison between Lundberg-Palmgren and Ioannides-Harris theories [32]*

As it can be seen in Tab. 2.1, many different models were proposed, integrating the previous models with new details, or relying on particular assumptions on initiation site, the influence of parameters and relation with statistical distributions.

Some studies, such as the one from Kudish and Burris [124] proposed an improvement to previous models (Tallian [11]) considering the contribution of operational parameters. But also in this case some assumptions were made to give validity to the model. The Kudish-Burris model, for example, considers the propagation stage as dominant in fatigue life, neglecting the nucleation phase. But, as stated initially, these models are all derived from empirical evidence and thus relative to the specific assumptions made on top of the tests. So a clear physical explanation of RCF behavior cannot be taken from this model. For this purpose, the research models are a more suited way.

The deterministic models are a more theoretical approach, based on the constitutive principles of the materials studied and failure mechanics. They also have to make assumptions on the importance of initiation or propagation phases during fatigue life, due to the different processes involved. Few models take into account both. In Tab. 2.2 is reported a historical overview of this research models, by year. Keer and Bryant [35] were the first to propose a model based on 2D fracture mechanics. They took crack propagation as the dominant phase in bearing fatigue life. The resulting equation relates the number of cycles for fatigue to the maximum Hertzian pressure by this formula:

$$N = \frac{b^{1-m/2}}{\beta_0 p_{max}^m} \tag{2.27}$$

where p_{max} is maximum Hertz pressure, b is contact half-width, and β_0 and m are crack growth parameters. Anyway, this initial model was found to be largely underestimating fatigue life by orders of magnitude with respect to the engineering models.

It goes with the theoretical nature of these approaches, that many studies differ from one another by their basic assumptions or crack modeling techniques. Bhargava et al. [128] considered plastic strain accumulation as the critical parameter for spalling. Cheng et al. [72][129] instead took micro-mechanical theory to develop a dislocation pileup model for crack initiation. The same basic idea was taken by Vincent et al. [130] whose then added the contribution of residuals stresses to the model. Many developments were made depending on the particular side of the phenomenon to be analyzed and a summary can be found in Tab. 2.2.

Zhou et al. [113][131] otherwise considered both initiation and propagation phases in their formulation, expressing the total life as:

$$N = \frac{AW_c}{(\Delta\sigma - 2\sigma_k)^2 D} + \int_{a_i}^{a_f} \frac{da}{c\Delta K^n} \tag{2.28}$$

where among the usual material parameters (A,c,n), the damage accumulation variable D is introduced. This comes from continuum damage mechanics integration in the RCF life study. This is the approach that will be considered for the numerical part of this work.

Table 2.2: Chronological list of deterministic bearing life prediction models [32]

Year	**Researchers**	**Model Description**
1983	Keer and Bryant	First deterministic analysis of RCF life. A 2D fracture mechanics approach is used for life estimation in rolling/sliding Hertzian cylindrical contacts assuming initiation life is small compared with propagation life. Eq. 2.27. Life computed greatly under predicted compared with LP-based models [35].
1989	Zhou et al.	Model included both crack initiation and propagation lives. The total life is related to the specific fracture energy per unit area, material friction stress, the damage accumulation factor, and the stress intensity factor range at the crack tip [131].
1990	Bhargava et al.	Model based on plastic strain accumulation in strain hardening materials under cyclic RCF [128].
1992	Sehitoglu and Jiang	Multiaxial fatigue crack initiation model for RCF [132].

Table 2.2 – continued from previous page

1994	Cheng et al.	Micro-mechanical model based on dislocation dynamics (pileup). Crack nucleation was assumed to take place in slip bands at the grain level. Initiation life is related to the critical shear stress amplitude and its threshold [129].
1997	Melander	A FEA study of cracks subjected to RCF including crack face friction due to closure [133].
1998	Vincent et al.	Crack initiation model based on dislocation pileup and accounted for full stress tensor and residual stress field. Dislocation emissions were assumed to occur due to the presence of subsurface inclusions that lead to stress concentration and localized slip/plasticity [130].
1998	Lormand et al.	Extension of model by Vincent et al. [130] to include crack propagation via Paris law, driven by Mode II loading. A Coulomb stress was included to account for crack face friction due to closure [134].
1999	Harris and Yu	The inclusion of surface traction along with Hertzian normal pressure was shown to significantly increase subsurface octahedral shear stress (von Mises stress), but not the maximum stress range [135].
1999	Jiang and Sehitoglu	Elastic-plastic FEA that included effects of ratcheting under RCF in conjunction with a multiaxial fatigue damage criterion was used to compute initiation life for line contacts. Maximum damage corresponded to depth where orthogonal shear stress range was maximum, in accordance with LP theory [136].
2001	Ringsberg	Elastic-plastic FEA, multiaxial fatigue crack initiation model based on a critical plane approach, and fatigue damage accumulation concepts were used to develop a procedure for life prediction under RCF loading [137].
2006	Liu et al.	Multiaxial fatigue damage due to RCF using the critical plane approach, with applications to railway wheel contact fatigue [138].
2007	Liu and Mahadevan	A unified multiaxial fatigue damage model for RCF using the critical plane approach for isotropic and anisotropic materials [139].

2.10.1 Continuum Damage Mechanics

This approach was proposed by Kachanov in 1958 [140]. As reported by Bhattacharya [141], "continuum damage mechanics [...] deals with the distribution, characterization and growth of microstructural defects in terms of macroscopic state variables" [142], [143]. The basic concept of this theory is the evolution of damage inside the material, with relative loss of integrity. A damage variable is then evaluated as the density of defects: cracks in the specific case or voids more generally. This variable is evaluated over a small volume and usually a tensor.

Anyway, "if the weighted fractional loss in cross-sectional area is the same in every orientation within the material" [141], then the value of the damage variable is independent of the orientation and can be considered isotropic. In this case, it becomes a scalar variable, identifying the degree of degradation of a single volume element. Its value can range from zero, undamaged element, to one or a maximum critical value (D_c) proper of each material. At the critical value, the section is considered to lose continuity. What is considered as a failure by this approach doesn't necessarily mean the complete fracture of the component, but has only local validity.

Lemaitre [144] introduced thermodynamic potential functions in the theory, later developed by Bhattacharya and Ellingwood [141][145]. The consideration of dissipative and irreversible accumulation of damage was united to the laws of thermodynamics in order to evaluate the time evolution of variable D as a non-decreasing function of time. The Helmholtz free energy is considered to evaluate the maximum work in an iso-thermal process on a deformable body. For a reversible process, this energy is stationary and its variation is expressed by:

$$\delta\Psi(t_2) = \delta\Psi(t_1) + \delta\int_{t_1}^{t_2}(\dot{W} - \dot{K_E})dt - \delta\int_{t_1}^{t_2}\Gamma dt \tag{2.29}$$

If an initial thermodynamic equilibrium is assumed at time t_1, the variation of Eq. 2.29 can be expressed by:

$$\delta\Psi(t_2) = \int_{t_1}^{t_2}\delta I_1(t)dt - \int_{t_1}^{t_2}\delta I_2(t)dt \tag{2.30}$$

Applying the commutative property of integration and considering the term $\delta\Psi(t_1)$ to vanish. This is possible if damage occurs through a series of equilibrium states, a condition that has been validated for ductile damage under load [146]. In this way, it can be demonstrated [141] that the second term in Eq. 2.30 vanishes. A set of coupled partial differential equations results from the former formulation:

$$T_i + \psi_D\frac{\partial D}{\partial\epsilon_{ij}}n_j = 0 \tag{2.31}$$

In the conditions of multiaxial stress and strain, characteristic of rolling contacts, the solution of this equation results difficult. But after a precise study of the critical mechanisms governing the case under study, the uniaxial solution can be used:

$$\frac{dD}{d\epsilon} = -\frac{\sigma_\infty}{\psi_D} \tag{2.32}$$

This solution is important not only because easy to use, but also because it is comparable with experimental fatigue data.

In this way, the dependency on the specific dimensions and parameters of the considered crack was avoided. The only material parameter to be derived experimentally is the critical damage value. This can be obtained by a simple static tension test [141]. A wide range of values has been found, from 0.15 to 0.85, depending on the material [147]. The concept of effective stress and the principle of strain equivalence are used for this purpose.

$$\tilde{\sigma} = \frac{\sigma}{1 - D} \tag{2.33}$$

$$\tilde{E} = E(1 - D) \tag{2.34}$$

Where the stress distribution and the constitutive law are affected by the presence of damage. σ is the nominal stress and E is the undamaged Young's modulus. In this way, applying gauges around the damaged zone to measure the change in stiffness of the material, the variable D can be derived empirically.

3 Analytical-numerical model in Matlab: multiaxial stress state and fatigue life simulation

3.1 Model Introduction

In this section, a numerical model for the prediction of fatigue initiation life in rolling contacts is presented. A theoretical approach starting from Hertzian theory and continuum damage mechanics [16][5] has been taken.

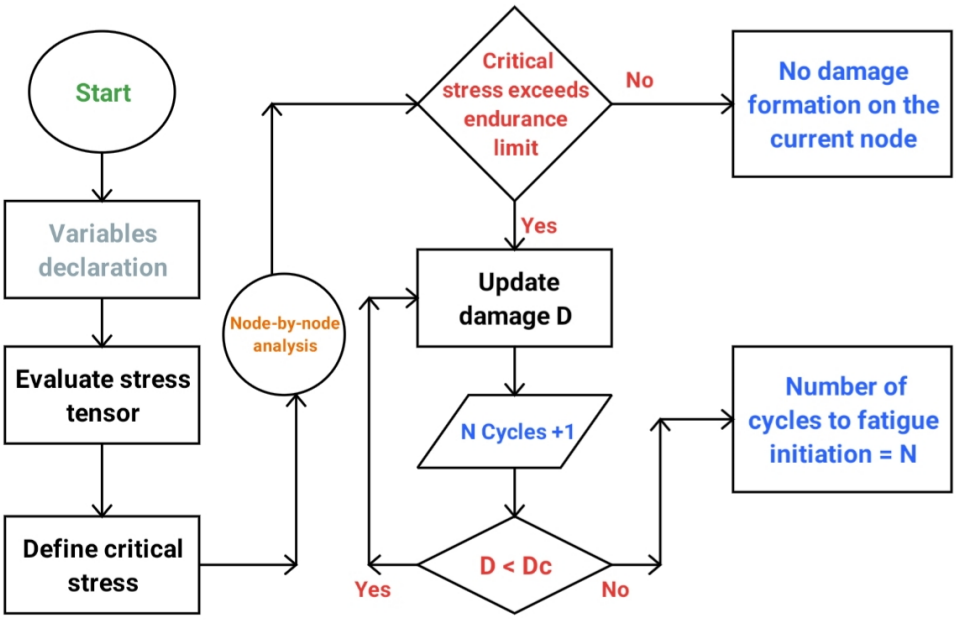


Figure 3.1: *Flowchart of numerical model passages for damage evolution and fatigue initiation life evaluation*

The evaluation of the stress tensor has been performed following the general Hertzian theory of contacts. The tensor evaluated is then independent of the properties of the material. The only parameters to give contribution are the maximum pressure, the contact dimension (half-width) and the friction coefficient. The values obtained have been compared with literature references on similar models [16].

Material's property take a role in the damage evolution evaluation. Each material has a different threshold for plastic deformation and can withstand different values of stress. The different thresholds act as a filter for damage generation, resulting in fatigue initiation to happen at different stresses for different materials. Other parameters, as the cyclic hardening modulus and exponent or the true failure stress, which act on damage evolution per each cycle, are also material dependent and influence the amplitude of each damage increment.

From this process, the final result is the number of cycles needed to reach a critical

value for damage. The critical damage is a threshold for material loss of continuity. It can be argued then, that once this value is reached, a crack starts to form.

Once the number of cycles needed for crack initiation has been evaluated, fatigue life can be then consequently derived by knowing the relationship between initiation and total life in the specific material and application.

A flowchart of the process passages followed in the model is shown in Fig. 3.1.

3.2 Non-conforming Contact

The contact between two bodies can be identified as *conforming* or *non-conforming*. In the case of an exact fitting between the two surfaces, the contact is defined as conforming. Examples are flat sliders and journal bearings. On the other hand, when the surfaces in contact have dissimilar profiles, the contact is defined as non-conforming. This alternative is verified in many cases, where either *line* or *point* contacts are concerned. The main examples of point contacts are ball bearings, whereas line contacts are usually generated by roller bearings.

Another case giving rise to line contact is when the surfaces are conforming in only one direction. This has been the case under exam in the present work. The roller bearings' profile is slightly curved (Fig. 3.2) and the contact result to be elliptical.

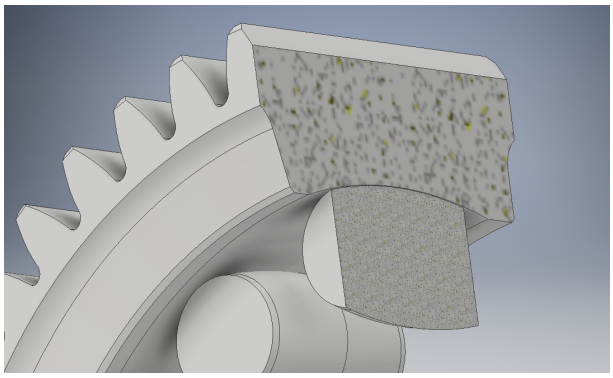


Figure 3.2: *Section of the CAD model indicating the contact region geometry*

A precise definition of a flat contact area may be difficult in non-conforming contacts. But, as in the case examined, this region lies approximately in a plane and the presence of slight curvature and warping are neglected. Only in highly warped surfaces, such as deep groove ball bearings, the contact area deviates consistently out of the plane defined by the gear's race surface [5]. Hence shear traction gives a significant contribution to the normal equilibrium.

In the case object of this model, the relative curvature of bearings and gear race is such to admit a flat surface approximation. Moreover, the contact happens over an area which is sensibly smaller than the radii of curvature of the bodies. This is verified when deformation is small and stresses are concentrated in the area of the contact region. In these conditions, the stresses have a rapid decay far from the point of interaction and thus, the only region to be analyzed is close to the contact.

From these assumptions on dimensions of the bodies and of the contact region, it's acceptable to approximate them to semi-infinite elastic solids. The contact then occurs on a plane surface, represented by an elastic two-dimensional half-space.

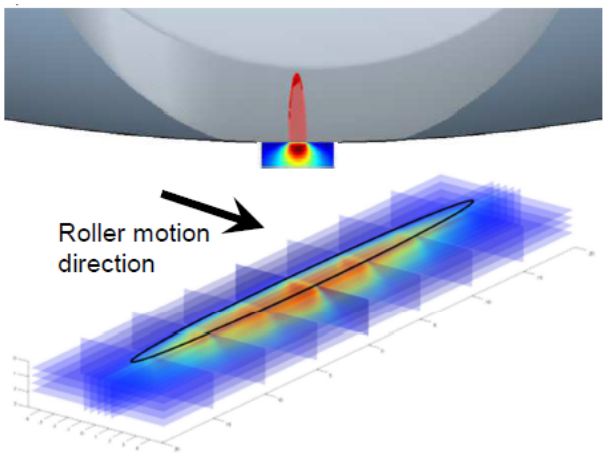


Figure 3.3: *Hertzian stress field under elliptical contact [7]*

It can be seen that the elliptical region has an elongate configuration that makes it comparable with a line contact symmetry.

Thus, only the central, most critical section is going to be considered in this numerical model. An approximation to a non-conforming line contact and reference to circular symmetry has been used, depending on the specific case.

3.3 From 2D to 1D Domain

Usually rolling contact analysis is modeled as a 2D problem [16][35][56][137][141][148], because of its intrinsic nature, consisting of a linear motion of a body over a surface. This is well reported in FEM and numerical problems as the one by Beheshti et al. [16].

In this article, taken by reference for the present model, the race on which rollers are moving is modeled as a 2D space. The load is therefore moving on the top of the domain and the stresses are evaluated inside as a function of two variables x, z. This setup is reported in Fig. 3.4, where the stress tensor motion along a cycle is shown on a 2D domain.

It can be seen clearly that, due to the nature of the load and stresses generated, the stress tensor generated inside the material is not directly affected by motion, but is simply dragged along the domain with the movement of the load. Therefore, it's evident how symmetry is resulting along the axis of motion, where the stress tensor remains the same along all the cycle.

The lack of variation in the in-depth behavior suggested a reduction from a two-dimensional to a one-dimensional domain (Fig. 3.5).

In the present model, the domain (Fig. 3.5(b)) it's a line of nodes, in the depth direction z. The symmetry along the moving direction has been eliminated and the

significant dimension of depth it's conserved in order to perform the fatigue analysis. The domain is then divided into evenly spaced nodes, acting as finite elements with independent properties.

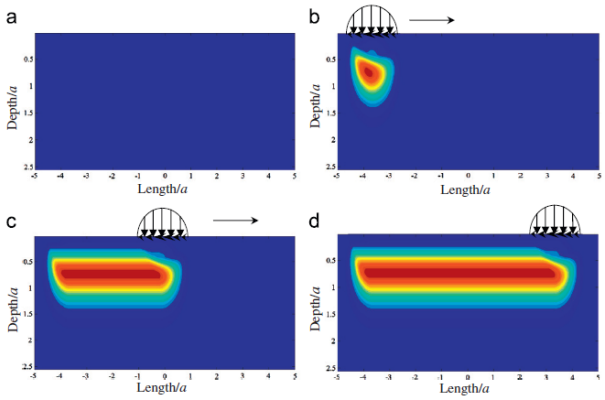
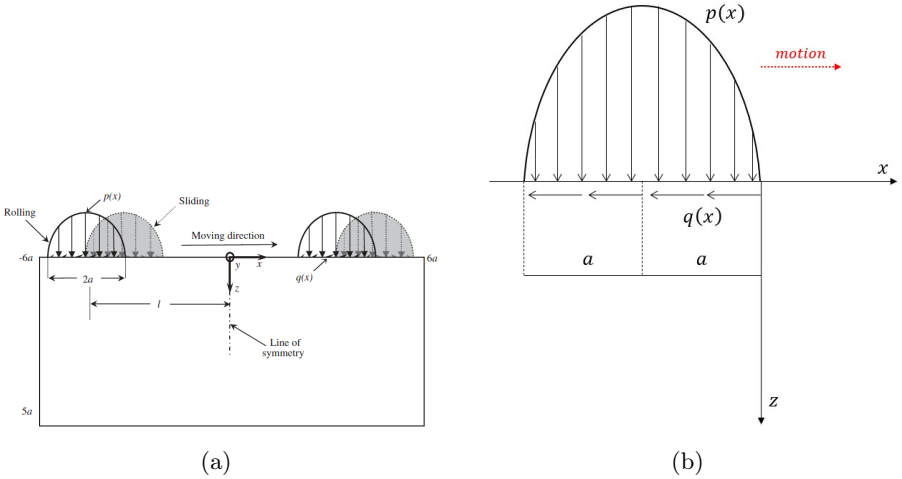


Figure 3.4: *Damage evolution contour for all x, at different depths for SAE 4340 [16]*



(a) (b)

Figure 3.5: *Schematic representation of the analysis domain in: (a) the 2D model used by Behesti et al. [16], (b) rendition of the current model*

The stress calculation and the subsequent evaluation of damage evolution has been performed at any node singularly. By this approximation not only the calculation time is reduced, but also redundant operations are omitted.

Once the type of domain to be used has been introduced, the motion of the load has to be formulated in a different way from the reference 2D model. In this work, two variables have been created, one for the z coordinate, which resembles the depth position of the nodes to be analyzed, and one for the x coordinate, identifying the width dimension of the load.

In order to simulate its motion, the stress generated it's evaluated step-wise, in the domain, for each relative position of the load distribution and the line domain. In this way, every single calculation on the x direction corresponds to a step-wise movement

of the load over the domain.

Once the cycle is concluded we obtain a complete passage of the rolling contact over a vertical section of the race underneath. Evaluating the stresses generated, the entire stress history felt at each cycle is known. The analysis is performed in the subsurface region, where depth is the only variable.

The stress tensor will be then evaluated for each position of the load (x), with respect to the line domain, and at each node (z) of the same domain. The result will be a 2D tensor representing the amount of stress felt at each depth per each step of the load motion. Validation of the stress evaluation will be taken from the shape and value of these tensors, representing the principal stresses acting on the material.

It can be concluded that exploiting the symmetry of the system and of the stress state, the reduction from a two-dimensional to a one-dimensional model is possible. The stress evaluation has always been performed in two dimensions, due to the load distribution, but the analysis is performed node by node in a one-dimensional domain.

This reduction reduces the weight and time of numerical calculations and makes the model more suitable for fatigue life predictions and comparisons.

3.4 Load Distribution

In a general case, the interaction transmitted through the contact can be resolved into a localized force P, normal to the contact surface S. This force is usually directed toward the contact and thus compressive over the entire domain.

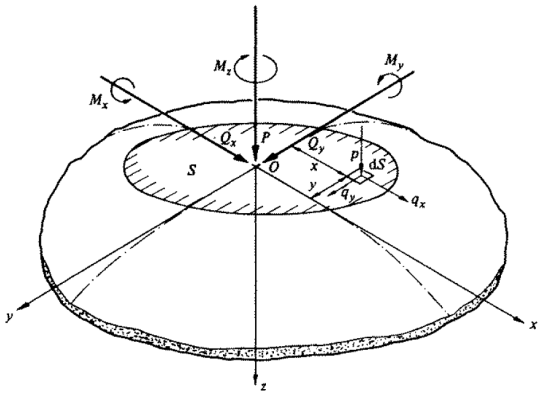


Figure 3.6: *Forces and moments acting on the contact area [5]*

An equivalent normal force P can be evaluated by the equilibrium formulation: integrating the normal traction p, which identifies a pressure, over the entire area of contact S (Fig. 3.6).

$$P = \int_S p dS \tag{3.1}$$

Depending on the geometry of the contact, as seen before, various integration approaches are available in theory as there are many approximations to simplify the mathematical procedure [5]. Here, the specific case taken into consideration is one

of a non-conforming contact with a slight curvature and one-dimensional conformity on the axis perpendicular to the direction of motion. Later, this configuration led to specific considerations on the symmetry of the interaction, allowing to lighten the general case and obtain simple solutions.

The case of the generic contact load distribution is reported in Fig. 3.7. Usually, the load is zero at the extremities and reaches its maximum in the proximity of its center, depending on the symmetry of the surfaces. The letter a and b identify the half-width of the distribution. In particular, b can be used as the second parameter in an asymmetric distribution or to indicate the width in the y direction, in case of specific configurations (eg. elliptical contact). Both parameters are directly dependent on the contact geometry.

In the case of symmetric distributions, as in the present study, only the letter a is used as a fundamental parameter. Therefore a circular symmetry is assumed, or a line contact with an axis much bigger than the other.

In the results section, Chap. 3.10, it's shown how the dimension of the contact surface (a) has a direct influence on many important factors, such as the position of the maxima reached by the stresses.

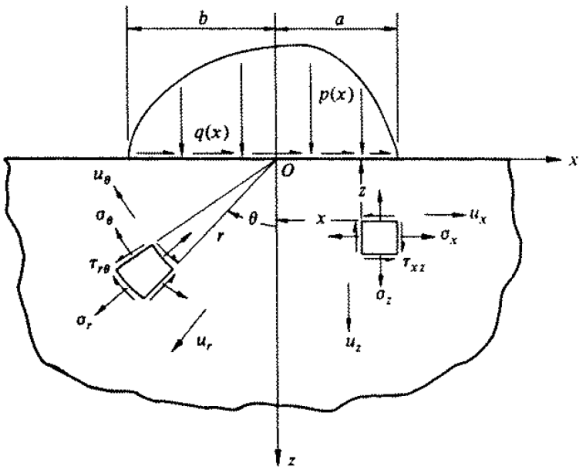


Figure 3.7: *Generic Hertzian load distribution and stress tensor reference [5]*

The roller bearings considered in this study have symmetry that admits the specularity of the load distribution along the rolling direction. It can be seen how the dimensions of the contact along the x, y axes are different. Therefore the case can be identified as an elliptical rolling contact. The parameter 'b' will then identify the dimension along the y direction, perpendicular to the direction of motion of the roller.

According to theory, [5], the elliptical case has a load distribution which is qualitatively similar to a circular one, considered as the most simple. The resulting equation for the pressure distribution is:

$$p(x, y) = p_0[1 - (x/a)^2 - (y/b)^2]^n \tag{3.2}$$

Which is a two-dimensional formulation taking into account the asymmetric configuration in the contact surface (Fig. 3.3). The parameter a will identify the smaller axis of the ellipse, parallel to the rolling direction (along x), and b the larger axis, along y.

The exponent n can take different values according to the approach considered when establishing the total equilibrium. In Hertzian theory $n = 1/2$ [5].

The resulting total load, over the whole contact surface, can be expressed as:

$$P = 2\pi abp_0/3 \tag{3.3}$$

It's obvious that these equations (Eq. 3.2 and 3.3) don't suit our model, because both are bi-dimensional. This load distribution results in a two-dimensional model with the introduction of y domain, that has been previously neglected due to geometry.

Analyzing closely the symmetry, it's resulting that this latter dimension is not the principal direction of motion and of stress distribution. Being $b > a$ and resulting in less pronounced stress gradients, a further approximation can be introduced.

By neglecting the influence of b dimension, in order to have a one-dimensional simulation, the final model is reduced to the line contact case, usually generated by cylindrical surfaces. Clearly, the bodies shown in Fig. 3.2 are not cylinders, having a curvature along the main axis. Anyway, this curvature is not very pronounced and is present in both the rollers and the gear race over which rolling occurs.

Therefore, we have conformity along y direction, where lies the major axis of the ellipse. In the central section, where the most critical stresses are concentrated, the elongated dimension along y allows the assumption of a linear symmetry, with constant behavior in this direction. This assumption allows to focus only on x axis, where the non-conforming contact is more pronounced, stresses gradients are higher and also tangent traction is generated by friction.

In this case, the contact domain is reduced by one dimension and happens over a symmetrical domain of dimension $2a$. This assumption is considered acceptable in Hertzian theory when b becomes large compared with a, in elliptical contacts as the one under exam [5].

The general pressure distribution has been studied on x axis only, along the direction of motion.

The resulting equation for the pressure at a generic point x of the distribution is then:

$$p(x) = \frac{2P}{\pi a^2}(a^2 - x^2)^{1/2} \tag{3.4}$$

where a is the aforementioned half-width of the load distribution, x is the position coordinate and P is the maximum normal force applied on the contact surface. It should be noted that P is a scalar value, indicating the maximum force applied in the contact. The shape instead is given by the second part of the equation, concerning the position of the point in which the load is evaluated.

This formulation agrees with the shape indications given before, dropping to zero at the edges and having a maximum at $x = 0$.

The maximum pressure can be defined as:

$$p_0 = \frac{2P}{\pi a^2} \tag{3.5}$$

Therefore:

$$p(x) = p_0(a^2 - x^2)^{1/2} \tag{3.6}$$

Now the scalar parameter is no more a concentrated force, which is purely theoretical and unrealistic, but a pressure indicating the maximum value reached by the load.

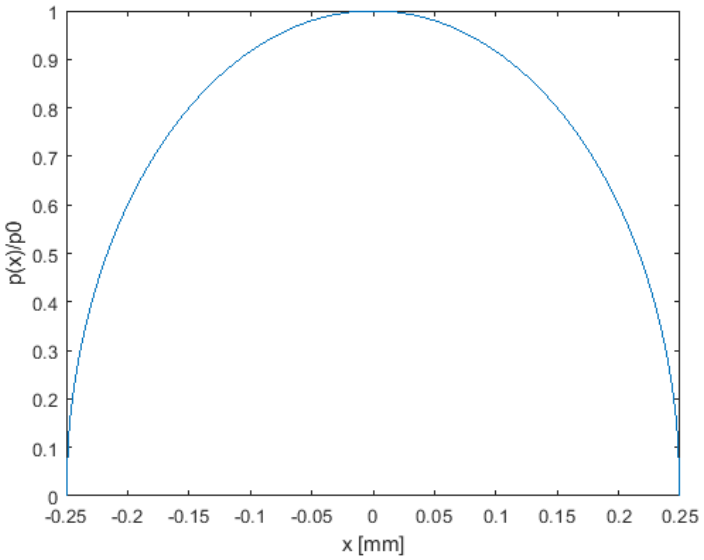


Figure 3.8: *Load distribution representation, from Matlab code*

This expression results to be useful if the data for the analysis come from experimental runs and FEM analysis. In most references also, the main reported load parameter is not a force $[N]$ but directly a pressure $[MPa]$ coming from real cases. This variable is called 'maximum Hertzian pressure' [16] and will vary depending on the application and on the loads transmitted through the system.

It can be concluded that the pressure generated by a line or elliptical rolling contact has a parabolic shape, as depicted in Fig. 3.5 and 3.8. This profile satisfies the symmetry imposed by the chosen configuration and is well defined by theory.

3.5 Friction

The presence of friction between the bodies gives rise also to forces lying in the plane tangent to the contact surface.

These tangential forces are directly dependent on the magnitude of the normal load, indicated generally as P although should be distributed, through the friction coefficient, mu:

$$Q \leq \mu P \tag{3.7}$$

In the system considered, the only load present is the pressure distribution analyzed previously. Therefore, the only tangential force present will be parallel to the direction of motion, lying on the x axis, as indicated in Fig. 3.5(b). This frictional traction will give a contribution to the stresses, by shifting the position of the critical depth toward the surface [16].

The value of the friction coefficient is specific for each system and lubrication regime. Usually in bearing gears the values are very small (≤ 0.1), as confirmed

by Beheshti [16] in the initiation life evaluation. In models studying the influence of lubricants on crack propagation, such as from Ancellotti et al.[148], the friction coefficient between the specimens is assumed to be 0.1 or lower.

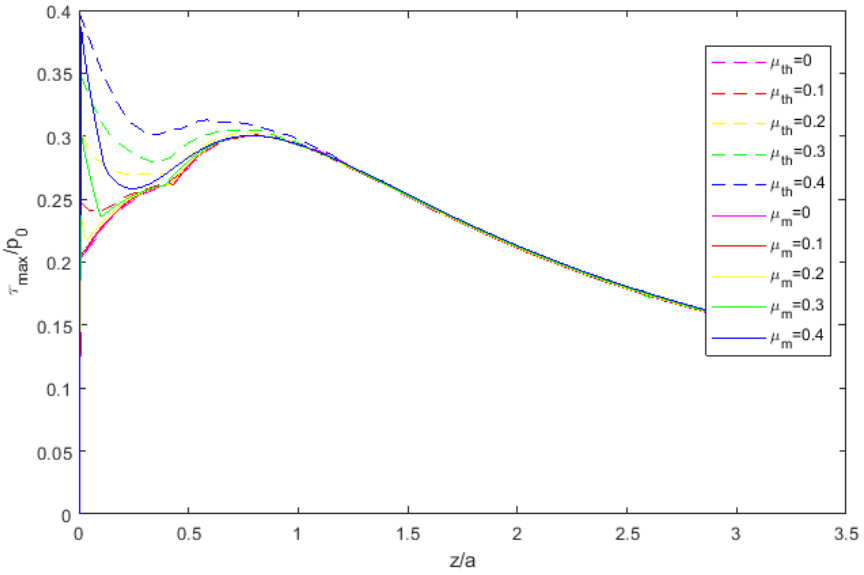


Figure 3.9: *Comparison between representation of critical shear value for varying friction coefficients, from Beheshti [16] (dashed lines) and the current model*

Anyway, from the urgency of validation with reference [16] and to demonstrate the approximations made on the friction, a study on the effect of μ on the maximum shear stress has been made.

In Fig. 3.9, different values of contact friction have been analyzed to evaluate the effect on the position and the magnitude of stresses and fatigue life. It can be seen good accordance with the trend of the graphs, especially at low friction values. The difference at shallower depths can be due to different usage of variables such or, mainly, the way to evaluate the maximum shear stress. Anyway, the following fatigue analysis will be performed in the low-friction regime.

At values approaching of $\mu = 0.3$ and higher, the maximum stress occurs is subject to a shift toward the surface (Fig. 3.9), as also anticipated in the State of the Art. This effect is due to the growing relevance assumed by the traction force ($q(x)$) when friction increases. The consequence on the critical shear stress is an increment of the maximum value in the near-surface region, when μ approaches 0.3, which results in a stress peak and damage formation at shallow depths.

This explains the necessity of a good lubrication regime, to avoid aggressive spalling on the surface and high wear. On the other hand, the results in Fig. 3.9 show how the maximum shear peak located in the subsurface region is not affected by friction. This is in accordance with the model developed by Beheshti [16], although being different from the classic integrated formulas introduced in the past [5][149].

3.6 Stress Components and Maximum Shear

As explained in the previous sections, during contact between surfaces surface tangential traction $q(x)$ and a normal pressure $p(x)$ are generated. These components are distributed arbitrarily, depending on the kind of contact, along the domain. In order to evaluate the stresses generated from these actions, an integral formulation is generally chosen.

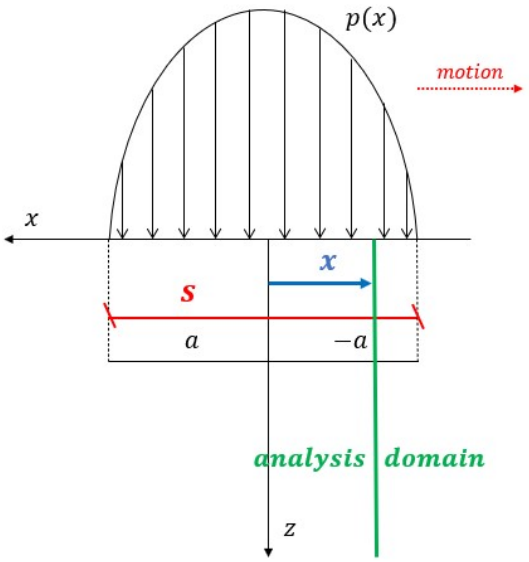


Figure 3.10: *2D representation of the integration calculation and variables*

Considering an infinitesimal element of the analysis domain with width ds, the actions acting on it can be expressed as concentrated infinitesimal forces. The contribution of the normal pressure is expressed by pds, where the contribution of the tangential traction is expressed by qds. The position of the infinitesimal element is expressed with $(x-s)$, where x is the position of the element, with respect to the center of the load distribution or the origin, coincident in our case (Fig. 3.10). Instead, s is the integration variable spanning over the whole load to take into account the entire contribution. For this reason, it spans from $-a$ to a over each calculation.

Integrating over the analysis domain, or loaded region to account for the contribution of all the load, the principal stresses components can be evaluated [5].

In the specific case of an elliptical contact the equations are the following:

$$\sigma_x = -\frac{2z}{\pi}\int_{-b}^{a}\frac{p(s)(x-s)^2ds}{[(x-s^2)+z^2]^2} - \frac{2}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)^3ds}{[(x-s)^2+z^2]^2} + \sigma_\Theta \tag{3.8}$$

$$\sigma_z = -\frac{2z^3}{\pi}\int_{-b}^{a}\frac{p(s)ds}{[(x-s^2)+z^2]^2} - \frac{2z^2}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)ds}{[(x-s)^2+z^2]^2} \tag{3.9}$$

$$\tau_{xz} = -\frac{2z^2}{\pi}\int_{-b}^{a}\frac{p(s)(x-s)ds}{[(x-s^2)+z^2]^2} - \frac{2z}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)^2ds}{[(x-s)^2+z^2]^2} \tag{3.10}$$

In which the x and z coordinates identify the position of the point in which the stress should be analyzed and s is the integration variable that spans through the pressure distribution. The last term in Eq. 3.8, accounts for the residual stresses, usually compressive, induced by thermal treatments.

This formulation gives some singularity problems at the origin if integrated, but a direct evaluation in integral form has been performed successfully in this model. It should also be noted that these equations are purely theoretical and assume the material to be defect-less.

As previously introduced, it can be seen that the stress tensor is multiaxial and composed of both normal stresses and shear, evaluated in the xz plane.

This configuration is confirmed by Depouhon et al. [7], which analyses a similar bearing gear system for the same application as the one under investigation. The total stress tensor is found to be multiaxial and also the shear contribution is found to have two critical directions.

The multiaxial nature of rolling contact stress tensor is introduced also in the Hertzian theory [5]. In the case of small deformations produced by the surfaces, the contact can be considered elastic, allowing some general considerations and specific stress tensor shapes. As reported by Johnson [5], the contact between two cylindrical bodies was considered also by Hertz as the limit of the elliptical case, when $b >> a$.

The problem then becomes two-dimensional, over a domain of width $2a$. Developing the results found for the general line loading, and introducing the elastic approximation for small deformations, the equations for the load distribution evaluation result to be equal to Eq. 3.4 and 3.6, presented previously.

Therefore, considering the stresses to be zero outside the contact region [5], integration along the z axis is possible, giving:

$$\sigma_x = -\frac{p_0}{a}[(a^2 + 2z^2)(a^2 + z^2)^{-1/2} - 2z] \tag{3.11}$$

$$\sigma_z = -p_0 a(a^2 + z^2)^{-1/2} \tag{3.12}$$

Which are considered to be the principal stresses acting along the axis beneath the center of the load distribution. This equations are consistent only for position $x = 0$, when the load is exactly over the line domain considered.

Should be noted that, in the case of plane strain, a third stress factor can be considered, dependent on the Poisson's ratio.

$$\sigma_y = \nu(\sigma_x + \sigma_z) \tag{3.13}$$

This formulation can be applied at any kind of equation, integral or approximated. In any case, is not usually indicated because it doesn't give any significant contribution to rolling contact fatigue or crack formation.

The shear factor τ_{xz}, acting perpendicular or parallel to the rolling direction, is not considered in this approximation. Another shear component is considered instead:

$$\tau_l = p_0 a[z - z^2(a^2 - z^2)^{-1/2}] \tag{3.14}$$

Called 'principal shear stress' by Johnson, is reported to have a defined maximum depth assessed at $z = 0.78a$, where a is the load dimension parameter. At this depth, the maximum shear value is reached, and it is $(\tau_l)_{max} = 0.30p_0$, linearly dependent on the maximum pressure applied at the contact.

This second shear factor has been highlighted also by Depouhon et al. [7] in the Airbus study on the bearing gear failure which is considered as the starting point for this thesis work.

The higher depth and more pronounced peak of the so-called 'maximum shear' makes it the most critical factor to be monitored when performing a fatigue analysis.

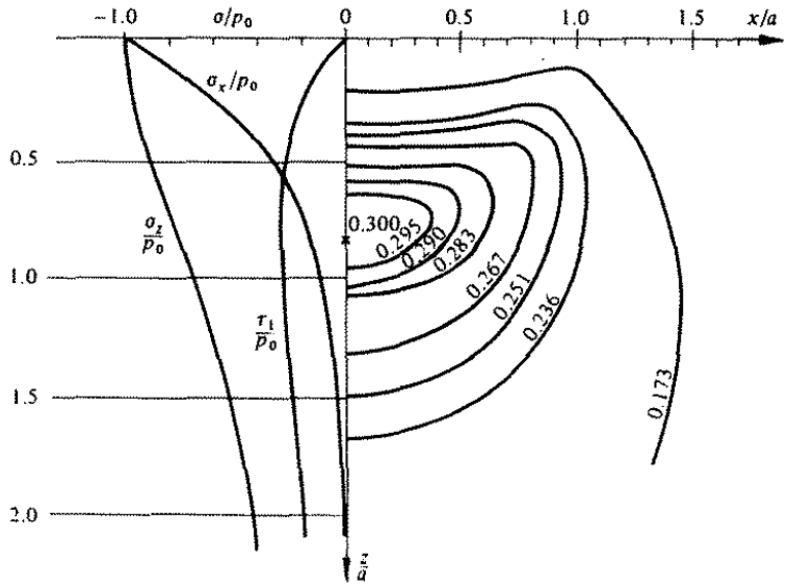


Figure 3.11: *Representation of in depth profile of (a) principal stresses on the symmetry line, (b) level curves of maximum shear stress [5]*

As indicated in Fig. 3.12, the critical value of shear stress lies in a plane which is perpendicular to the one of motion, coincident with the analysis domain along z direction. This component lies in the xy plane and can be evaluated in different ways.

Most of the studies on rolling contact fatigue are based on the choice of specific critical stress, acting as the driving factor for crack formation. Different reference values have been proposed as the Von Mises equivalent stress and the alternate shear among the others. More recently, in studies such as the one by Beheshti [16], the maximum shear stress has been more frequently addressed as the chosen critical value.

Cheng [72] demonstrated that the selection of the critical stress parameter can be considered arbitrary. This happens when considering conditions of non-frictional Hertzian line contact. In this casem the position and value of the critical shear can be evaluated only as a function of the maximum contact pressure and of the contact width.

It has been reported that from the beginning of fatigue life evaluation different critical shear values have been chosen. Starting from Lundberg and Palmgren [8] whose proposed the maximum orthogonal shear stress to be the critical value and passing through Zaretsky [112] which proposed the maximum shear stress and then the Von Mises equivalent stress proposed by Ioannides and Harris [9] and the octahedral shear stress by Zhou [113]. Some experimental tests were run to validate one proposal or another, but the results showed significant inconsistencies which created further confusion.

The choice is then left at the specific situation, depending on the contact mechanics

and the operational conditions in which each study is performed. In the particular case of non-frictional Hertzian line contact instead, it has been demonstrated [72] that the four previous criteria can be considered equivalent if considered in relation to the proper fatigue limiting stress.

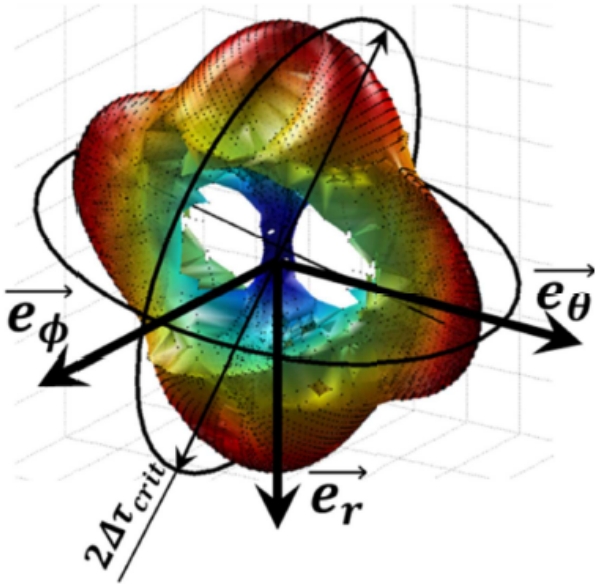


Figure 3.12: *Shear stress envelope at maximum stress position, evaluated by Depouhon et al. [7]*

All critical shear stresses can be defined as a function of the Hertzian pressure p_0:

$$\tau_{orth} = 0.25p_0 \tag{3.15}$$

$$\tau_{max} = 0.3p_0 \tag{3.16}$$

$$\sigma_{Mises} = 0.322p_0 \tag{3.17}$$

$$\sigma_{oct} = 0.263p_0 \tag{3.18}$$

And due to this direct relation, a critical factor can be defined as:

$$(\tau_{crit} - \tau_f) = m(p_0 - p_u) \tag{3.19}$$

where m and p_u, limiting pressure, are constants. Thus, per each case and critical shear values chosen, the corresponding fatigue limiting stress τ_f is evaluated, only depending on the Hertzian pressure applied.

In the same way, also the critical depth can be defined independently from the chosen critical stress:

$$z = ka \tag{3.20}$$

where k is also a constant, considered 0.78 by Cheng et al. [72], and a is the half contact width.

In the current work, the chosen critical shear stress is the maximum shear. Beheshti [16] introduced a complex way to evaluate this factor, passing through the computation

of equivalent stress and then through conversion using the Mohr's circle. Instead, taking reference from Johson [5] and other studies, a similar and simpler way to define τ_{max} is available.

Because the stresses evaluated with Eq. 3.8 and 3.9 are already the principal stresses [5], the maximum shear stress can be calculated simply as the half-module of their difference:

$$\tau_{max} = \frac{1}{2}|\sigma_x - \sigma_z| \tag{3.21}$$

According to Eq. 3.16, the peak will reach a value of $0.3p_0$ at a depth equal to $0.78a$.

To perform the stress equation integration in Matlab, the trapezoidal (*trapz*) integration function has been preferred to the direct integration procedure to avoid the singularity which is present in point of coordinates $x, z = 0$. An integration variable s is defined in the x domain, spanning in the interval $(-a, a)$ (which is often represented weighted in graphs over the half-width dimension a, resulting to span over a $-1, 1$ interval). This means that all the load distribution is considered when calculating the stress at each depth z per each position x of the load, with respect to the analysis domain.

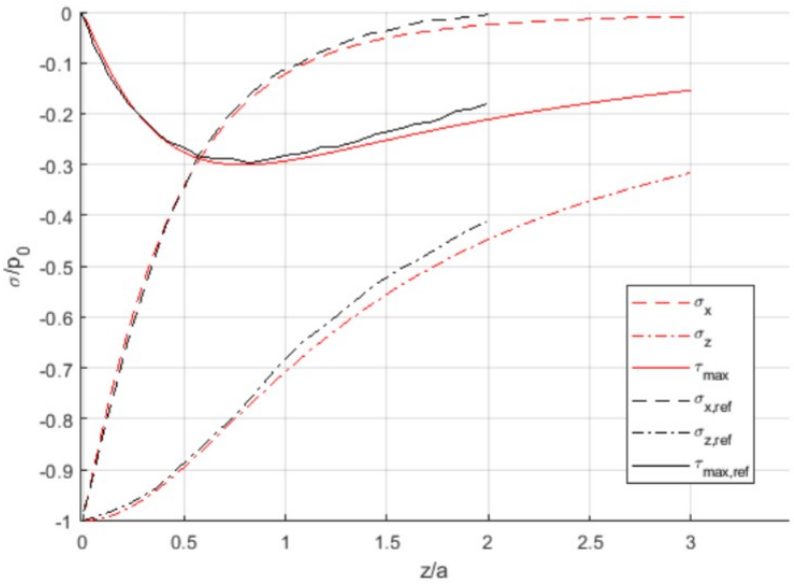


Figure 3.13: *Comparison between the reference [5] and model results for the stress tensor components σ_x, σ_z and τ_{xy} evaluated at $x = 0$ for $\mu = 0$*

The results obtained are two-dimensional, representing the stress felt by the line analysis domain over an entire cycle, due to the presence of the whole load distribution.

Each node of the x dimension represents the position of the load with respect to the analysis domain, where at $x = 0$ the load is approaching the line and at $x = 2a$ the load has completed an entire cycle.

The surface graphs of the four evaluated stresses can be seen in Fig. 3.16 and 3.15 and 3.17, for a representative friction coefficient of 0.1. The different perspectives are chosen to show in the best way the properties of each component.

In Fig. 3.13, it is evident the good resemblance of the tensor, evaluated at $x = 0$, with the theoretical predictions, showing correct peaks for the principal stresses and the critical shear. It should be noted in particular the almost exact correspondence for the maximum shear distribution, which is taken as the critical factor for the fatigue analysis. This result is obtained by deriving the τ_{xy} straight from the absolute value of the half difference between the two principal stresses, as in Eq. 3.21.

Both the principal stresses respect the value of the peak at position $z = 0$ and the gradient seems to be perfectly comparable. Moreover, the critical shear factor, chosen as the maximum shear stress, has been flipped to the negative side to allow a direct comparison of the two graphs. The peak is correctly assessed at a value of $\tau_{max} = 0.3p_0$, reached at a depth of $z = 0.78a$.

Another way to obtain the maximum shear stress is proposed by Beheshti [16], where all the three principal stresses components along x, y, z directions are considered. The critical stress is then selected as:

$$\tau_{max} = max[|\frac{\sigma_1 - \sigma_2}{2}|, |\frac{\sigma_1 - \sigma_3}{2}|, |\frac{\sigma_2 - \sigma_3}{2}|] \tag{3.22}$$

where the stresses considered are the principal ones, directly evaluated with Eqs. 3.8 and 3.9 and 3.13. This formulation has already been used for the stress distribution in Fig. 3.9. The comparison of the results from the two formulations is shown in Fig. 3.14, where the difference is minimal and limited to the near-surface part of the domain.

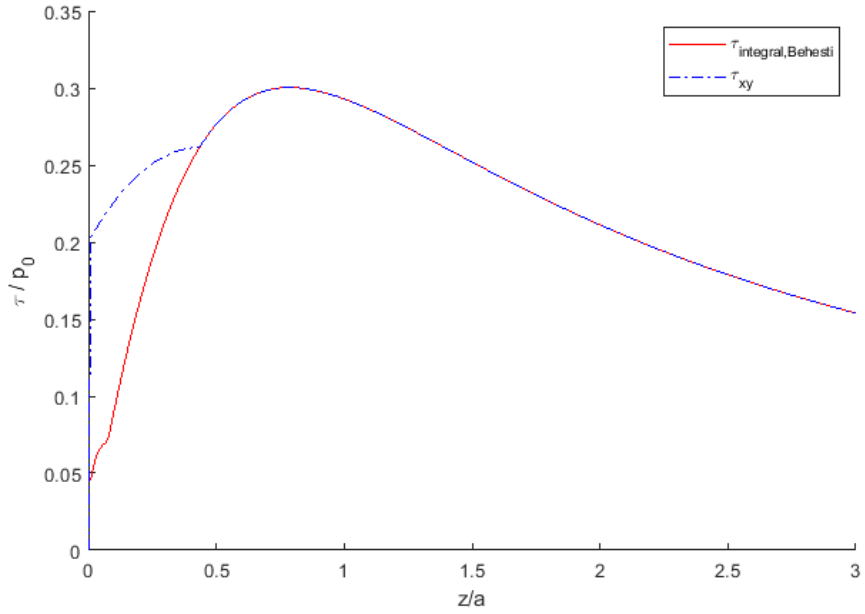


Figure 3.14: *Comparison between the model results for the critical shear stress vector, evaluated at $x = 0$ for $\mu = 0$, for the Beheshti [16] and general formulation [5] mode*

Looking at the two-dimensional stress distribution is also important to notice the shape of τ_{xz}, which undergoes a complete sign inversion. The two shear factors were

the ones indicated by Depuohon [7], giving rise to the double critical direction of the shear tensor envelope in Fig. 3.12.

The results shown in this section for stress evaluation and friction influence on the critical shear shape are considered to be satisfying. Thus, this model of the code is to be considered validated by theory [5][16].

3.7 Integrated Equations for Stress Tensor Evaluation

Many studies attempted to give a simple solution for the integral Eqs. 3.8 and 3.9 and 3.10, representing the principal stress components generated by a rolling contact. The most relevant results were achieved by McEwen [149], Poritsky [150], Smith and Liu [151] and Sackfield and Hills [152][153].

Because of the direct inclusion of friction effect in the model, the McEwen [5][149] formulation was taken in this work. Proceeding from the elastic approximation that led to the Eqs. 3.11 and 3.12 and 3.14, a way to express the stress components in a generic point of a 2D domain was developed. Here the second dimension (x) was considered to be the relative position of the load and the analyzed line domain.

The complex integration of the general equations has been solved by introducing two-dimensional parameters m, n which have the same sign of z, x respectively:

$$m^2 = \frac{1}{2}[((a^2 - x^2 + z^2)^2 + 4x^2z^2)^{1/2} + (a^2 - x^2 + z^2)] \tag{3.23}$$

$$n^2 = \frac{1}{2}[((a^2 - x^2 + z^2)^2 + 4x^2z^2)^{1/2} - (a^2 - x^2 + z^2)] \tag{3.24}$$

These parameters are a function of the position (x, z) of the point, in which the stresses is evaluated, and of the contact dimension a. It can be noted how the contact half-width still fills an important role, being the only 'external' variable appearing in the equation.

The principal stresses are then obtained by [149]:

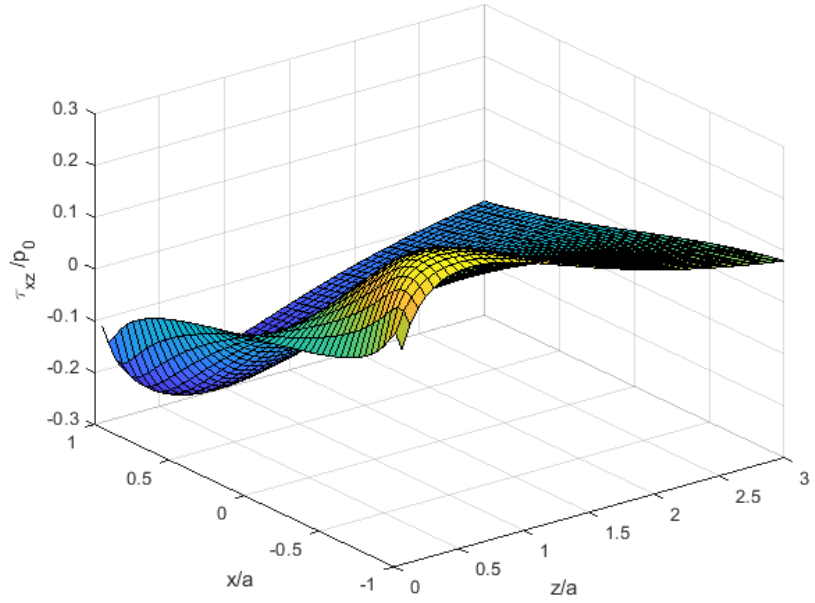
$$\sigma_x = -\frac{2P}{\pi a^2}[m - 2z + 2\mu(x - n) + m(\frac{z^2 + n^2}{m^2 + n^2}) + \mu n(\frac{z^2 - m^2}{m^2 + n^2})] \tag{3.25}$$

$$\sigma_z = -\frac{2P}{\pi a^2}[m(1 - \frac{z^2 + n^2}{m^2 + n^2}) - \mu n(\frac{z^2 - m^2}{m^2 + n^2})] \tag{3.26}$$

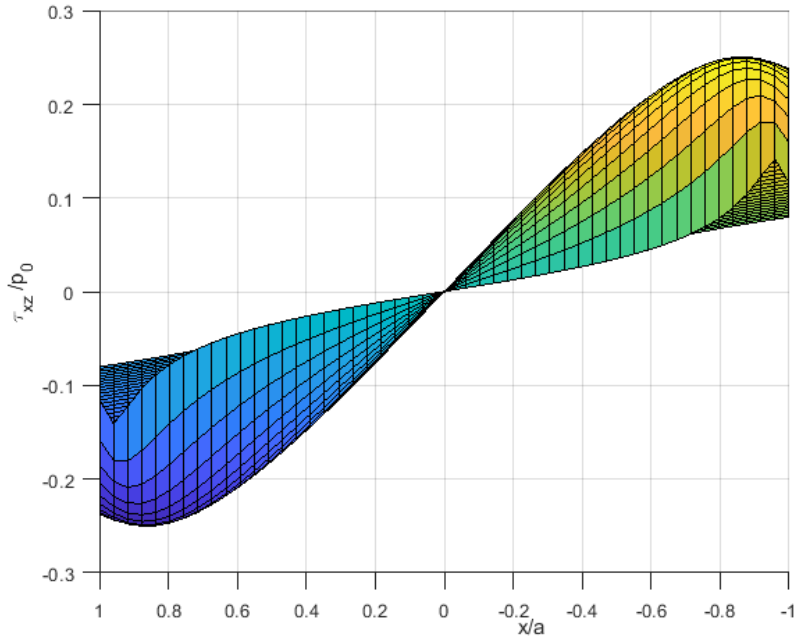
$$\tau_{xz} = -\frac{2P}{\pi a^2}[\mu(m - 2z) - n(\frac{z^2 - m^2}{m^2 + n^2})] \tag{3.27}$$

It should be noted that in this notation, the parameter a appears only in the load because is already present in the m, n parameters. The load is indicated in the initial general way already treated in the previous chapter. All the generalizations and the expression as a function of the maximum Hertzian pressure are still valid.

It can be noted a close resemblance in values and shape with the tensor evaluated with the general integral functions (Fig. 3.16 and 3.17). An appreciable difference is in the module of τ_{xz}, (Fig. 3.19(a)), which has two peaks with the same sign, whereas the same factor evaluated with the integral equations underwent a sign inversion (Fig. 3.15). This effect can be attributed to the McEwen formulation, which considers

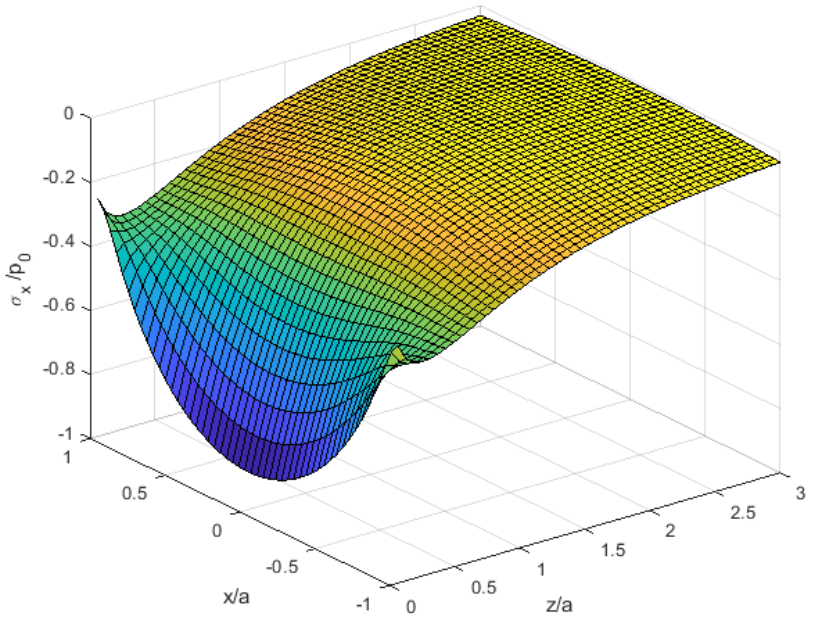


(a)

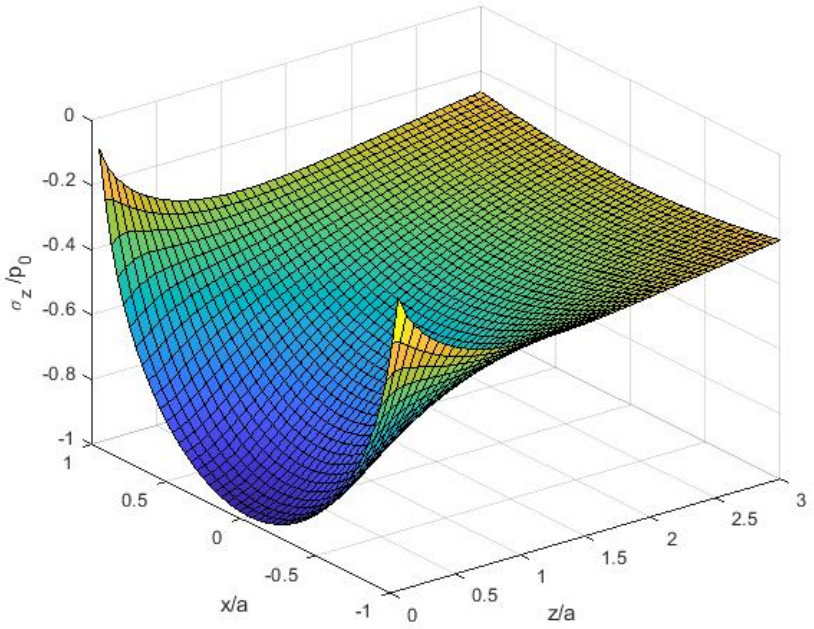


(b)

Figure 3.15: *Surface graph of Matlab stress tensor evaluation for τ_{xz} in: (a) isometrci view, (b) front view*

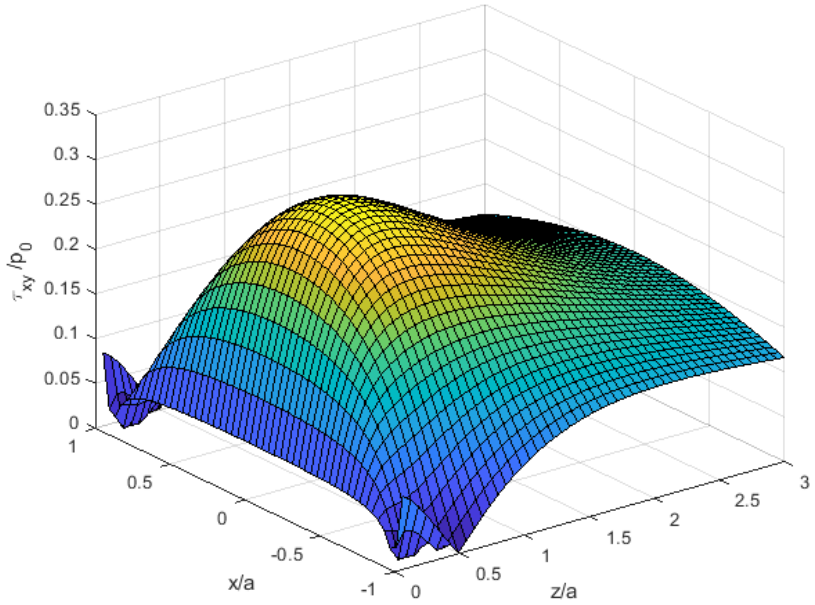


(a)

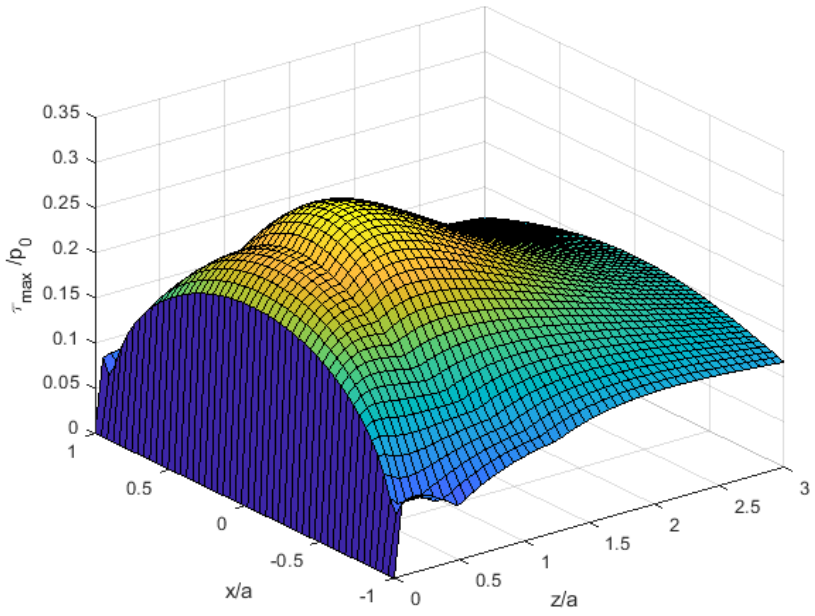


(b)

Figure 3.16: *Surface graphs of Matlab stress tensor evaluation for: (a) σ_x, (b) σ_z*

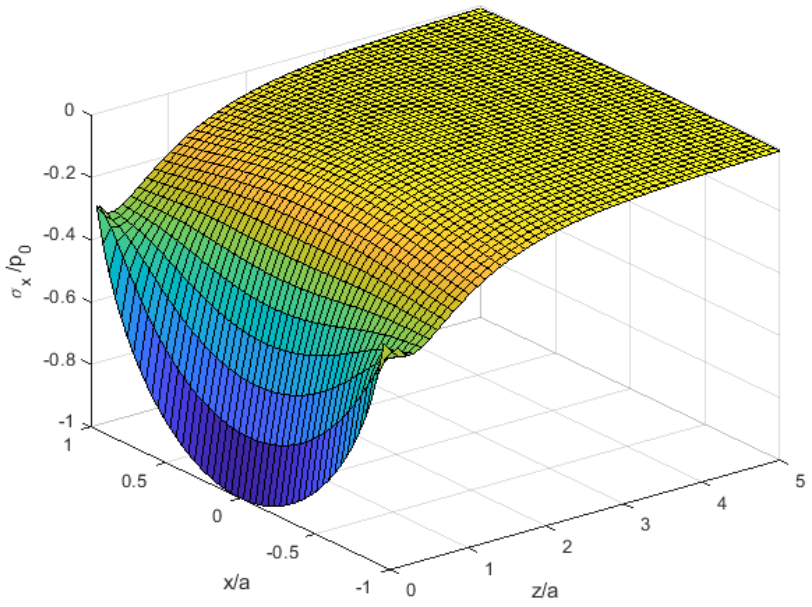


(a)

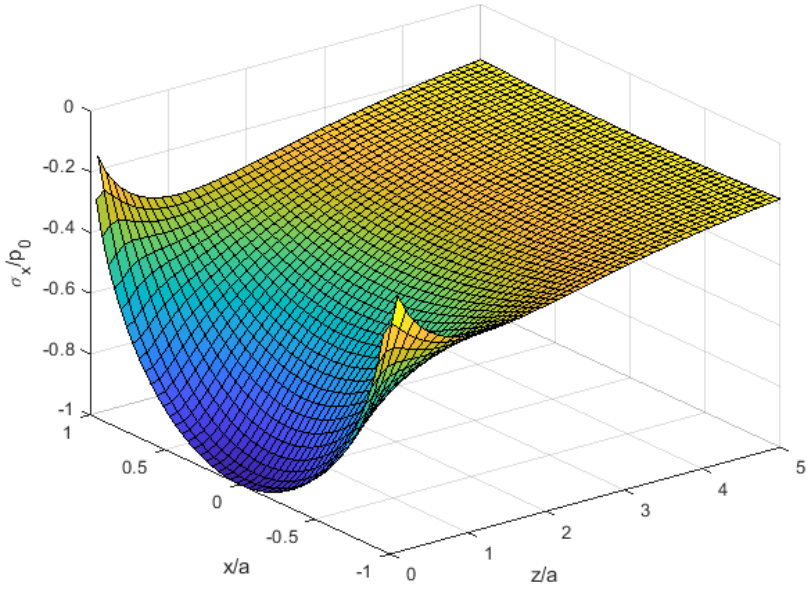


(b)

Figure 3.17: *Surface graphs of Matlab stress tensor evaluation for τ_{max}, evaluated in: (a) general formulation [5], (b) Beheshti formulation [16]*

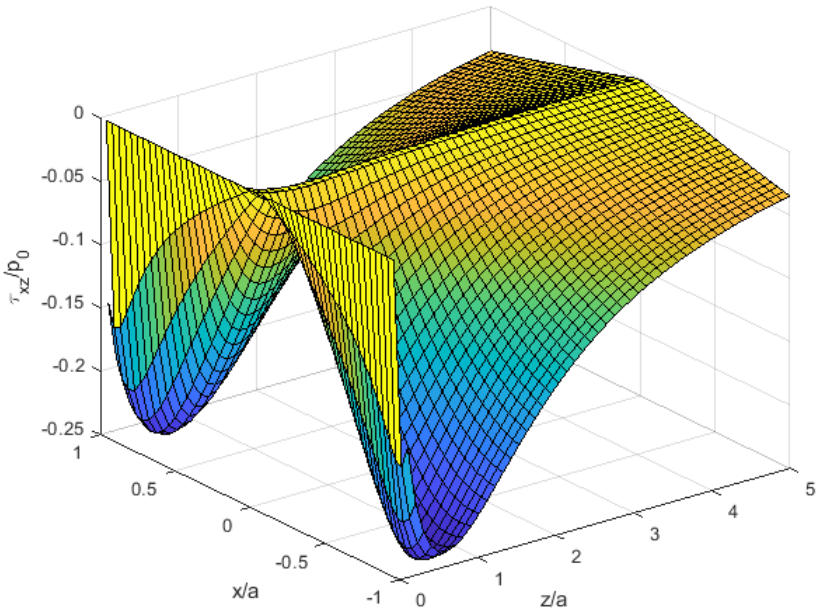


(a)

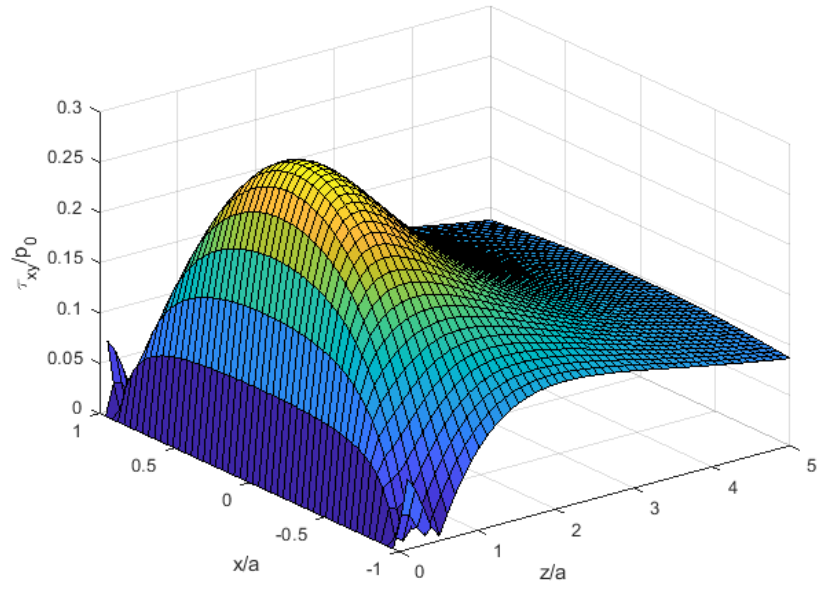


(b)

Figure 3.18: *Surface graphs of Matlab stress tensor evaluation with McEwen formulation, for: (a) σ_x, (b) σ_z*

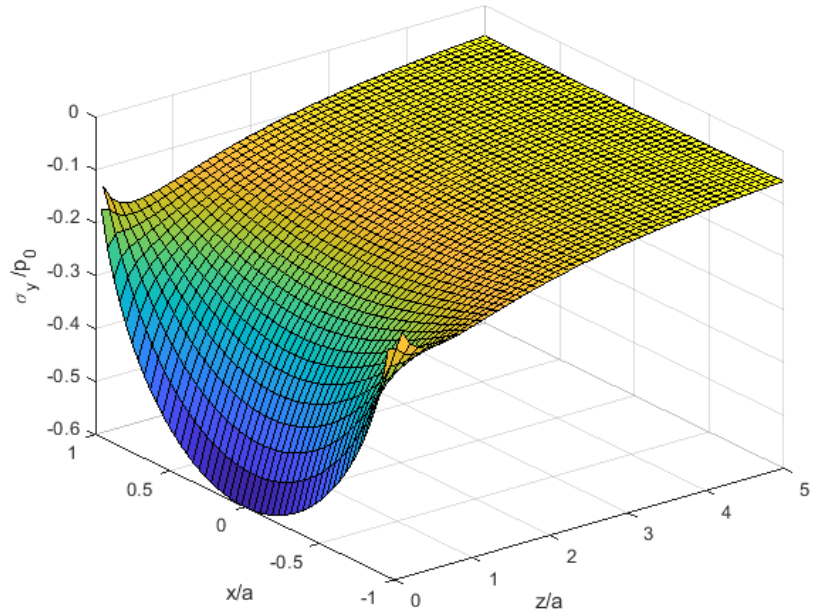


(a)

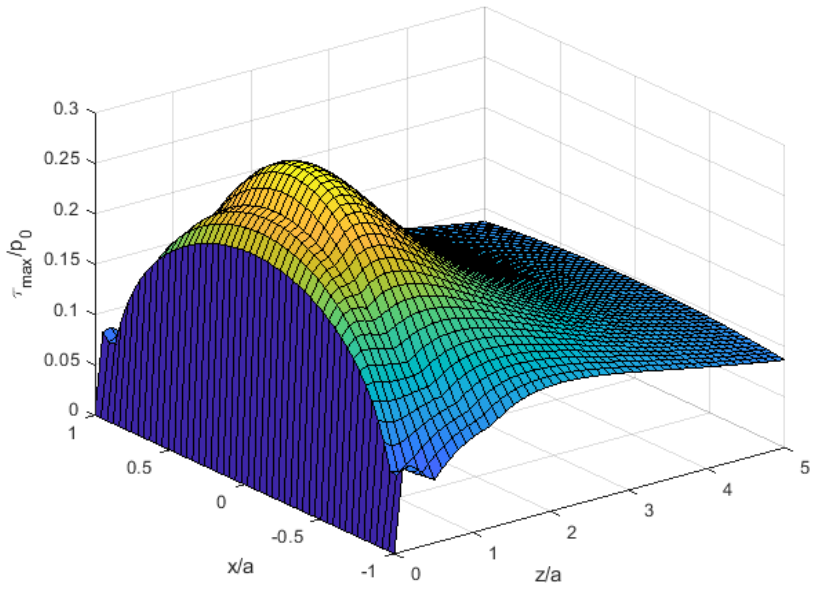


(b)

Figure 3.19: *Surface graphs of Matlab stress tensor evaluation with McEwen formulation, for: (a) τ_{xz}, (b) τ_{xy}*



(a)



(b)

Figure 3.20: *Surface graphs of Matlab stress tensor evaluation with McEwen formulation, for: (a) σ_y, (b) τ_{max}*

position values x, z always elevated at the second power. This results in the loss of information on the sing of the shear stress factor. Anyway, pure shear is not considered as critical in our analysis.

The most relevant difference instead, is in the τ_{max} distribution, which reaches the peak at minor depth with respect to the equivalent values obtained with the integral result. The τ_{xy} value instead (Fig. 3.19(b)), remains essentially unvaried and similar to the one evaluated in the previous section (Fig. 3.17(a)).

A direct comparison between the critical shear profiles from the different evaluations is shown in Fig. 3.21. A good superposition of the maximum shear stress peak is conserved when considering the $\tau_x y$ value from McEwen calculations.

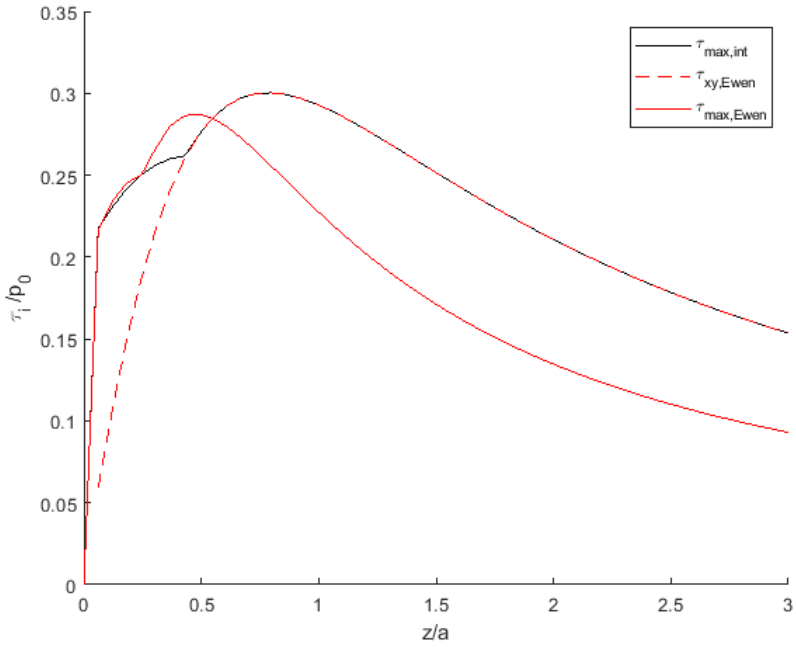


Figure 3.21: *Comparison of resulting critical stress profile, from integral and McEwen's calculation methods*

Should be noted that the McEwen equations require a much more limited time than the general integration, for obvious reasons. But the discrepancy shown in the critical shear stress component τ_{max}, makes it less reliable for the current analysis.

The greatest limit of McEwen's model is instead the friction behavior (Fig. 3.22). As shown in Fig. 3.21, the comparison at null friction is good, if considering the tau_{xy} factor as critical stress. When increasing the friction coefficient instead, McEwen's model fails, returning a sensible overestimation of the resulting stress on the whole domain. This results to be a crucial factor against the choice of using this stress calculation model.

In conclusion, it has been stated that the McEwen method is lighter in a matter of calculation time and reliable if using τ_{xy} as critical stress. But the steep variations at friction increment are not reliable for a more comprehensive analysis.

So, the chosen method for the evaluation of the fatigue initiation life will be the

general integration formulation, which is more complete and gives better results in accordance with theory [5][16][154]. From this point on, all the calculations of the stress tensor, for further computations, will be performed with this method.

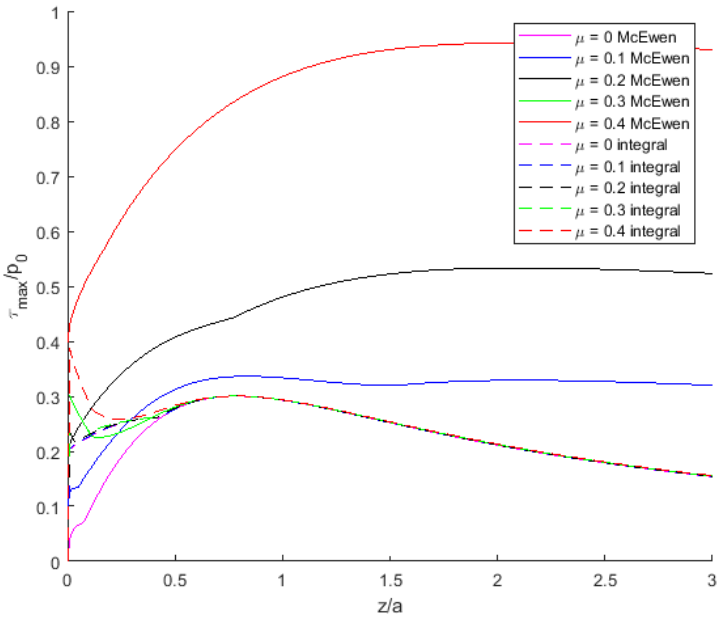


Figure 3.22: *Comparison on the critical shear stress profile when increasing the friction coefficient for the McEwen and general integration evaluation modes*

3.8 Continuum Damage Mechanics

As presented in the state of the art, continuum damage mechanics is a theory derived from thermodynamics potentials to evaluate the evolution of damage. An important advantage of this approach is the complete independence from empirical growth and crack parameters. Therefore, as stated by Beheshti [16], the "macroscopically obtained material parameters [141]" are the basis for the damage analysis.

The parameter identifying the amount of damage in the material is expressed by the variable D, which is a macroscopic state variable. The damage expressed in this way derives from the definition given by Lemaitre [155]. Variable D sums up the quantitative concentration of micro-cracks and voids present in a given section of a plane inside the material. Taking infinitesimal elements of these planes, the damage can be intended as a value associated with a specific node.

Moreover, the damage is assumed to be isotropic due to the nature of the materials considered. Hence, the variable D can be considered as a scalar quantity associated with a specific region or element of the domain. This brings to a variation in the material's properties such as Young's modulus. The Poisson's ratio is instead considered

independent from the damage evolution process [156] and will be taken as constant in this model.

The value of the damage variable D will increase at each rolling contact cycle if the stress on each element reaches a threshold specific to the material. The aim of the computation is then to define the number of cycles (N) required to start the crack formation and give rise to fatigue initiation. This happens when D reaches a critical value called D_c, while $D_c \leq 1$ always.

The critical damage value D_c is considered to be an intrinsic property of the material [157] and should be determined by experimental test. As introduced by Beheshti [16] and Bhattacharya [141], a simple tension test can be performed to determine the material's properties used in this analytical-numerical model. Among these properties, also the D_c value is defined from the elastic and mechanic properties of the material in such tests.

This value acts as the upper limit for damage evolution. Once variable D reaches this threshold, the fatigue initiation phase is considered to be completed for the specific node. It should be clear that the achievement of the critical value for damage is not equivalent to fracture. At this value, the element is considered to be degraded and the continuity of the material is lost in that section. It can be stated then, that a defect is formed, meaning the initiation of a crack in a previously intact material.

Following from the thermodynamic approach proposed by Bhattacharya and Ellingwood [145], which has been verified experimentally [141], is possible to integrate the Eq. 3.8 and 3.9 and 3.10 and the damage value can be evaluated at each cycle. The system is considered to evolve through a sequence of equilibrium states, according to thermodynamics principles.

From the solution of the Helmholtz energy function for a uniaxial load, an equation for fatigue damage growth can be derived:

$$\frac{dD}{d\epsilon} = \frac{-\sigma_\infty}{\psi_D} \tag{3.28}$$

verified if

$$\sigma_\infty \geqslant S_e \geqslant 0, \quad \dot{\epsilon} > 0 \tag{3.29}$$

otherwise, the equation results equal to zero because no damage is generated. The initial damage for dD calculation is considered to be the one from the previous cycle (D_{i-1}).

The term ψ_D indicates the free energy per unit volume and can be expressed per each cycle i as:

$$\psi = \int_{\epsilon_{0i}}^{\epsilon} \sigma d\epsilon' - (\gamma - \gamma_{i-1}) \tag{3.30}$$

where γ is the surface energy needed for the formation of defects. This parameter can be obtained theoretically by considering defects to be spherical voids, uniformly distributed inside the material, a linear micro-scale force-displacement relation and the condition of void formation is when the stress on its boundary is equal to the true failure stress σ_f.

Then a simple way to evaluate the surface energy can be expressed as a function of known parameters, as:

$$\gamma = \frac{3}{4}\sigma_f D \tag{3.31}$$

The strain factor ϵ_{0i}, in Eq. 3.30, is the threshold strain needed for the onset of damage, as reported in Fig. 3.23, associated with the endurance limit S_e.

The integral formulation in Eq. 3.30 can be solved by applying a Ramberg-Osgood type equation [141] at the hysteresis loop for each cycle i, obtaining:

$$\Delta\epsilon_i = \frac{\Delta\tilde{\sigma}_i}{E_i} + 2(\frac{\Delta\tilde{\sigma}_i}{2H_i})^{M_i'} \tag{3.32}$$

where the effective stress range is $\Delta\tilde{\sigma}_i$, E_i is Young's elastic modulus, and the factors H_i, M_i' are the cyclic hardening modulus and exponent, already introduced before. The total strain $\Delta\epsilon_i$, is given by the sum of the elastic deformation $\Delta\epsilon_{e_i}$, the first term of the equation, and the plastic strain range $\Delta\epsilon_{p_i}$ represented by the second term. Their sum $\Delta\epsilon_i$ is also called the cyclic strain range.

It should be noted that all the quantities in Eq. 3.32 carry the subscript i, meaning that they may vary from one cycle to another. However, in the considered case, the stress-strain curve can be considered as stable [141] and the factors E, H, M' can then be considered as constants through the whole fatigue analysis.

Within a single cycle also the lower extremes of the loop $(\epsilon_{min}, \sigma_[min])$ are obviously constant, so it can be written that $d\epsilon = d\Delta\epsilon$ and $dD/d\epsilon = dD/d\Delta\epsilon$ inside the same cycle.

Assuming that $d\Delta\epsilon \simeq d\Delta\epsilon_p$ the damage incremental equation can be expressed as:

$$\begin{aligned}\frac{dD}{1-D} = &\frac{[K'(\Delta\epsilon_p)^{1/M'} - K'(\Delta\epsilon_{p1_i})^{1/M'}]d\Delta\epsilon_p}{[\frac{K'^2}{E}(\Delta\epsilon_p^{2/M'} - \Delta\epsilon_{0_i}^{2/M'}) + \frac{K'}{1+1/M'}(\Delta\epsilon_p^{1+1/M'} - \Delta\epsilon_{p0_i}^{1+1/M'})} \\ &-\frac{K'^2}{2E}\Delta\epsilon_{p1_i}^{1/M'}(\Delta\epsilon_p^{1/M'} - \Delta\epsilon_{0_i}^{1/M'}) - K'\Delta\epsilon_{p1_i}^{1/M'}(\Delta\epsilon_p - \Delta\epsilon_{0_i}) + \frac{3}{4}\sigma_f]\end{aligned} \tag{3.33}$$

Where parameters K', M' are the cyclic hardening modulus and exponent, respectively. The solution of this equation (Eq. 3.33) gives the amount of damage D_i at the end of the i-th cycle, where the initial condition is considered to be $D = D_{i-1}$. In each cycle, the maximum strain range is considered to be given by the plastic strain, therefore $\Delta\epsilon_p = \Delta\epsilon_{pm_i}$ as it can be deducted from Fig. 3.23.

To generate damage inside the material a specific threshold should be exceeded by the selected critical stress.

$$\sigma_{max} \geqslant S_e \tag{3.34}$$

This value, indicated by S_e, is called 'endurance limit' [16] and it is also an internal property of the material, addressing the minimum stress value for the generation of plastic deformation leading to internal degradation. If this condition is satisfied, then the damage variable D_i, is expressed by:

$$D_i = 1 - (1 - D_{i-1})F_i \tag{3.35}$$

Otherwise, if critical stress is lower than the stress threshold, no new damage is generated. The index i referring to the generic i-th cycle and variable F_i is defined as:

$$F_i = \frac{(1+1/M)^{-1}\Delta\epsilon_{0i}^{1+1/M} - \Delta\epsilon_{p1i}^{1/M}\Delta\epsilon_{0i} + C_i}{(1+1/M)^{-1}\Delta\epsilon_{pmi}^{1+1/M} - \Delta\epsilon_{p1i}^{1/M}\Delta\epsilon_{pmi} + C_i} \tag{3.36}$$

where

$$C_i = \frac{3\sigma_f}{4K} - \frac{\Delta\epsilon_{0i}^{1+1/M}}{1+1/M} + \Delta\epsilon_{p1i}^{1/M}\Delta\epsilon_{0i} \tag{3.37}$$

Factors σ_f indicates the true failure stress of the material, M is an exponent for the cyclic hardening and K is evaluated by:

$$K = 2^{1-1/M}H \tag{3.38}$$

where H is the cyclic hardening modulus. All these factors are intrinsic material properties that are also obtained with the simple test already mentioned.

These equations are the solution to Eq. 3.33 and are the final formulation used in the code to evaluate the fatigue damage evolution.

The various $\Delta\epsilon$ variables instead indicate the strain values obtained from a hysteresis loop in the stress-strain curve under a loading cycle (Fig. 3.23), as described by Beheshti and Khonsari [158]. The compressive stresses and the unloading part of the loading history are neglected because are not considered to give a contribution to the damage evolution. Only the loading part of each cycle in which the endurance limit is exceeded is responsible for material degradation. The same assumptions were adopted also by Kachanov [140] and Lemaitre [14].

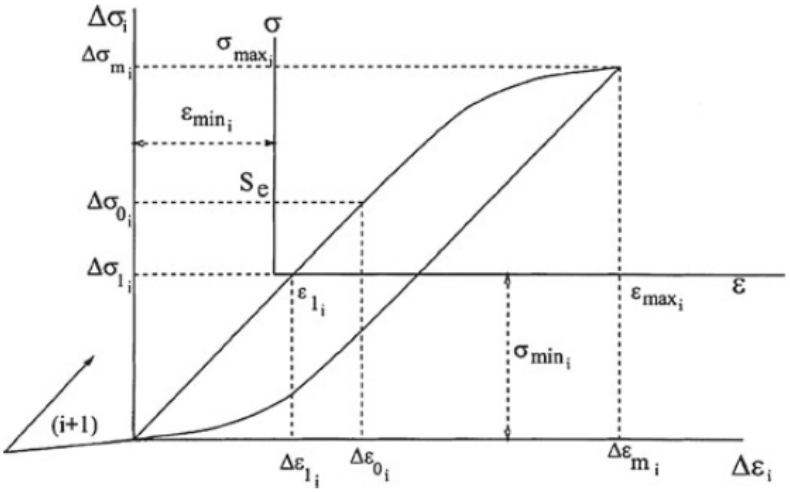


Figure 3.23: *Hysteresis loop of a rolling contact cycle in stress-strain coordinates [158]*

In order to explain the meaning and origin of these deformation factors, Bhattacharya et al [141] proposed a comparison with the empirical Coffin-Manson law. Prediction of fatigue life in a constant-amplitude load regime, controlled by strain, is expressed as:

$$\frac{\Delta\epsilon}{2} = \frac{\sigma'_f}{E}(2N)^b + \epsilon'_f(2N)^c \tag{3.39}$$

where parameters $\sigma'_f, \epsilon'_f, b, c$ are all obtained empirically as material's properties. Also in this approach, the strain amplitude is divided in elastic ($\Delta\epsilon_e/2$) and plastic ($\Delta\epsilon_p/2$), which correspond respectively to the first and second term of Eq. 3.39. By Coffin-Manson's theory, the elastic strain range addresses the fatigue initiation stage, and the plastic strain the propagation one. The prediction is not completely satisfactory

[141], in particular in the low-cycle region associated with fatigue initiation, where instead CDM theory is more accurate.

Anyway, the case considered by the present model is of stress-controlled cycling. Considering the stress range being, $\Delta\sigma_m = \sigma_{max} - \sigma_{min}$, to remain constant over all the fatigue lifespan, a numerical solution of Eq. 3.32 will give the values of the strain ranges, appearing in Eq. 3.36 and 3.37, as function of $\Delta\sigma_m$:

$$\Delta\epsilon_{pm_i} = (\frac{\Delta\sigma_m}{K(1 - D_{i-1})})^M \tag{3.40}$$

$$\Delta\epsilon_{0_i} = (\frac{\Delta\sigma_{1i}}{K(1 - D_{i-1} + \frac{S_e}{K})})^M \tag{3.41}$$

$$\Delta\epsilon_{p1_i} = (\frac{\Delta\sigma_1 i}{K(1 - D_{i-1})})^M \tag{3.42}$$

where all the factors are known.

The maximum plastic strain range $\Delta\epsilon_{pm_i}$ is associated with the maximum stress range. The intermediate limit strain range $\Delta\epsilon_{p1_i}$ is associated with the onset of damage generation and the endurance limit S_e. The smaller strain range parameter $\Delta\epsilon_{0_i}$ is the origin of the reference system for a tension-compression hysteresis cycle. In our case the critical stress τ_{max} has only positive values (Fig. 3.20(b)), not undergoing a complete inversion. $\Delta\epsilon_{0_i}$ is then considered as zero because the compression part of the cycle is not present.

As it can be seen in Eq. 3.40 and 3.42 and 3.41, all the strain parameters are function of the damage D_i at each cycle. Therefore, they change constantly with the material's degradation evolution. This property is fundamental to reproduce the non-linear behavior of damage evolution.

As a matter of fact, it has been proved [159]–[161] that, in industrial machinery components, the major part of the service life may be spent without any relevant effects on the mechanical properties, from damage. This happens because the damage grows with a pronounced accelerating rate in the proximity of the end of life. This is reproduced in the above equations, where the strain ranges undergo an accelerating increment with damage growth.

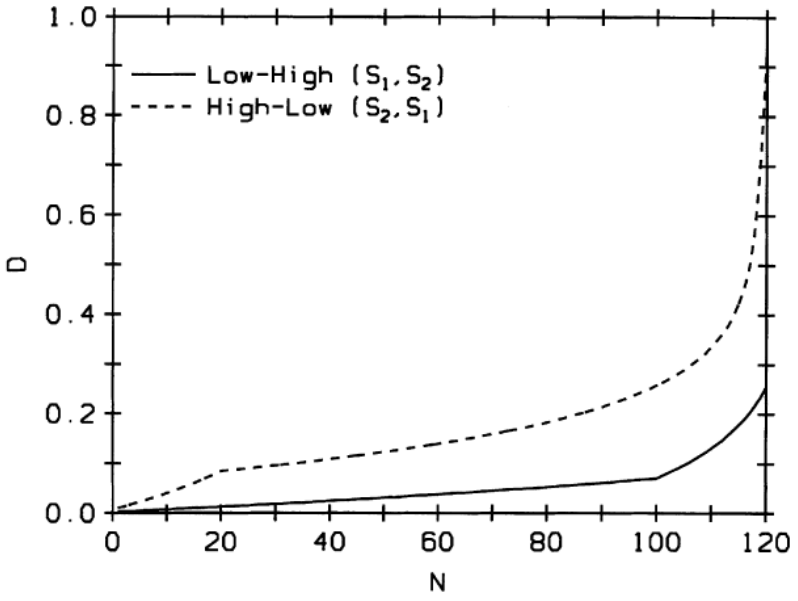
The model's results shown in Fig. 3.25 are qualitatively comparable with the references in Fig. 3.24. The shape is similar and shows a pronounced acceleration in damage evolution rate in the proximity of the crack formation.

In the calculation, an upper limit for damage evolution is set at the critical value D_c for the given material. In this way, the damage increment will continue until the maximum value of the damage vector D, defined as an independent value per each node of the domain, reaches its critical threshold. At that point, the analysis is stopped.

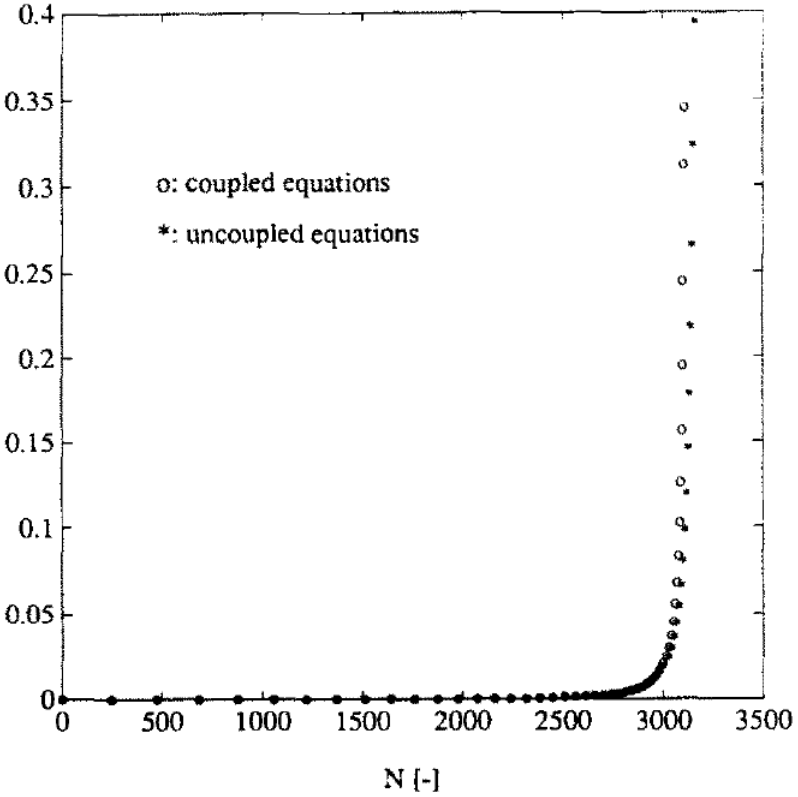
This process is performed on all the one-dimensional domain, in this way, each time the calculation runs, the calculation corresponds to a complete cycle (passage of the load) over the entire domain, because the stress was evaluated step-wise for an entire cycle.

The number of cycles is updated after each complete calculation. Due to the lack of a compressive part in the stress cycle, the parts of Eq. 3.37 and Eq. 3.36 where is present the factor $\Delta\epsilon_{0_i}$ are omitted for sake of simplicity.

The maximum value at each depth of the maximum shear stress distribution (τ_{max}) is collected in a vector, which then identifies the critical stress felt at each position

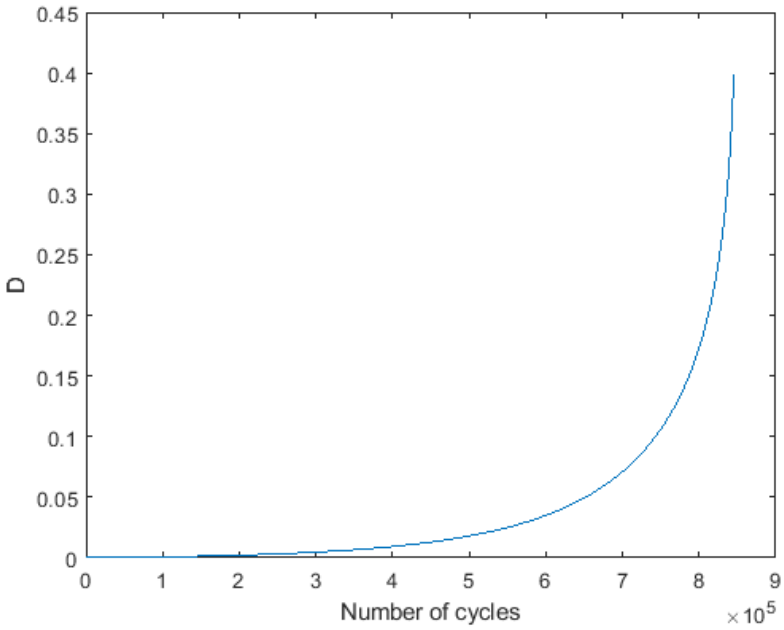


(a)

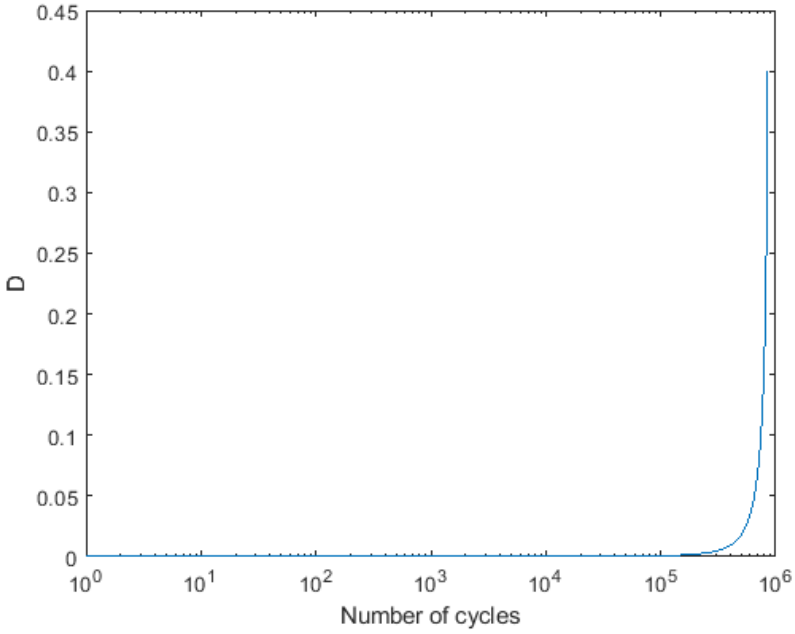


(b)

Figure 3.24: *Qualitative example of damage (D) evolution over the number of cycles for: (a) Bhattacharya [141] and (b) Paas et al. [159]*



(a)



(b)

Figure 3.25: *Plot of damage (D) evolution over the number of cycles, obtained from the model with load* $P = 2560$*, on* $AISI52100$*. (a) linear plot (b) logarithmic scale*

during an entire cycle. The value obtained is assumed as critical stress, at any node, for the damage evaluation, and is then compared with the endurance limit stress (S_e) as a condition to update the damage (Eq. 3.34).

Once the stress endurance limit is exceeded, the damage evaluation for the given node is performed, recalling the formulation expressed in Eqs. 3.40, 3.42, 3.41, 3.37, 3.36, 3.35.

The final result of the evaluation is a number given from the iteration counter, indicating the cycles necessary to reach the critical damage vale in a specific node, which indicates the probable critical position of crack initiation for the given system.

3.9 Input Parameters

The data required for this model are the material properties:

- load half-width, a
- Poisson's ratio, ν
- endurance limit, S_e
- cyclic hardening modulus, H
- cyclic hardening exponent, M
- true fatigue limit, σ_f
- critical damage threshold, D_c
- elastic limit, $\Delta\sigma_l$

The contact half-width, being one of the most important parameters of the whole model, is taken from the main reference [16], with the dimension of $a = 250\ \mu m$. This reference unit measure and dimension is confirmed in all the references treating rolling contact fatigue of bearing gears, and reported in this work, as the general dimension of a bearing contact.

The Poisson's ratio ν, instead is to referred in some passages of the reference paper by Beheshti [16] but is also confirmed in the materials' tables that can be easily found on different sites [162][163], and its value for this alloys is always assessed around 0.3.

The other parameters, derived with a simple tension test, are taken from the reference study made by Beheshti [16], who performed a numerical analysis from which this study has taken its cue and reported data for bearing steels (SAE 4340 and $AISI$ 52100), which are two of the most commonly used steels for bearing gears. In Tab. 3.1 are reported the main properties for both the alloys, used in the model.

It can be noted how the SAE 4340 alloy is the slightly softer material and has a sensibly lower stress endurance limit. This factor results in fatigue at lower stresses. This is the principal discriminating factor to obtain a fatigue behavior which is specific to the material under exam. The stress tensor evaluation is unique, with the friction coefficient μ as the only external variable. The materials' properties has been the basis

Material	E $[GPa]$	H $[MPa]$	M	σ_f $[MPa]$	S_e $[MPa]$	D_c
SAE 4340	192.9	1812	7.1	1911	542	0.46
$AISI$ 52100	206.9	3443	6.22	2586	768	-

Table 3.1: Material's empirical data from Behesti [16]

of the continuum damage mechanics approach and act as a filter for the analysis, with the endurance limit S_e, and also as internal parameter with all the hardening factors.

A brief disquisition should be made on the critical damage threshold value. It can be noted in Tab. 3.1 that there is no critical damage value available for the $AISI$ 52100 alloy. For the analysis of this material, two theoretical values have been employed (0.1 and 0.4). These are the averages of the extreme values for most materials [155][164], obtained according to the "elastic energy equivalence" and "elastic strain equivalence" methods [164]. This uncertainty on the critical damage threshold can be seen as a limit of the sensibility of the model but, as it will be shown in the results section, Chap. 3.10, the span in final fatigue life results is acceptable.

The last evaluated parameter is the elastic stress range $\Delta\sigma_l$ used in the damage evolution section. This parameter has been deducted from the interpretation of the hysteresis loop, in order to satisfy a coherent description of the graph. The value is then assessed as the stress needed to have a deformation of 0.2% in the material, coming from the elastic region described by Young's modulus. The resulting values will be of 385.8 MPa for the SAE 4340 alloy and of 413.8 MPa for the harder $AISI$ 52100 alloy. These values satisfies the conditions depicted in Fig. 3.23 and give coherent results in the damage evaluation.

3.10 Results

3.10.1 Initiation life

The evolution of damage (D) in each element of the material is evaluated from the critical stress profile (τ_{max}) depicted in Fig. 3.21. At each cycle, the damage value D_i is updated in any position where the stress exceeds the material's endurance limit. After updating the strain range values ($\Delta\epsilon_{k_i}$), which are dependent on the damage value at the previous cycle (D_{i-1}), the factors F_i and C_i are calculated. The damage is then updated at the current cycle, and the number of cycles is increased by one after all the domain's depth is spanned.

With this procedure, the first point to reach the critical damage threshold is identified. This position is considered as critical for crack formation in the specific set of conditions (load, friction). It can be deducted easily that the critical point is the one subjected to the highest stress, which according to theory is placed at $z = 0.78a$, around 195 μm. This position remains similar in low-friction conditions and then will change at high frictions, where the surface becomes the most stressed region.

Beheshti [16] evaluated the initiation life at different loads for the $AISI$ 52100 alloy. That analysis (Fig. 3.26) is used to validate the model on the initiation and the numerical results are compared also with the experimental results obtained by Chen et al. [86].

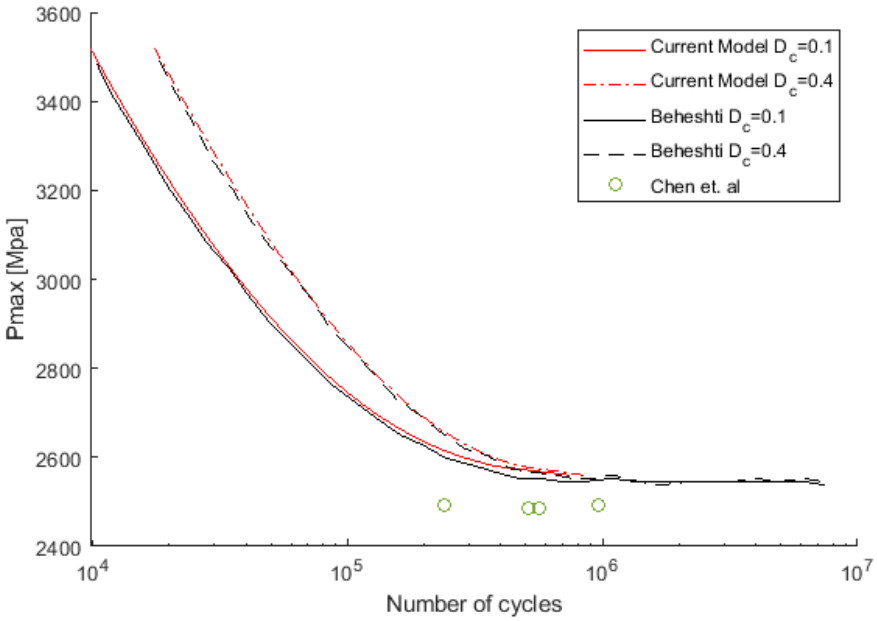


Figure 3.26: *Comparison of number of cycles to initiation for different maximum Hertzian pressure, on AISI* 52100*, between Beheshti and Chen references [16], [86] and model's results*

This study reported results for fatigue initiation life for a $GCr15$ bearing steel, very similar to the analyzed $AISI$ 52100. It can be seen a horizontal asymptotic slightly above the value of 2500 MPa, which is close to the limit of the experimental range measured also by Harris and Barnsby (2450 MPa) [165]. It should be also noted that the identification of the precise location of crack initiation is a complicated task and most of the experimental evidence is obtained after the complete failure of the component, causing a slight imprecision in the measurements.

Anyway, in Fig. 3.26 there's a good agreement between the numerical results and the experimental data. The number of cycles to failure increases when the maximum Hertzian pressure approaches the lower limit ($\sim$ 2560 MPa for $AISI$ 52100) where the maximum value of the critical stress (τ_{max}) approaches the endurance limit. Under this value, the analysis doesn't give any result and the life of the component is considered theoretically infinite.

It should be noted that the limit value is slightly higher than the experimental results obtained by Harris and Barnsby, but with an acceptable tolerance.

The graphs report two lines, corresponding to two different values of the critical damage variable D_c, which is not available in the open literature for the specific material. Therefore, as indicated in the previous section Chap. 3.8, the two limit values of 0.1 and 0.4 are analyzed and compared to identify a generic range in which the real initiation will result.

It can be seen good resemblance with the reference both in shape and values. The number of cycles to crack initiation span from above 10^4, for the higher stress values to a plateau which stabilizes around 10^6 cycles and then extends at higher values approaching the limit pressure. The lack of indications on the friction coefficient in

the reference [16] pushed to perform the analysis at a null friction value, because a better resemblance on stress values was verified in the low friction regime.

The good accordance in these results can be taken as validation of the current numerical model when evaluating the fatigue life of bearing gears.

As indicated by Beheshti [16], this analysis validates the model only in the specific case of roller bearings made of the specific *AISI* 52100 alloy.

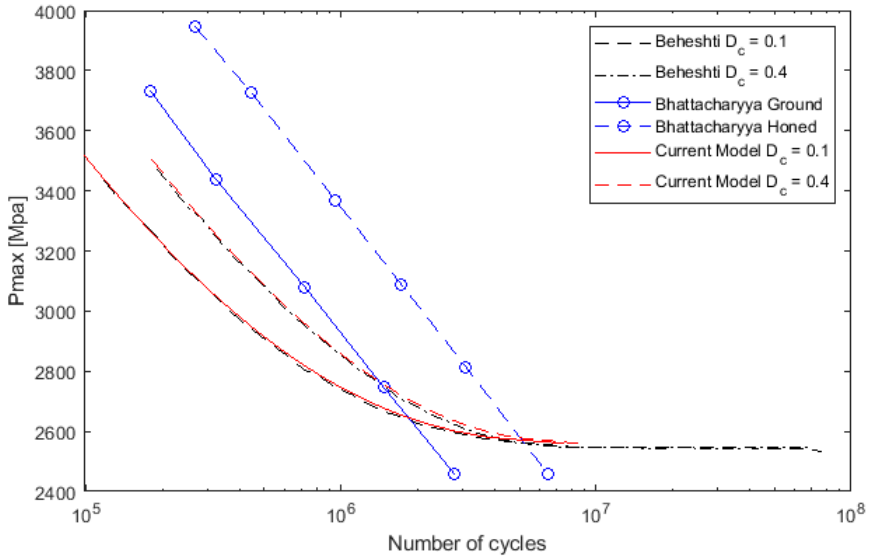


Figure 3.27: *Comparison of number of cycles to failure for different maximum Hertzian pressure, on AISI* 52100*, between Beheshti, Chen and Bhattacharya references [16] and model's results*

Moreover, a prediction of the total fatigue life, for the same specific system, is proposed in the reference study based on Chen [86] assumptions. Here, the crack nucleation stage is considered to cover around the 10% of the total fatigue life. The total life is then evaluated only shifting the curves of the initiation period, previously evaluated, to fit the proportion.

In fact, it can be seen that the shape of the curves is preserved and the only variation is a shift toward a higher number of cycles. This shift is exactly one order of magnitude, corresponding to the assumption of the initiation stage being about the 10% of the total fatigue life. This fitting results to return acceptable results if applied to the previously obtained data for fatigue initiation (Fig. 3.26), now shifted to the higher cycles region (Fig. 3.27).

This assumption made by Chen et al. [86] cannot be taken as a general statement. Numerous factors [54][57] affect the fatigue behavior of roller bearings, including the material properties, the loading conditions and all the operational parameters. The amount of time spent in the nucleation phase could then raise also to the 90% of the entire fatigue life [16], becoming the dominant stage to failure.

Anyways, the focus of this model is on the evaluation of the fatigue initiation step and not to estimate the total life to failure of the material.

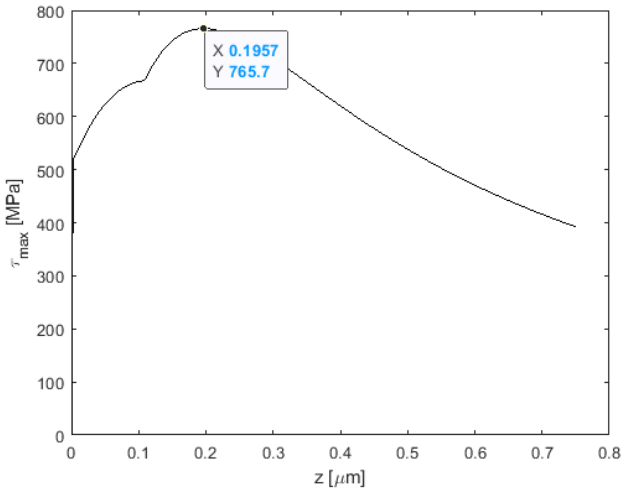
Tab. 3.2 and 3.3 report the resulting data of the failure depths, critical position for

Table 3.2: Number of cycles to initiation and critical depth for different load and friction values, on $AISI$ 52100

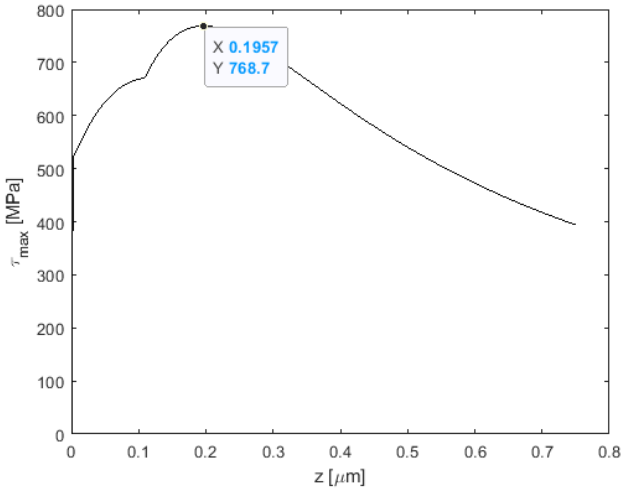
Load $[MPa]$	Friction coefficient, μ	Number of cycles, N	Critical depth $[\mu m]$
2560	0	845'123	197.32
	0.1	844'857	197.32
	0.2	844'591	197.32
	0.3	333'313	1.67
	0.4	24'997	1.67
2750	0	150'747	197.32
	0.1	150'744	197.32
	0.2	150'742	197.32
	0.3	120'881	1.67
	0.4	14'443	1.67
3000	0	63'544	197.32
	0.1	63'543	197.32
	0.2	63'542	197.32
	0.3	53'526	1.67
	0.4	7'548	1.67
3250	0	32'591	197.32
	0.1	32'591	197.32
	0.2	32'591	197.32
	0.3	27'859	1.67
	0.4	4'193	1.67

crack nucleation, and the number of cycles to initiation for different loading and friction conditions. The critical damage value used for the $AISI$ 52100 alloy is $D_c = 0.4$ in this case. Because more similar to the value of 0.46 reported for SAE 4340.

It is evident how the shift in the critical position happens at $\mu = 0.3$ and also a sensible decrement in the initiation life is registered for $\mu = 0.4$. This case is coherent with the stress profile shown in Fig. 3.9, where the stress at the surface for the maximum friction exceeds the peak at depth $z = 0.78a$ and results in a sensibly shorter life.



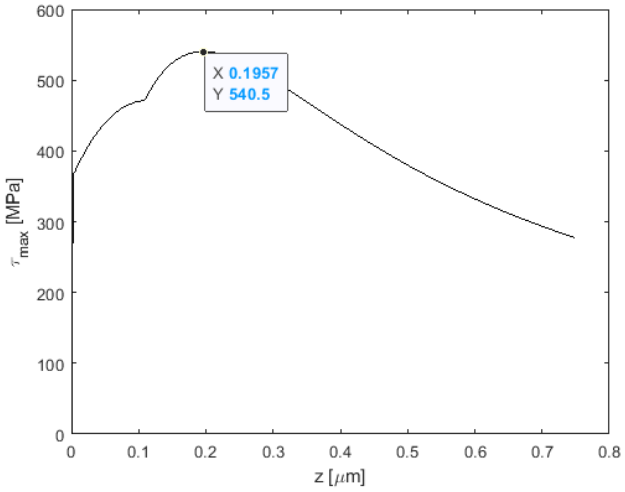
(a)



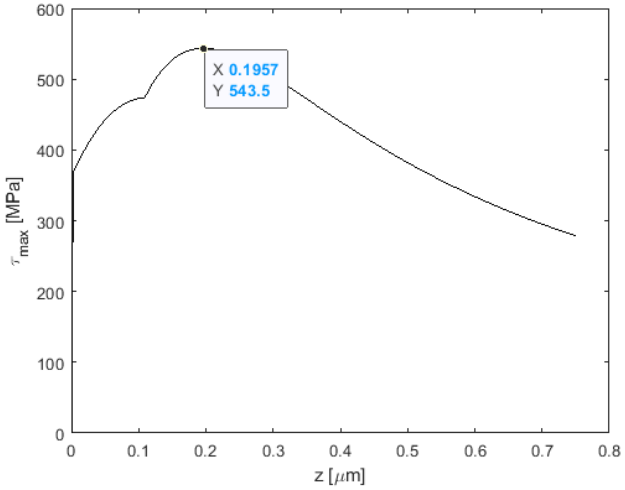
(b)

Figure 3.28: *Resulting maximum critical stress profile (τ_{max}) and peak value for AISI 52100 at (a)* 2550 MPa*, (b)* 2560 MPa*.*

This trend can also be seen in Fig. 3.30, where is shown the progressive initiation life decrement with friction and load. Obviously, the SAE alloy is subject to lower



(a)



(b)

Figure 3.29: *Resulting maximum critical stress profile (τ_{max}) and peak value for SAE* 4340 *at (a)* 1800 MPa*, (b)* 1810 MPa*.*

stresses, because of the smaller endurance limit of the material. The progression is then similar, although the weaker material is more sensitive to the load increment and has a fast life decrement from $5 * 10^5$ to 10^4 with only 400 MPa added. The stronger $AISI$ 52100 alloy instead decays to an initiation life of 10^4 in almost 800 MPa, so double with respect to the SAE 4340 alloy.

Table 3.3: Number of cycles to initiation and critical depth for different load and friction values, on SAE 4340

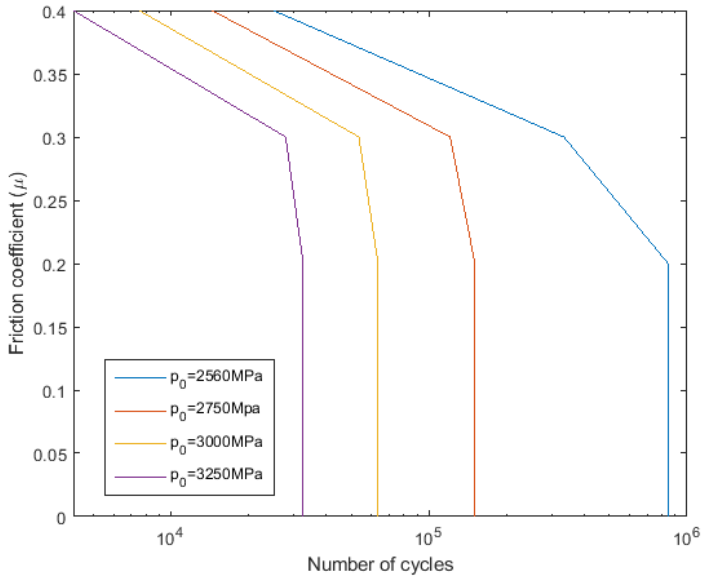
Load $[MPa]$	Friction coefficient, μ	Number of cycles, N	Critical depth $[\mu m]$
1810	0	509'695	197.32
	0.1	509'623	197.32
	0.2	509'552	197.32
	0.3	239'843	1.67
	0.4	15'241	1.67
1950	0	102'427	197.32
	0.1	102'425	197.32
	0.2	102'423	197.32
	0.3	81'360	1.67
	0.4	8'120	1.67
2200	0	30'050	197.32
	0.1	30'050	197.32
	0.2	30'049	197.32
	0.3	25'149	1.67
	0.4	3'000	1.67
2400	0	14'006	197.32
	0.1	14'006	197.32
	0.2	14'006	197.32
	0.3	11'837	1.67
	0.4	1'476	1.67

The initiation life approaches values of 10^6, when the resulting stress approaches the endurance limit. The load range for both the alloys has been chosen by selecting the minimum load at which the critical resulting stress will exceed the endurance limit ($S_e = 542\ MPa$ for SAE 4340 and $S_e = 768\ MPa$ for $AISI$ 52100). In Fig. 3.28 and Fig. 3.29 are shown the maximum loads at which the resulting τ_{max} doesn't generate any damage, because lower of S_e. Are also shown the chosen values, slightly exceeding the threshold.

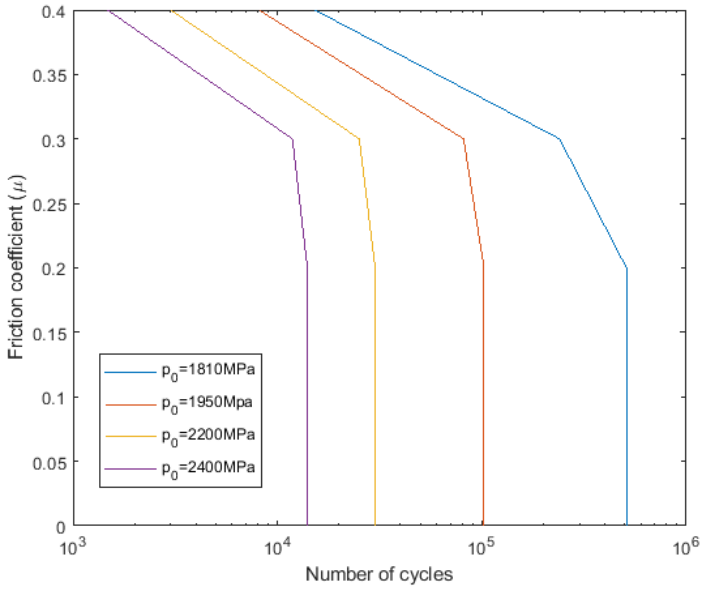
The load progression on the two materials has then been chosen to obtain a similar decrease in the order of magnitudes, showing the different sensibility and strength of the alloys.

3.10.2 In depth analysis

In order to evaluate the behavior in a complex case, as the one presented by the Airbus case [7], an in-depth analysis can be relevant. The evolution of damage at different depths can be the prologue to secondary crack formation. By the analysis of the damage evolution time in different zones of the material, it can be stated which of them are more or less prone to failure. Even if the crack propagation phase is controlled from different mechanisms with respect to the initiation, the presence of damage can act as an enhancer of the fatigue failure process.



(a)



(b)

Figure 3.30: *Representation of data from Tables 3.2 and 3.3. Number of cycles at failure for different maximum Hertzian pressure and friction coefficients, on (a) AISI* 52100 *and (b) SAE* 4340

The analysis is aimed at evaluating the initiation life for the entire z domain and the process follows the same structure used before for the critical node and initiation life evaluation. The stress tensor is evaluated through all the domain and the critical stress tensor is assumed to be the maximum shear stress (τ_{max}). Once this tensor is obtained, the maximum value at each depth is collected in a tensor (max_t) indicating the maximum stress acting on any node of the 1D domain, during each cycle.

For each node where $max_t > S_e$, the initiation life is evaluated and stored. So, the number of cycles to initiation is now a vector with the same dimension of the domain, collecting the cycles to failure of all the nodes, when in the standard analysis it was a simple number.

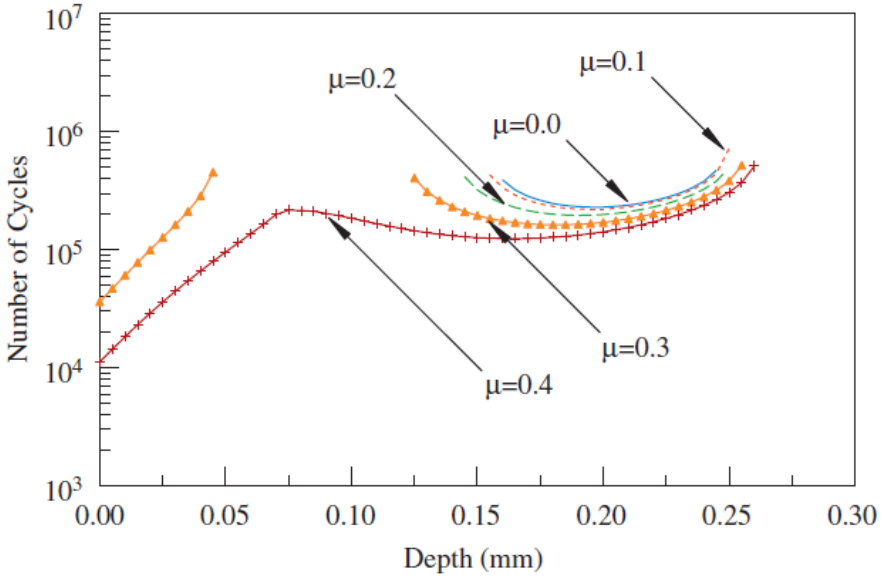


Figure 3.31: *Number of cycles to initiation at different depths, for* $P_max = 1800\ MPa$. *Reference values from Behesti [16]*

The results have been compared with the reference analysis from Beheshti [16], which considered the SAE 4340 alloy under a pressure of 1800 MPa (Fig. 3.31). Recalling the considerations on the minimum load able to generate damage in the current model, some consideration must be done. It is impossible, in this analysis, to use 1800 MPa because it has been shown that no damage will be generated. The minimum value of comparison used has been of 1810 MPa, which is the minimum pressure able to start the damage evolution process in the current model.

The results obtained using the lowest load are not completely satisfactory in the extension of the damaged zone (Fig. 3.33).

To obtain a more fitting comparison, the Hertzian applied pressure has been increased to 1850 MPa and the result is shown in Fig. 3.32. Here, the correspondence between the curves is clear when friction is low and good results are obtained. The divergence at higher friction values, shown previously, makes the correspondence fade. This effect can be attributed to a slight difference in the resulting critical shear stress, due to a difference in the effect of friction from the current model and the reference [16]. So, due to slight differences in the stress distribution, the correspondence is

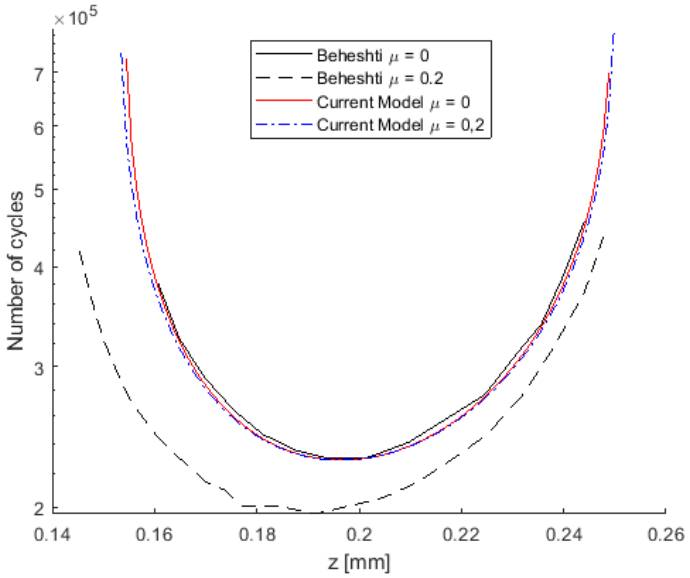


Figure 3.32: *Comparison of number of cycles to initiation at different depths between Beheshti [16] reference curve and the current model*

shown at not identical, but close, Hertzian pressure values (1800-1850 MPa).

As seen in Fig. 3.9, the peak values in the critical zone are coincident, but a slight difference in the distribution in non-critical zones is present. Anyway, the difference is acceptable and reduce in particular for low friction regimes, which are more relevant for the studied case.

Anyway, the friction affects only slightly the behavior around the critical position of 195 μm, as stated in the reference. A small reduction in the initiation life is present when increasing the friction coefficient, not varying sensibly the value of the number of cycles.

Despite the lack of a complete correspondence in the friction behavior, the evaluation of the fatigue initiation life results to be coherent in the critical and in the general value of the number of cycles. A more close correspondence is present at low friction coefficients, which are more interesting for the chosen application.

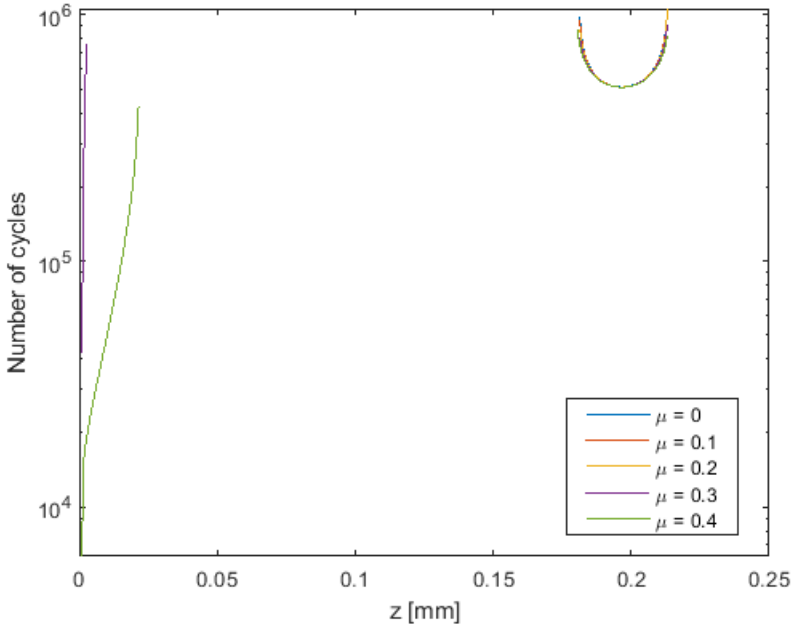
Due to the problem of crack propagation toward the bulk of the material, the evolution of damage in deep regions represents a critical situation. it can be seen how crack can be initiated at a depth of almost 300 μm. So, the propagation can start at an already deep point, becoming dangerous if penetrating further in the material.

A further fatigue propagation study may assume this depth value as a depth limit initiation spot. Usually, the thermally treated region has a depth of a few millimeters, whereas grinding process generates a treated layer of few tenths of microns ($\sim 20\mu m$ [7]). Once the crack propagates toward the limits of this region, a catastrophic failure can occur due to the change of crack propagation method and speed.

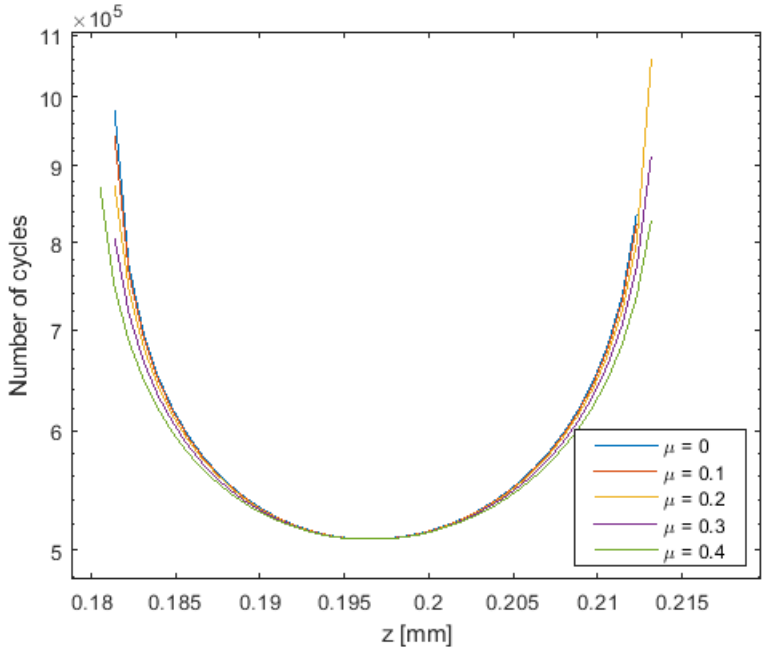
Useful information on the initiation life behavior at different depth has been obtained on a defect-free and untreated material on the SAE 4340 alloy (Fig. 3.33 and 3.34). A similar analysis has been performed on the more resistant $AISI$ 52100 bearing steel (Fig. 3.35 and 3.36). Also here can be seen a small damage formation for the lower stress, which can be attributed to a sharper stress distribution. Although,

the behavior for a maximum Hertzian pressure of 2700 MPa resembles the behavior shown in the reference, even if for a different alloy.

The considerations made on the slight difference due to the effect of friction can be still considered valid also for this second alloy, being the stress evaluation process not changed. The shape of the curves, indicating the different number of cycles to initiation through all the domain's depth, appears to occur at higher stresses for the more resistant alloy. The shape and values seem similar and comparable.

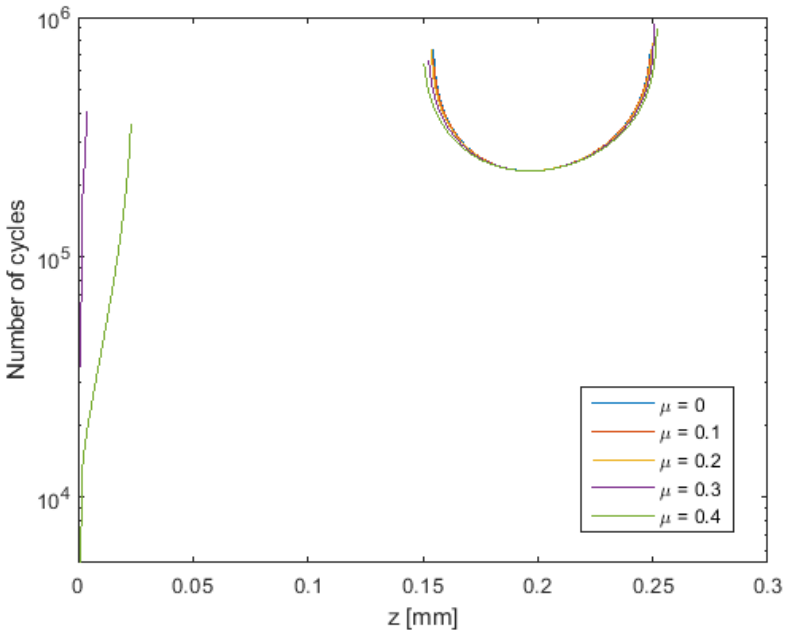


(a)

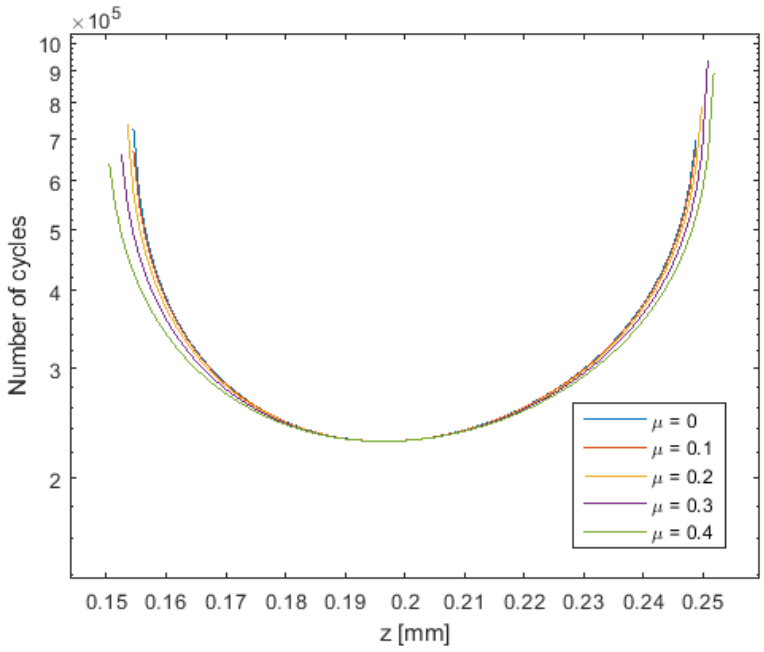


(b)

Figure 3.33: *Number of cycles to initiation at different depths for a pressure of 1810 MPa on SAE 4340. (a) general view, (b) detail on critical region.*

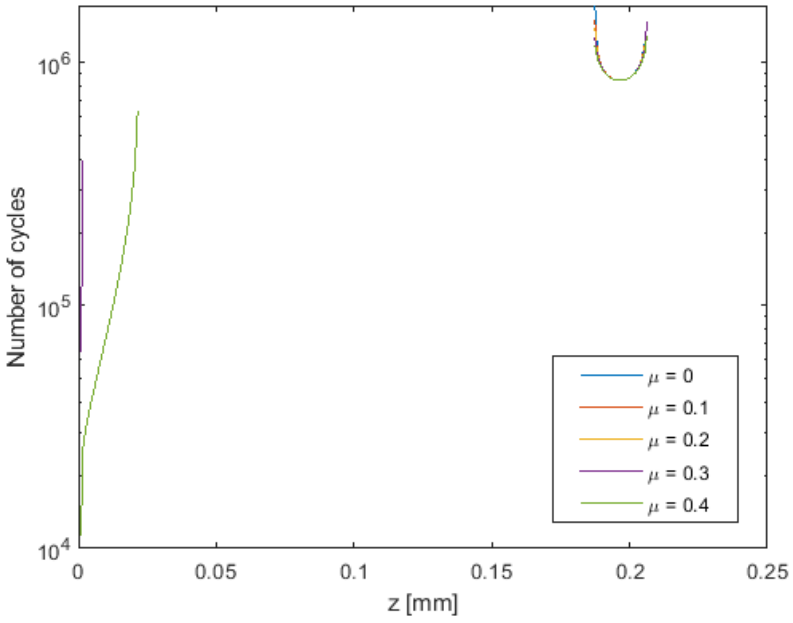


(a)

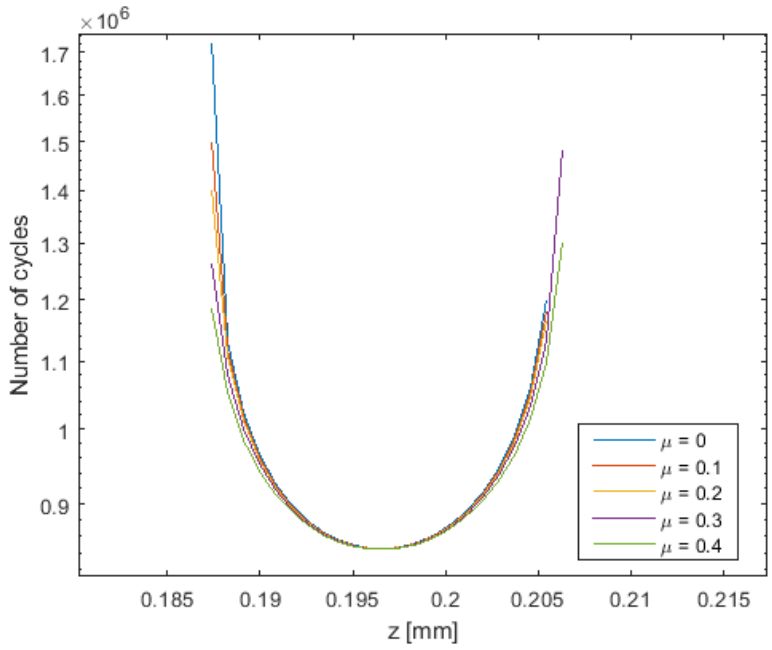


(b)

Figure 3.34: *Number of cycles to initiation at different depths for a pressure of 1850 MPa on SAE 4340. (a) general view, (b) detail on critical region.*

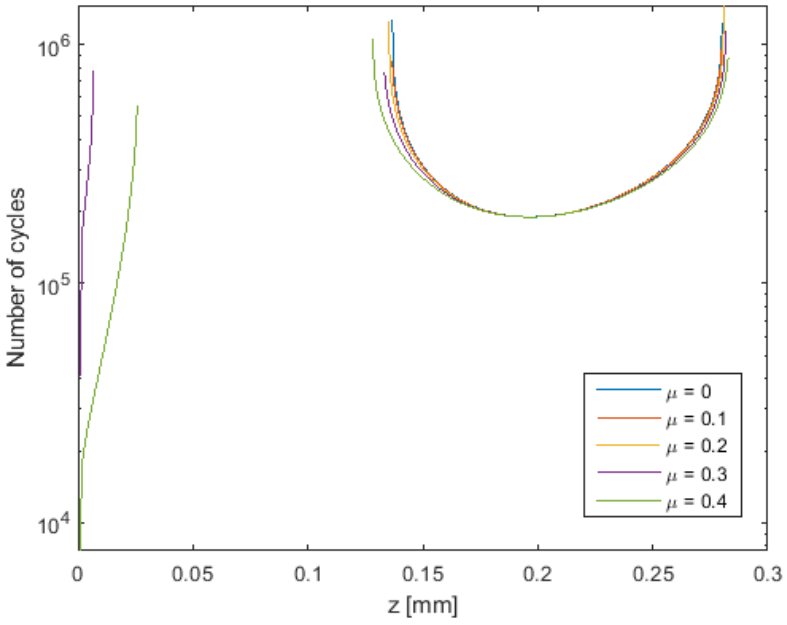


(a)

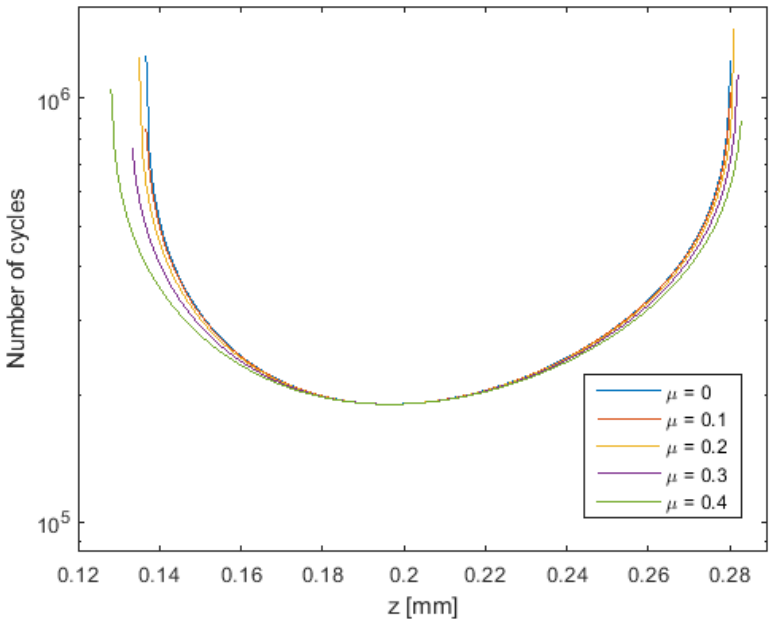


(b)

Figure 3.35: *Number of cycles to initiation at different depths for a pressure of 2560 MPa on AISI 52100. (a) general view, (b) detail on critical region.*



(a)



(b)

Figure 3.36: *Number of cycles to initiation at different depths for a pressure of 2700 MPa on AISI 52100. (a) general view, (b) detail on critical region.*

3.11 Residual stresses

Hardening mechanical processes such as grinding or thermal treatments like carburizing are widely used in the production of bearing steels. Due to severe wear phenomena and high mechanical stresses produced in the near-surface region, hardening is aimed at protecting and increasing the performances of this part of the component. Both mechanical and thermal treatments can be performed on a usual bearing steel, but mechanical grinding has a much shallower penetration, as indicated before, and is then neglected.

The most important treatment considered in this specific application is the carburizing process. The resulting stress field profile has been evaluated in many different ways [109]–[111], and has been schematically represented for the specific bearing gear component in the study from Depouhon [7](Fig. 3.37).

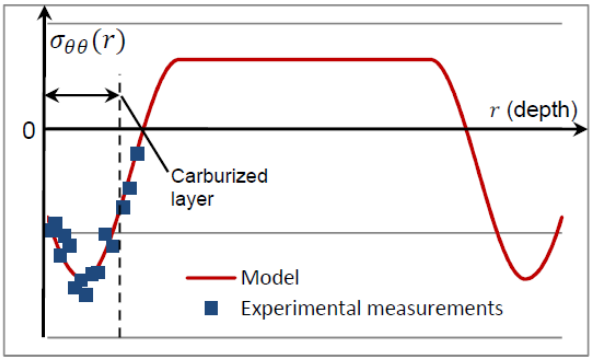


Figure 3.37: *Simplified residual stress profile, from experimental X-ray diffraction data [7]*

As an approximation, the compressive field may be assumed as constant in the analysis domain, which has an extension of 300 μm. The critical depth has been found to be $<$200 μm and then even smaller. The grinding process, indicated by Depouhon [7], gives a peak of residual stresses of around 1000-1100 MPa. This peak is located at a depth of around 20-50 μm, also far from the critical depths for low friction rolling contact.

According to Korsunsky et al. [166], the average value of the compressive field for bearing gears is found to assess around 600 MPa, extending even to a depth of 2-3 mm (Fig. 3.38), confirming the previous dimensional assumptions. Also in this study, some peaks at around 1000 MPa were found at shallow depths, endorsing the choice to neglect them. In other references by Hizli [111] and Reti [110], the levels of residual stresses in steels after carburizing are assessed around values of 400-600 MPa, in the order of magnitude of the stress to be counteracted [16].

The value for the critical stress field in this model referred to the analysis of the $AISI$ 52100 alloy, has been of 600 MPa and of 768 MPa, taken equal to the endurance limit of the material.

This constant field value has been added as a stress component in the evaluation of the tensile stress (σ_x), as indicated in the Eq. 3.8, recalled here.

$$\sigma_x = -\frac{2z}{\pi}\int_{-b}^{a}\frac{p(s)(x-s)^2 ds}{[(x-s^2)+z^2]^2} - \frac{2}{\pi}\int_{-b}^{a}\frac{q(s)(x-s)^3 ds}{[(x-s)^2+z^2]^2} + \sigma_\Theta \tag{3.43}$$

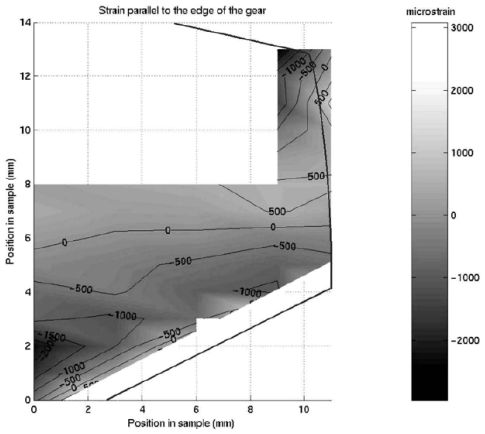


Figure 3.38: *Residual stresses in induction-hardened gear teeth mapped by neutron diffraction [166]*

The last term (σ_Θ) indicates the local value of the residual stress field, which as a first approximation has been assumed to be constant.

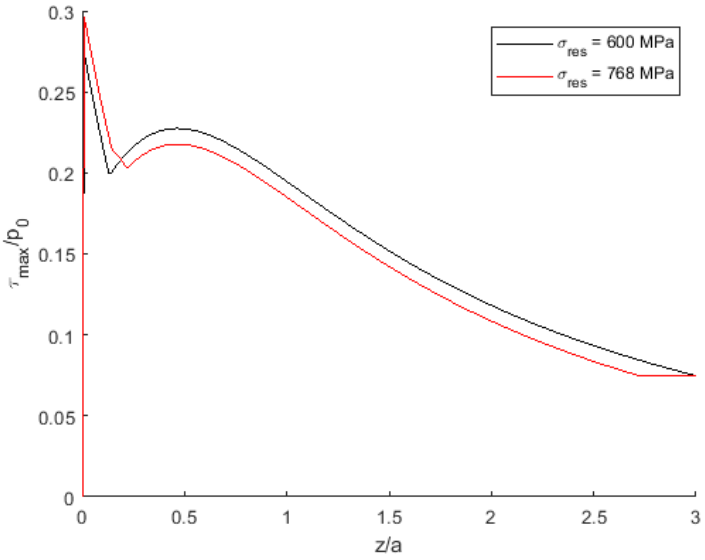


Figure 3.39: *Critical stress depth distribution after applying a constant compressive stress field of 600 MPa and 768 MPa, from a maximum Hertzian pressure of 2600 MPa, evaluated with τ_{max} as critical stress*

It can be seen in Fig. 3.39 that the whole distribution is lowered, with peaks at $<0.2\ p_0$, whereas the natural peak for the critical stress distribution is assessed at $0.3\ p_0$. Also, the position of the peak is shifted toward the surface, at an approximate value of $0.5a$, instead of the canonical $0.78a$ depth [5].

Moreover, an important peak in the proximity of the surface is generated, which

becomes dominant when increasing the value of the residual stress field (Fig. 3.39(b)). This stress distribution can be acceptable if looking at the reduction of the general stress intensity, but the presence of such a peak in the near-surface region can generate critical phenomena in that zone.

If considering the τ_{xy} factor as critical shear stress, the behavior changes sensibly. This maximum shear value showed good accordance (Fig. 3.14) on the principal peak with the τ_{max} evaluation method. The behavior of the critical stress distribution varies instead in a sensible way when applying residual stresses to the tensor evaluation (Fig. 3.40).

Differently from the previous results (Fig. 3.39), the principal peak is still placed at a depth of around $0.78a$ and the peak appearing in the near-surface region is less pronounced and the total value of the stress undergoes a more relevant overall decrement.

This second analysis seems much smoother and acceptable, due to the general trend of the stress distribution which decreases according to the purpose of the treatment. Moreover, the result shown in Fig. 3.39 suggests an increased value of the critical stress at the surface, exceeding the value of 0.3 p_0 obtained without the treatment. This obviously is unacceptable, because the residual stresses are generated to protect the component and decrease the stress felt by the material.

In conclusion, due to the better resemblance with the theoretical prediction of a reduction of the stress in the material, the maximum shear stress has been evaluated directly with τ_{xy}, with the formulation introduced previously (Eq. 3.21).

Still, the constant distribution of the residual stresses generates a critical peak in the proximity of the surface, which results in undesired criticality.

Maximum field value [MPa]	Layer depth [mm]	a	b	c
542	2.5	0.22067	-0.411675	-0.35
600	1	1.49	-1.09	-0.4
600	2.5	0.23885	-0.437125	-0.4
768	1	1.9434	-1.4434	-0.5
768	2.5	0.3109	-0.57735	-0.5

Table 3.4: Plotting coefficients of the parabolic residual stress distribution

To overcome this problem, a parabolic shape recalling the evaluations shown in Fig. 3.37 has been modeled. Keeping the peak values of 600 and 768 Mpa and of 400-500 MPa at the surface, chosen arbitrarily from the qualitative indications given in the graphs [7][16]. The resulting residual stress distributions are reported in Fig. 3.44, where the coefficients of the parabola have been derived imposing the curve to pass through the imposed points at the surface (400-500 MPa), at the peak (600-768 MPa) and at the limit of the carburized layer (1-2.5 mm).

The first line of Tab. 3.4, with a load equal to the endurance limit (S_e) of the SAE 4340 alloy, is obviously referred to that material, whereas the other lines refer to the more resistant $AISI$ 52100 alloy.

In Fig. 3.44 the y-axis indicates the intensity of the stress field and is indicated in GPa for the sake of simplicity, whereas the x-axis indicated the depth in mm. Obviously, the larger distributions (Fig. 3.44(b),(d)) show a smoother progression which results in less pronounced peaks in the critical stress distribution. The resulting effect on the tau_{max} critical distribution for both is shown in Fig. 3.43 and 3.41.

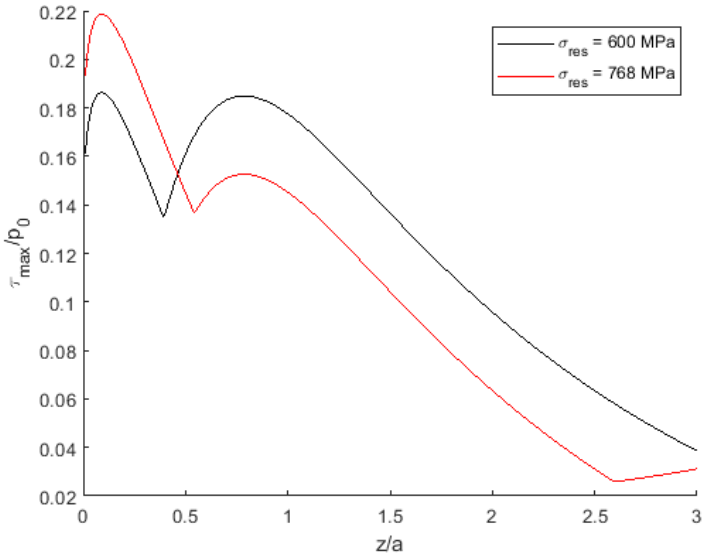


Figure 3.40: *Maximum shear stress depth distribution after applying a constant compressive stress field of 600 MPa and 768 MPa, from a maximum Hertzian pressure of 2600 MPa, evaluated with τ_{xy} as critical stress*

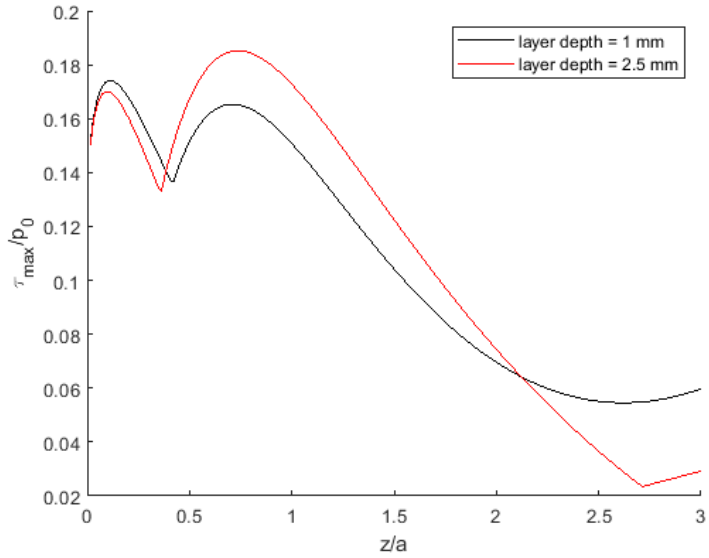


Figure 3.41: *Critical stress depth distribution after application of residual stress fields $\sigma_{res,max} = 768$ MPa with 1 mm depth and $\sigma_{res,max} = 768$ MPa with 2.5 mm depth, from an applied pressure of 2600 MPa*

As already anticipated, the intensity of critical stress is decreased due to the presence of the residual stress field. This acts as protection from fatigue. Also, a second peak

appears in the near-surface region, becoming more pronounced when increasing the maximum value of the residual stresses. When the carburized layer is distributed over a depth of 2.5 mm [166], this peak remains under the principal peak, given from the critical shear stress.

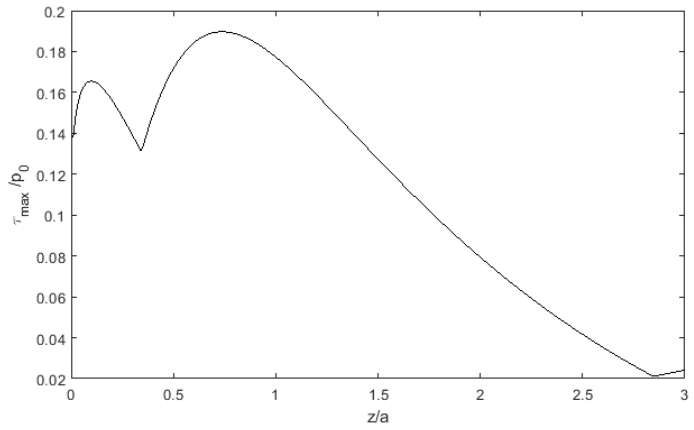


Figure 3.42: *Critical stress depth distribution after application of a residual stress field of* $\sigma_{res,max} = 542\ MPa$ *with 2.5 mm depth, from an applied pressure of 1900* MPa

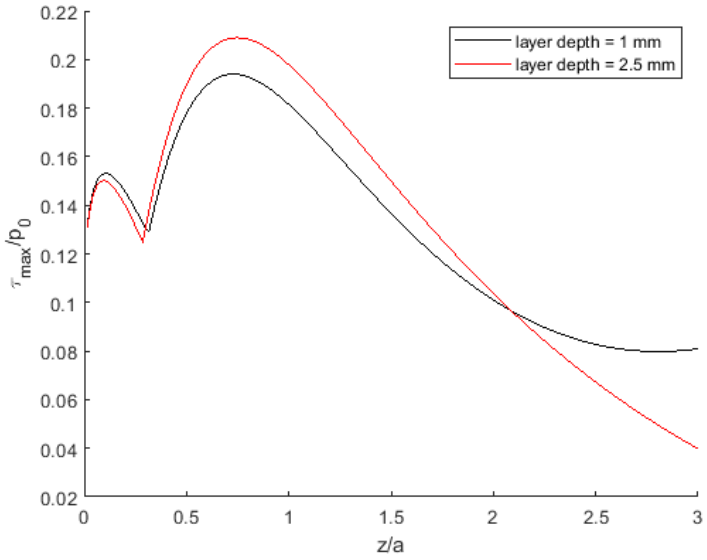


Figure 3.43: *Critical stress depth distribution after application of residual stress field of* $\sigma_{res,max} = 600\ MPa$ *with 1 mm depth and* $\sigma_{res,max} = 600\ MPa$ *with 2.5 mm depth, from an applied pressure of 2600* MPa

The maximum stress value is reached at a critical depth which is slightly lower than the $0.78a$ standard depth obtained in the previous chapters, but still similar ($\sim 0.75a$). It can be stated that a deeper penetration of the carburized layer is then

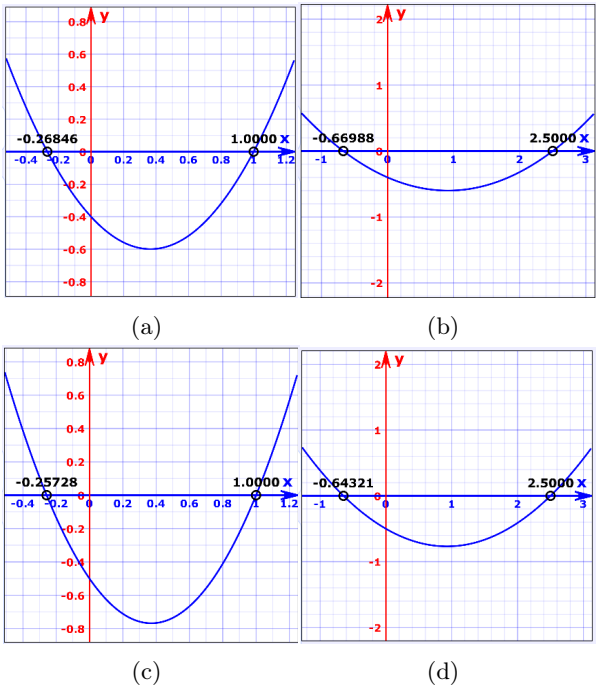


Figure 3.44: *Schematic representation of the parabolic residual stress distribution in the compressive region for: (a)(c)* $\sigma_{res,max} = 600\ MPa$*, (b)(d)* $\sigma_{res,max} = 768\ MPa$

preferable and safer for fatigue, because of the general reduction of the critical stress and absence of a dominant near-surface peak. For this reason, the following analysis on initiation life has been carried on taking into account the wider distribution for both the maximum given stresses.

The values chosen for the compressive residual stresses field are:

- 542 MPa for the SAE 4340 alloy ($= S_e$)
- 600 MPa for the $AISI$ 52100 alloy

due to the consideration of the emerging peak in the near-surface region that is produced at higher values of residual stress. For the same reasons, the carburized layer depth is assumed to be of 1 mm, to guarantee a smoother distribution and avoid peaks. Although it may seem safer to have a higher penetration of the treatment, with this distribution shape, the critical stress region is better covered with a shallower distribution, giving a higher decrease in the stress felt by the material (Fig. 3.43)

Considering the presence of the residual stress field, the loads able to generate damage should be higher than on the untreated material. The chosen values of residual stress for this analysis are the one reported in Tab. 3.4, lines 1 and 3.

The effect on the maximum value of the critical shear stress for both the materials is reported in Fig. 3.45. The stress endurance limit is reached at quite higher values in both cases. For SAE 4340, stress of 544 MPa is reached for a Hertzian pressure of 2637 MPa. So, a minimum Hertzian pressure threshold capable to generate damage is assessed at 2635 MPa for the sake of simplicity.

In the case of the $AISI$ 52100 alloy instead, a value of 768.4 MPa is generated by a pressure of 3480 MPa. For the same previous reasons, the minimum pressure threshold is fixed at 3480 MPa.

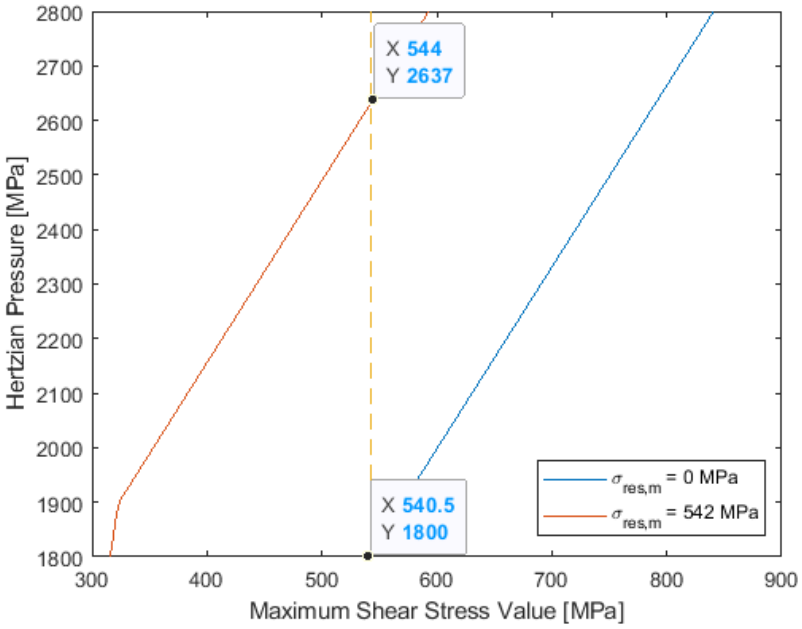
The analysis of fatigue initiation as a function of friction and depth can then be repeated, taking into account the effect of residual stresses, which operates a shift toward higher Hertzian pressures, but not changing the results in the value of the initiation life given by this model. The evaluation of the in-depth damage evolution has been performed for both the alloys for representational purposes. It should be noticed that this approximation doesn't take into account the specific material behavior and phase changes happening during carburizing and thermal treatments.

The resulting plot resembles obviously the one shown in Fig. 3.33 and 3.35. Also in this case (Fig. 3.47 and 3.48) a second plot with increased Hertzian pressure is presented, in order to show more comparable curves with the Beheshti [16] reference (Fig. 3.31). The details show also in the case of applied residual stresses, a similar behavior. The increase expected from the friction effect is still limited, but the low friction regime, considered for fatigue applications, remains acceptable.

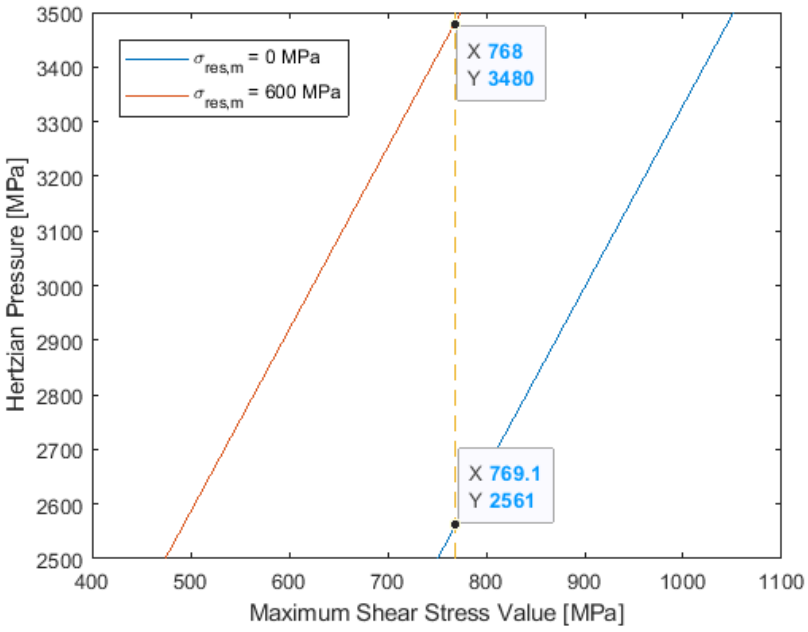
The main effect of the residual stresses clearly to increase sensibly the required pressure in order to start the damage generation, but there is also the formation of more superficial damage at lower friction values. This effect can be negative if seen from a neutral point of view, but it can also be useful. If superficial spalling is generated, in-depth crack propagation can be prevented because of the usual particle detection methods that are placed around the bearing gears.

In conclusion, the carburizing process is beneficial for fatigue because of the ability to reduce the critical stress felt by the material and by the compressive field applied in the crack formation zone, slowing further propagation.

In order to have a precise evaluation of the effects on the specific material, a direct experimental analysis on the component's internal stress profile should be performed.

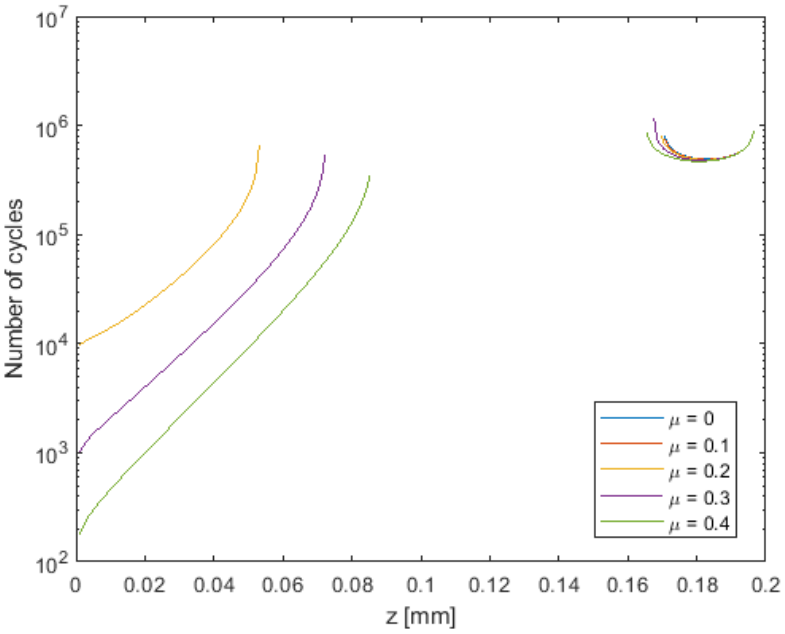


(a)

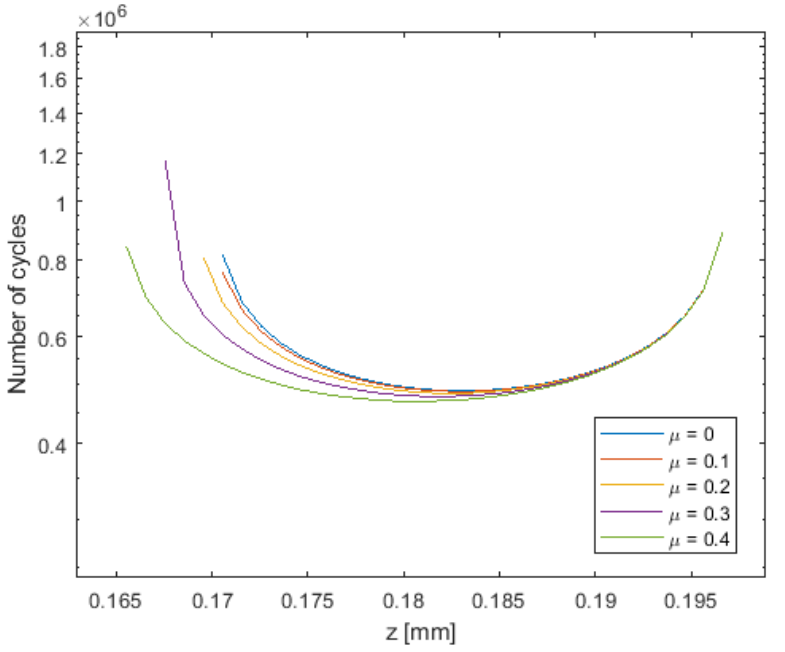


(b)

Figure 3.45: *Maximum critical stress value progression with load, after applying the residual stresses field with respect to: (a) SAE* 4340 *threshold ($S_e = 542\ MPa$), (b) AISI 52100 threshold ($S_e = 768\ MPa$)*

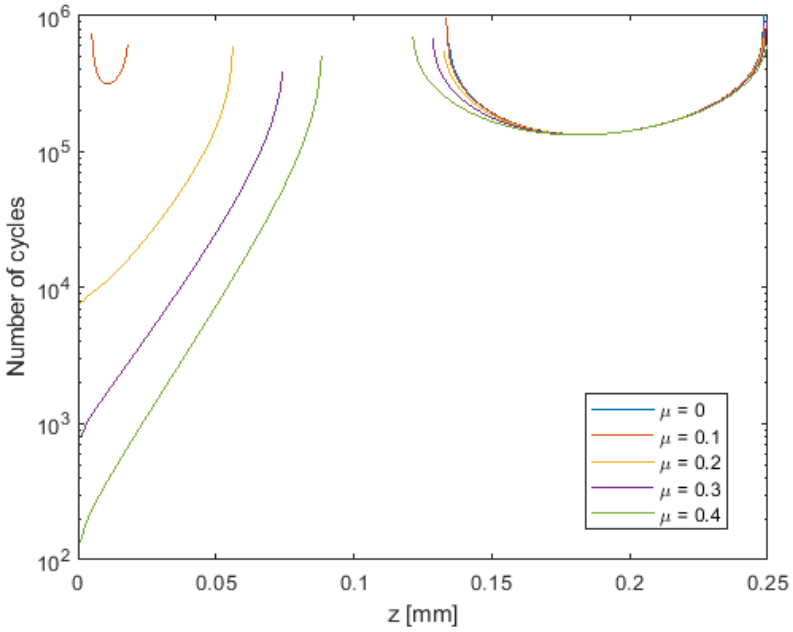


(a)

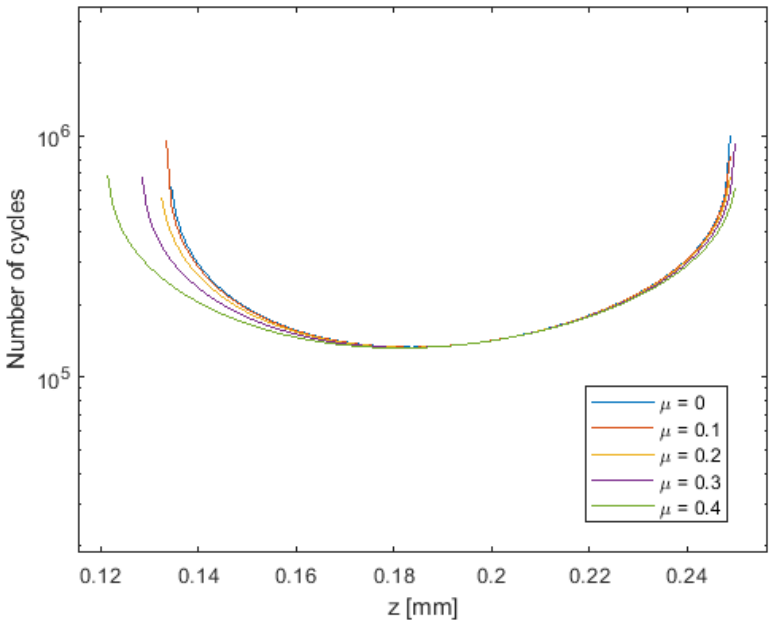


(b)

Figure 3.46: *Number of cycles to initiation at different depths, for* $P_max = 2635\ MPa$ *on SAE 4340. (a) general view, (b) detail on critical zone*

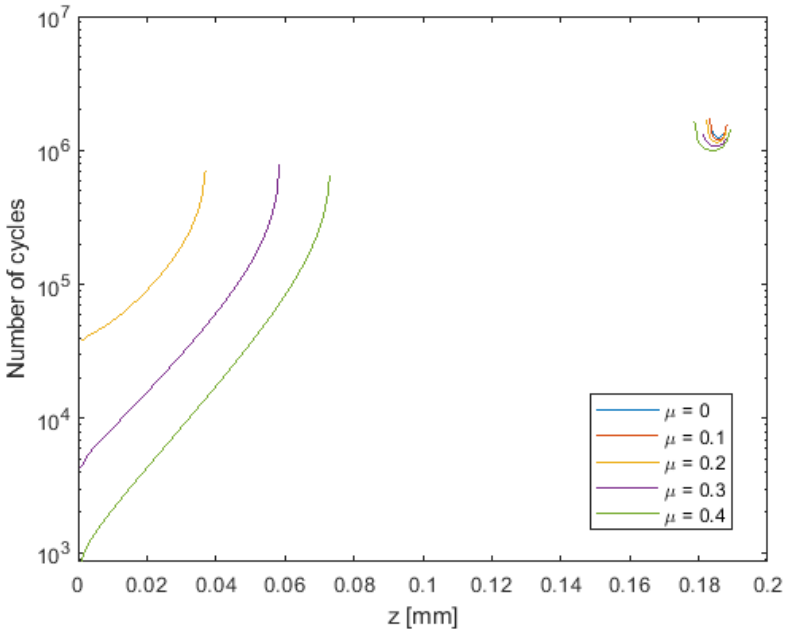


(a)

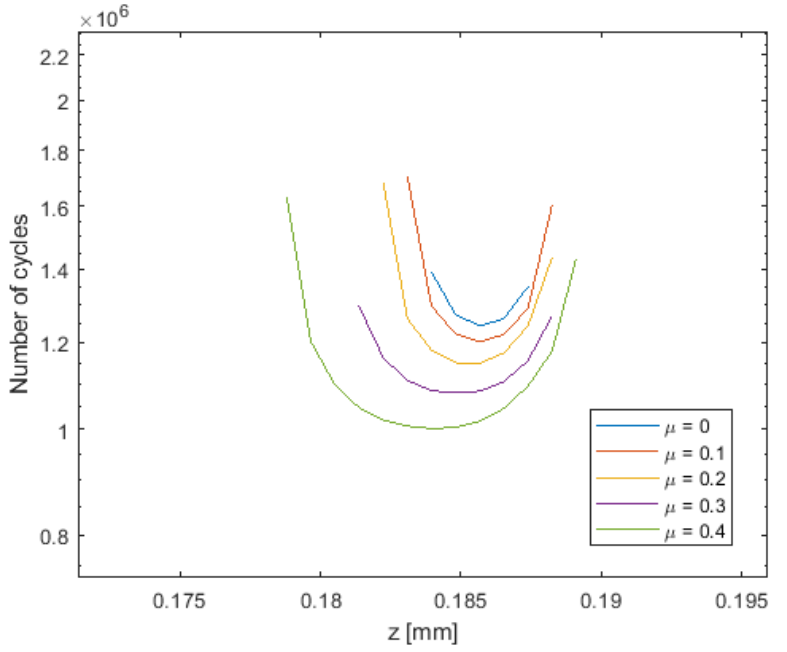


(b)

Figure 3.47: *Number of cycles to initiation at different depths, for* $P_max = 2735\ MPa$ *on SAE 4340. (a) general view, (b) detail on critical zone*

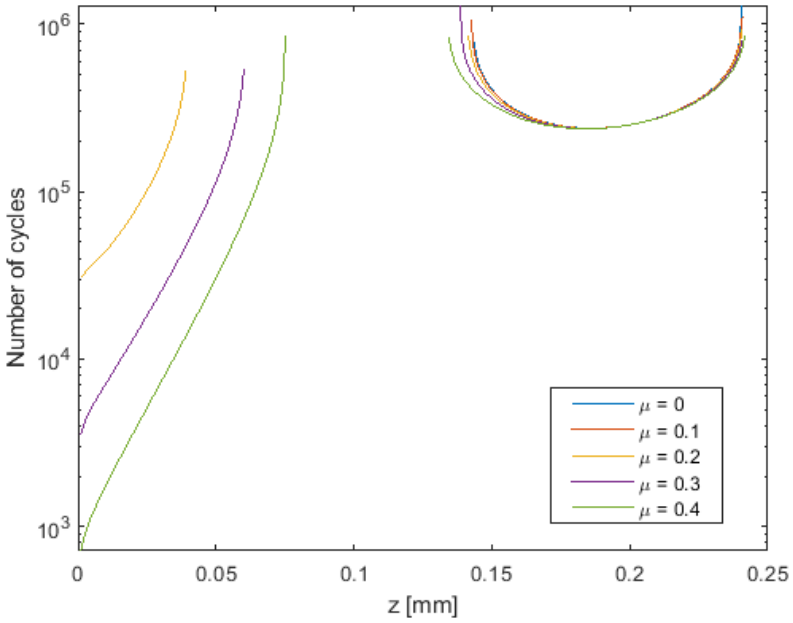


(a)

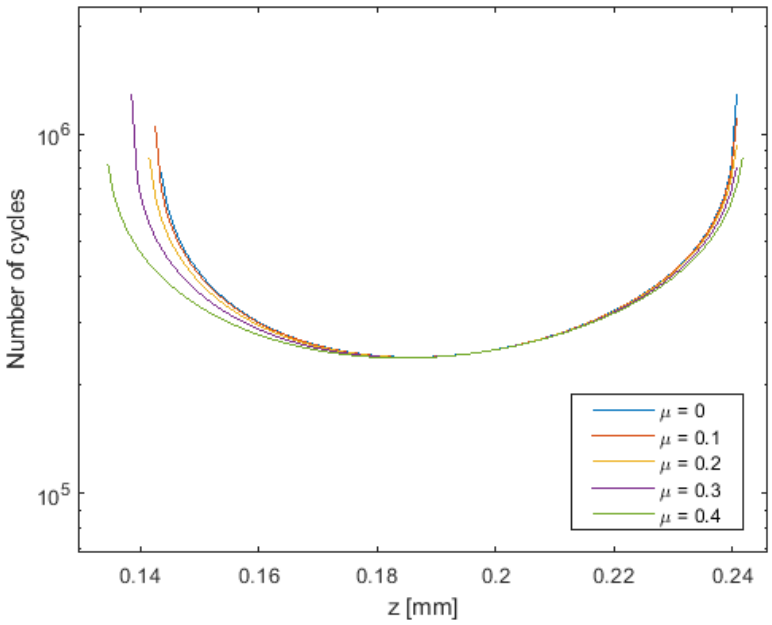


(b)

Figure 3.48: *Number of cycles to initiation at different depths, for* $P_max = 3480\ MPa$ *on AISI 52100. (a) general view, (b) detail on critical zone*



(a)



(b)

Figure 3.49: *Number of cycles to initiation at different depths, for* $P_max = 3580\ MPa$ *on AISI 52100. (a) general view, (b) detail on critical zone*

4 FEM Model

4.1 Introduction

A gearbox system is designed to transfer torque and rotational energy from the engines to rotating parts. In the case of helicopters' gearbox, the power produced by one or multiple engines is transmitted to the rotors through a shaft. The system is divided in multiple stages, in order to control the balance between rotational speed and torque, required from the rotors. The mechanism modeled in this work will be an epicyclic stage, which exploits multiple gears, called planets, which rotate around the central shaft. This shaft ends with a gear, called sun gear, that transmits the power coming from the engine to the planet gears, which are then kept in place by an outer fixed gear, called outer ring (Fig. 4.1).

Figure 4.1: *Schematic representation of an epicyclic gearbox stage [2]*

The plane gears are put into motion by the sun and are linked to the output shaft through a system of bearings, rolling on their inner race. The bearings are placed inside a cage, which rotates with them and transmits torque and rotation to a shaft, through a plate fixed to the cages of all the planet gears. This final shaft is the output shaft which is then linked to the main rotor (Fig. 4.2).

As seen in the introduction chapter, the fatigue problem investigated in this work, is referred to the contact section between the planet gear and the rolling bearings at his inner surface. The transmission of torque and rotation, coupled with the rolling contact condition, generates the stress state to be investigated.

The current model will be developed in Abaqus/CAE, starting from original 3D parts. From the whole model, only the sun gear, one planet gear with its bearings and the correspondent section of the the outer ring will be modeled. This reduction has been chosen for the sake of simplicity and to reduce the amount of calculation on the model, which will become impossible to analyze in its totality. Moreover, the symmetry of the epicyclic stage suggests this kind of simplification.

Both rotation and torque will be applied to the model and a preliminary analysis on the transmission feasibility will be shown. Initially the model will be a sample cylindrical shaft, to check the stress generation due to the application of torque and rotation. After this initial stage, the progressive assembly of the parts will be per-

formed. Starting from the two gears (sun and planet), the contact between the teeth will be implemented and checked. Then, the bearings and the fixed outer ring will be added to complete the section.

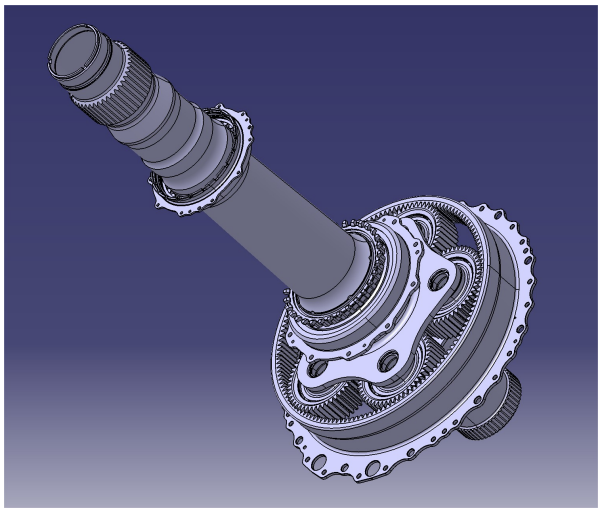


Figure 4.2: *Catia rendering of the epyciclic gearbox stage with input and output shafts*

Boundary conditions and contacts will be added coherently with the physical system functioning. After all the conditions will be checked and a satisfying mesh is obtained, the evaluation of the stresses in the critical regions, identified as the contact points between the bearings and the planet's inner race.

The torque and rotations applied will be derived from references on the engine power and rotational speed reduction by the gearbox. The computation of the reduction ratio will be performed following theoretical indications using the gears' teeth ration between the sun, planet and outer ring.

4.2 Model Parts

The gearbox model is realized starting from the drawings of the actual component of the AgustaWestalnd's helicopter AW169. This model is used mainly for civil applications and transportation of crews up to 13 elements (Fig. 4.3). The versatility of this helicopter makes it widely used and needing of monitoring.

From drawings given by the manufacturer, the single parts were extracted using a series of CAD software. Catia was the initial one, on which the drawings were produced, then Autodesk Inventor was used to separate the general assembly and obtain the single parts. In the end, these parts were imported into Abaqus for the analysis.

The whole gearbox stage is shown in Fig. 4.4, with and without the output shaft and relative plate linking to the planet gears. The structure is clear and the equal interaction of the planets with the sun and the output shaft can be deducted from the

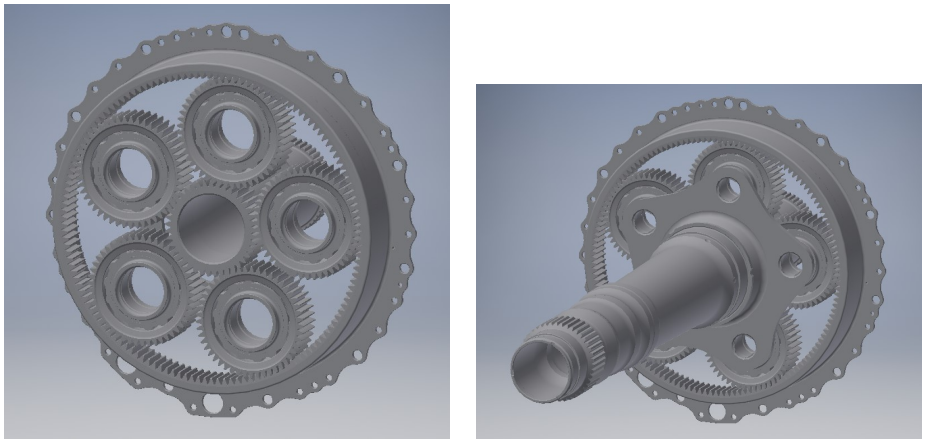
geometry of the configuration.

(a) (b)

Figure 4.3: *AgustaWestalnd AW169 during operations [167]*

The symmetry of the components and of the transmission system suggests that torque and rotation will be equally distributed among the planet planet gears. From this assumption, only one planet gear, with the relative bearings and outer ring section, will be analyzed in the FEM model.

Starting with the sun gear (Fig. 4.5), which is fixed on the terminal part of a shaft coming from earlier stages of transmission and from the engine. This is the component at the centre of the planetary system, causing its name, transmitting rotation to all the planet gears at the same time.



(a) (b)

Figure 4.4: *3D model of the epicyclic gearbox system showing: (a) the interaction between sun and planets, (b) the transmission system to the output shaft*

Because of the analysis on torque and rotation propagation, shown in the previous section, the input shaft in this element can be neglected. This operation has been proven to not alter the forces and motions in play. The upside of removing the shaft is to reduce the elements to be analyzed in the final stage and avoid possible meshing problems at the interface between the shaft and the gear.

Torque coming from the engine will be applied to the sun gear and will be trans-

mitted to the planet gears through teeth contact.

The planet gear (Fig. 4.6) will be added in contact with the sun and the outer ring. This component has a curvature on the inner race which permits the rolling contact of the bearings. The contact generated will have an elliptic section, with an elongated major axis, which allows the simplification to a line contact. From this consideration and utilizing Hertzian contact theory, the stress tensor evaluation for the numerical model will be derived.

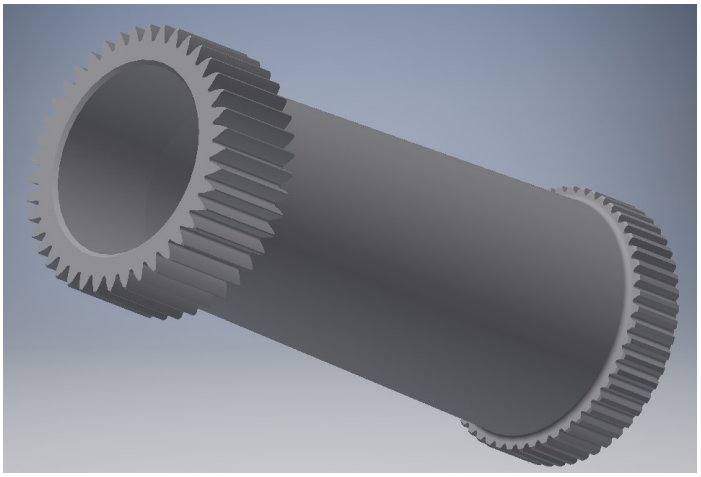


Figure 4.5: *Sun gear 3D representation*

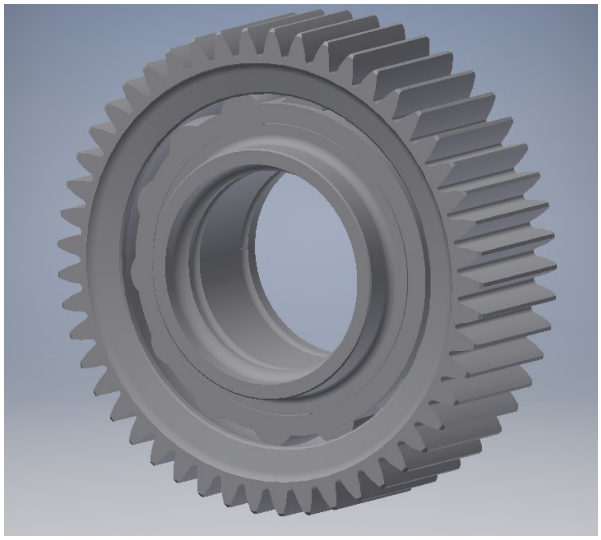


Figure 4.6: *Planet gear 3D representation*

The outer ring instead is the only fixed gear in this model. Acting as housing and track for the planet gears, the outer ring is fixed to the shell of the gearbox. Due to the approximation chosen for the current FEM analysis, only a section of this gear will be represented in the final model. Anyway in Fig. 4.7 it shown in its integrity.

The flange meant to be fixed to the outer shell will obviously be eliminated in the following model.

Inside the planet gear are contained the bearings, in their turn housed in a cage to guarantee the spacing between them and the transmission of rotation and loads. The cage-bearings system is represented in Fig. 4.8. This housing cage is disc in which the bearings are singularly situated and kept in place. Due to its functional, passive, role in the contact phenomenon and its relevant dimensions, the cage will be neglected in the FEM model. In place of the solid structure, a dimensionless beam structure will be modelled, guaranteeing the same functions.

Figure 4.7: *Ring gear 3D representation*

The roller bearings are instead reproduced in the final model, acting as major contact agents with the inner race of the planet gear. Kept in place by the cage-like structure, they act as intermediary for the load and rotation transmission from the planet gear to the output shaft. The structure will be copied from the original drawing in Fig. 4.9.

The components here represented will taken as parts to begin the FEM model. After modifying the geometries to satisfy the conditions for a smooth mesh in Abaqus, the assembly of the parts will give as a result the reproduction of a section of the epicyclic transmission. In the following sections the procedure to obtain the final model will be explained, showing the constraints and parameters used. A progressive construction of the model will allow to check step-by-step the validity of the system and of the conditions that will be assumed.

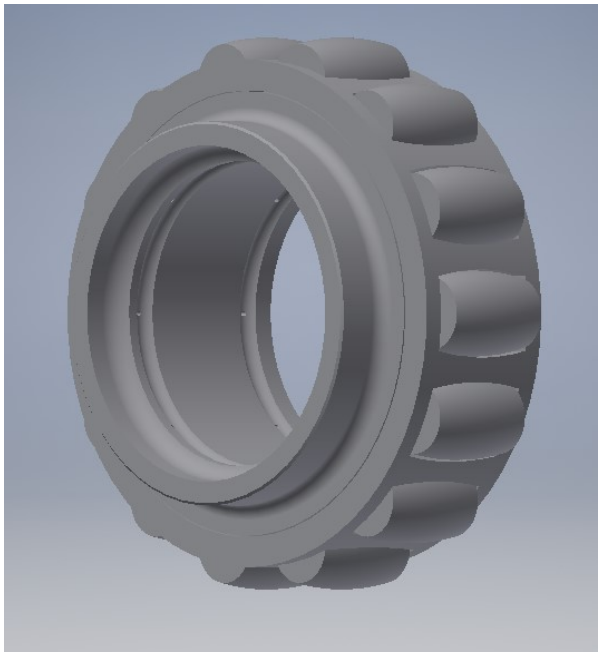


Figure 4.8: *Bearing cage 3D representation*

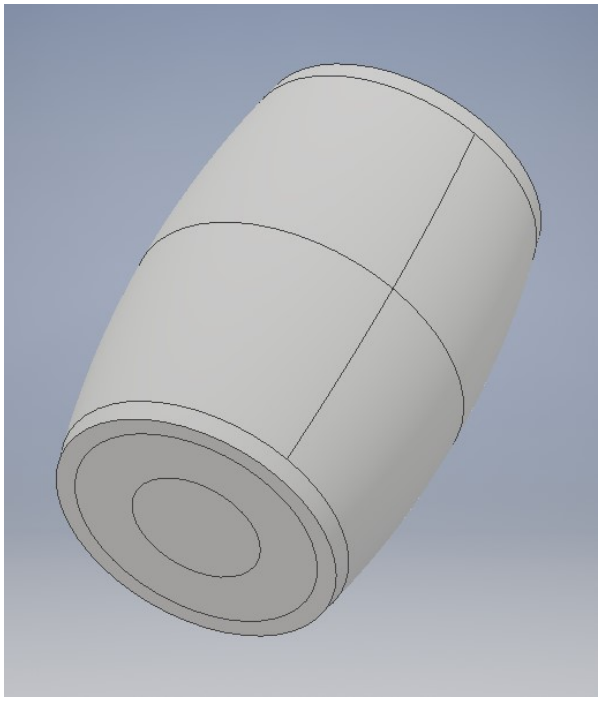


Figure 4.9: *Bearing 3D representation*

4.3 Torque and Rotation

In order to define the torque to be applied in the finite elements model, performance data of the real engines used on the Agusta Westland AW169 will be taken as reference. Before calculating the resulting torque, anyway, the transmission ratio of the planetary configuration should be defined.

The torque ratio from input to output, in an epicyclic gear system, can be evaluated with a straight forward calculation. The formulation takes into account the ratio between the rotational velocities of the different components (ω_i), relative to the number of teeth (N_i) of the gears.

The general formulation will be:

$$\tau = \frac{\omega_A - \omega_P}{\omega_B - \omega_P} \tag{4.1}$$

where A, B identify the sun and outer ring, P the planet carrier, linked to the output shaft, and S is referred to the planet gear, which doesn't take directly part in the calculation (Fig. 4.10).

The number of teeth and dimensions of the planet gears are determined with a direct relationship with the dimensions of the sun and outer ring gears. Indeed planet elements are not directly mentioned in calculations, preferring to refer to the carrier.

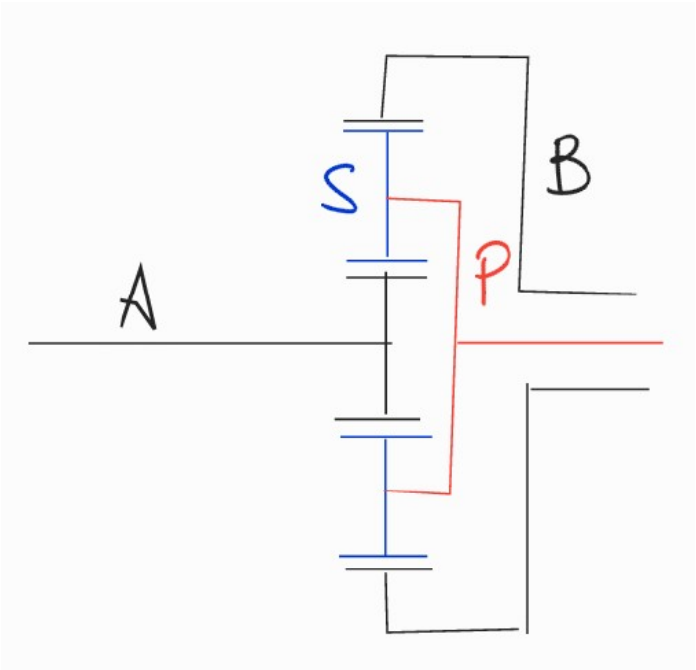


Figure 4.10: *Schematic representation of the section of the epicyclic gearbox.*

In the considered system, the supporting element is the external ring, which is fixed to the gearbox walls. The specific ratio τ is defined as:

$$\tau = -\frac{N_B}{N_A} = -R \tag{4.2}$$

The number of planets and teeth on the planet elements is a directly related to the number *tau* and though, this ratio becomes the defining parameter of the system.

In the case of a fixed outer ring:

$$\omega_B = 0 \tag{4.3}$$

Then, considering Eqs. 4.1 and 4.2 together, an equation can be derived, able to define the torque and rotational speed conversion, from sun to planet gears. The teeth ratio R will act as fundamental parameter.

$$\begin{aligned} \tau = \frac{\omega_A - \omega_P}{-\omega_P} = -\frac{\omega_A}{\omega_P} + 1 \\ \frac{\omega_A}{\omega_P} = 1 - \tau = 1 + R \end{aligned} \tag{4.4}$$

By knowing the geometric configuration of the system is then possible to know the reduction ratio to evaluate the rotational velocity transmission and reduction.

In the epicyclic gearbox under analysis, the number of teeth are:

$$\begin{aligned} N_A = 41 \\ N_S = 49 \\ N_B = 141 \end{aligned} \tag{4.5}$$

with the notation previously adopted.

The fundamental ratio τ and R can be then evaluated as:

$$-\tau = R = \frac{N_B}{N_A} = 3.439 \tag{4.6}$$

The engine power, for this specific model, will be applied at gear A and the output shaft rotates with the same speed of P.

Once characterized the mechanical system, the operational parameters of the engine should be defined. The AgustaWestland AW169 is equipped with two Pratt & Whitney Canada PW210A engines with turboshafts with FADEC system [168]. This engine model is attested in the 1,000 HP Thermo Power Class [169], approximate values at take-off, measured at sea level in standard day, static conditions and uninstalled.

The available maximum energy for the helicopter is then double by the presence of two twin engines. The rotational output speed made available from the engine ranges from 6,000 to 14,000 rpm.

It has been chosen arbitrarily to consider a low output rotational speed at the rotors, in order to simulate the maximum torque and stressing condition on the gearbox. In particular, the output speed ω_p has been imposed arbitrarily, following this principle. The selected data will then be:

$$\begin{aligned} \omega_p = 500[rpm] \\ \omega_s = 2'219.5[rpm] \\ P_{engine} = 1'000[HP] \end{aligned} \tag{4.7}$$

where ω_s has been obtained through Eq. 4.4.

The conversion between power and torque, for a rotating shaft is instead straight forward:

$$P[W] = \frac{T[Nm] \cdot \omega[rpm] \cdot 2\pi[rad]}{60[s/min]} \tag{4.8}$$

where the power to be considered is the sum of the two engines.

The conversion from horsepower $[HP]$ to Watt $[W]$ gives the following result:

$$P_{tot} = 2P = 2'000[HP] = 1'491.4 \cdot 10^3[W] \tag{4.9}$$

The output torque can now be obtained by solving Eq. 4.9, imposing the rotor's rotational speed as $\omega = \omega_P$ and assuming a unitary efficiency of the transmission.

From the output torque, by dividing in five equal contributions and applying Eq. 4.4, the torque applied by the sun at the single planet gear can be derived.

$$\begin{aligned} T_{p-total} &= 28'483.64[Nm] \\ \tau_p = \frac{1}{5} T_{p-total} &= 5'696.73[Nm] \\ \tau_s = \frac{\tau_p}{4.439} &= 1'283.34[Nm] \end{aligned} \tag{4.10}$$

Considering a complete conversion and an output rotational speed at the shaft of 500 rpm. It is obvious that halving the rotational output speed, the torque applied will be doubled.

The value τ_s obtained in Eq. 4.10 will be the torque applied at the sun in the final model, generating force transmission and stresses inside the system. This value can be reduced by 10-25% to account for the efficiency losses along the transmission and a usage lower than 100% of the power, which is not constantly yielded during the work life of the helicopter.

On the other end of the system a rotation will be applied. This rotation will be enough to generate the contact between a few teeth of the gears, in order to have a complete interaction between all the parts, in particular between the rollers and the planet's inner race.

4.4 Torque and Rotation Transmission

The initial step of the analysis will be to check the behavior of Abaqus/CAE when evaluating the transmission of torque and rotation. This control will be performed on a simple solid shaft modelled as a cylinder.

The dimensions of the object and of the applied loads are chosen arbitrary and non relevant in this stage. The aim of this analysis is to check that torque will be transmitted through the object without any loss or decay in the generated stress field and rotation. This result will allow to apply a torque transmission in the complex assembly in the following steps, relying on a constant propagation of forces and generated stresses.

The simple geometry described previously will be constrained in two circular hinge sections, on which the displacements will be fixed and the only remaining degree of

freedom will be the rotation along the central axis (Fig. 4.11). The loads will then be applied at reference points located on the lateral faces and coupled to the entire face. Both faces have an applied torque, although not clear in Fig. 4.11 because of Abaqus' colors and transparency choices.

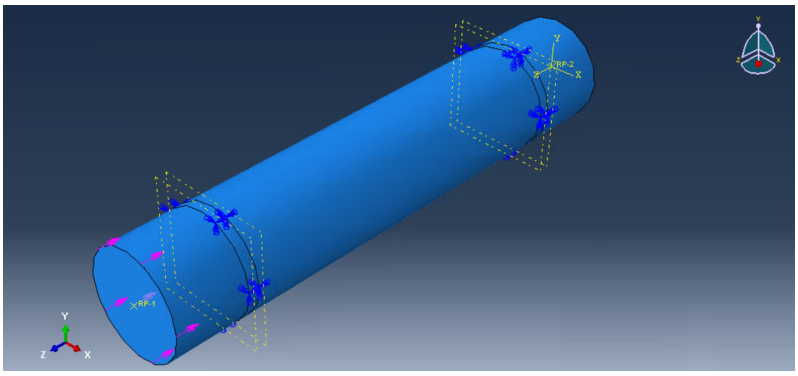


Figure 4.11: *Abaqus representation of the shaft with constrain sections, delimited by construction planes, and reference points on the lateral faces*

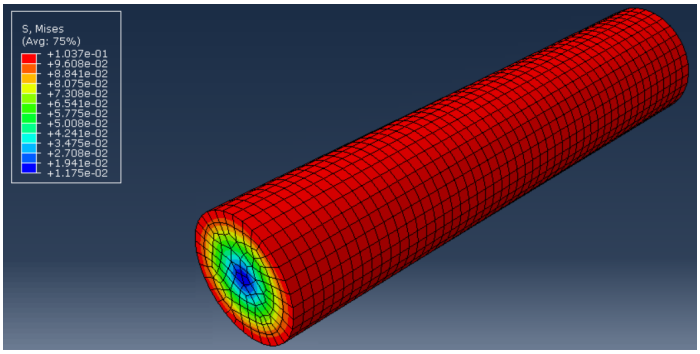
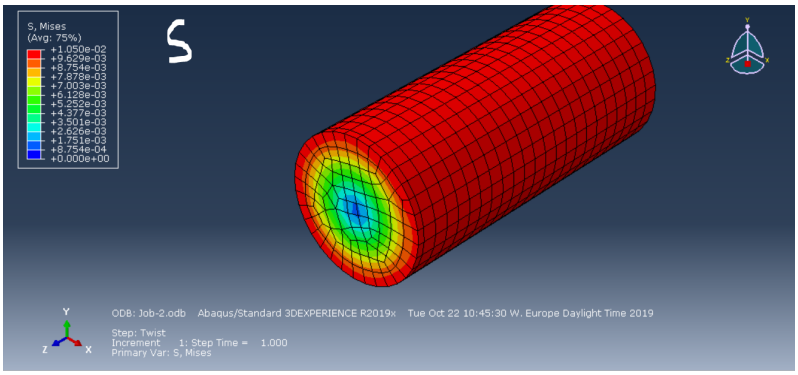


Figure 4.12: *Von Mises stress distribution resulting from a generic torque applied at the ends*

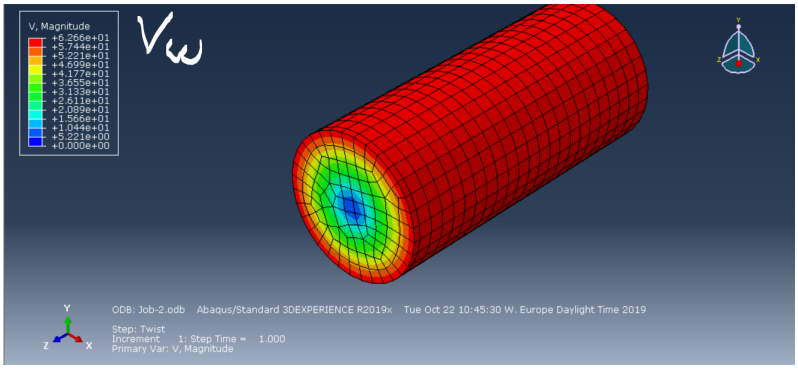
The torque and motion will then result to be acting on all the section. As it can be seen in Fig. 4.11, the rotation is applied to the whole face on one end of the shaft, whereas the load is applied to the reference point, coupled with the same region. The direction of the applied rotation and the one generated by torque are concordant, in order to give a coherent dynamics of the body.

The mesh used in this step is not particularly refined, with a generic Hex element type and a Sweep meshing technique. The resulting state of stress and rotational velocity profile are constant and radially distributed, as expected (Fig. 4.12). It can be seen that qualitatively the profile are similar, whereas the values are obviously different.

From this verification it can be stated that concordant torque and rotation generate radial a stress and velocity distribution that can be assumed as constant for parallel sections along the rotation axis. This assumption allows to neglect the input and output shafts in the complex final system, reducing the elements to be analyzed to the gears and bearings only.



(a)



(b)

Figure 4.13: *Section profile results for: (a) Von Mises stress, (b) rotational velocity*

4.5 FEM Model

4.5.1 Sun Gear and Input Shaft

The first step toward the construction of the complete model, was realized with the sun gear and the input shaft, coming from the engine. This assembly will be responsible of a simple torque and rotation transmission along a cylindrical section, terminating in the sun gear (Fig. 4.5).

The initial gear-like section of the shaft, present in Fig. 4.5 is eliminated in the FEM model because not relevant in the analysis. Moreover, it can be seen in Fig. 4.14 that to promote a better meshing and avoid geometrical problems, some sections with draft angles and complex geometries have been simplified. This process has been applied especially on the face of the gear.

Anyway, due to the urgency of reduction of mesh nodes in the final, complete, analysis, the input shaft will be removed from further models. This decision is justified from the assumptions of constant torque and rotation transmission along a cylinder, verified in section 4.4.

This operation also allows to avoid the presence of interfaces between different geometries, such as the one between the shaft and the gear.

The section will then be reduced to the single sun gear, which will carry the torque

from the previous stages of the transmission.

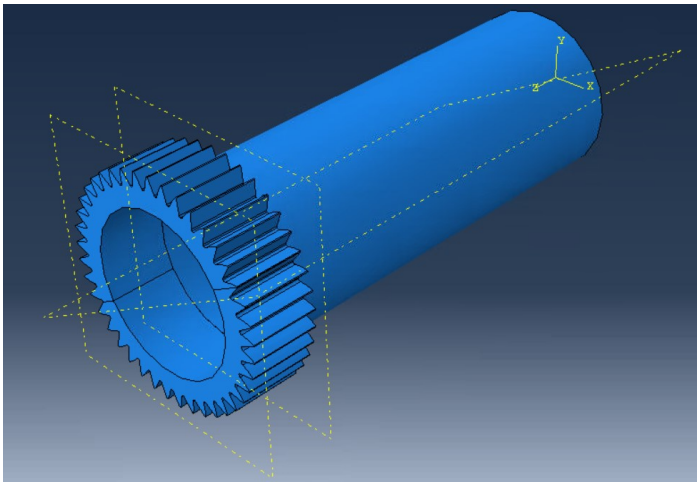


Figure 4.14: *Abaqus representation of the assembly between input shaft and the sun gear*

4.5.2 Sun and Planet Gears

After having removed the input shaft, the sun is added to the assembly, in contact with the planet gear (Fig. 4.15). This interaction, obviously, occurs at the teeth, where torque and rotation are transmitted from the sun to the planet gear.

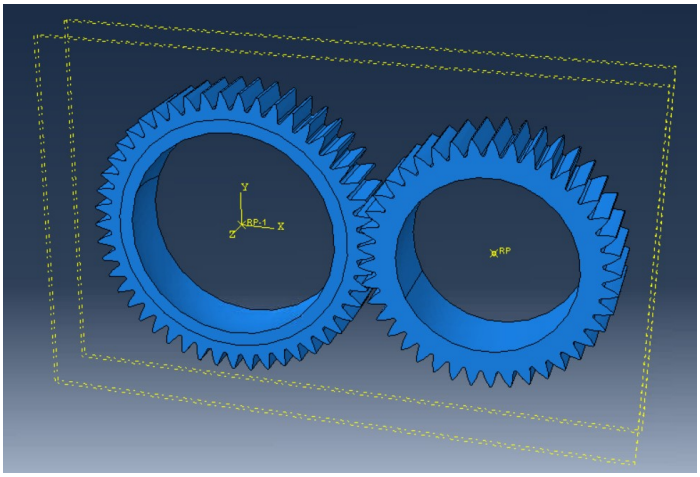
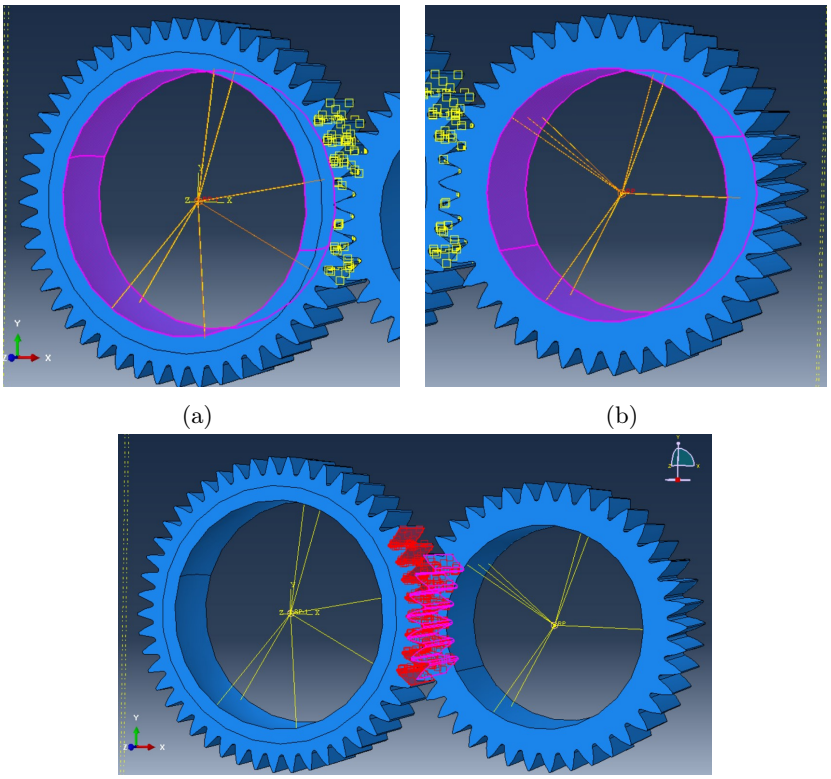


Figure 4.15: *Abaqus representation of the assembly between planet gear and the sun gear*

In this step of the construction of the model, only the contact effectiveness is verified. So, during the analysis of this specific system, the stresses will be purely arbitrary and the focus will be on the good functioning of contact instance.

In order to guarantee the alignment and correct placement of the components, the relative positioning of the gears has been extracted from the original scheme using Solid Works. Then a rotation is applied to one of the gears to remove any eventual overlap of the parts.

Both the gears are pinned to their central axis, where torque and rotation are applied through a coupling interaction to the inner races. This interaction is shown in (Fig. 4.16 (a,b)), where the involved inner surfaces, in purple, are coupled with the points at the center of the gears. The coupling is indicated by the yellow lines which link the involved parts and converge in the reference points.



(a) (b)

(c)

Figure 4.16: *Boundary conditions on gear components: internal coupling for (a) planet gear and (b) sun gear, (c) contact at teeth*

This setting will be carried onward for the whole analysis, going through more complex stages. The reference points at the centre of the gear will then assume a fundamental role in the whole model.

The surface-to-surface contact is applied on the outer surfaces of the teeth of both gears, limited to the region where contact is likely to occur, to reduce the calculation weight (Fig. 4.16). The extension of this contact zone has been increased in following steps, where the rotation angle became higher in order to allow a wider interaction, for stress evaluation. To guarantee the correct interaction between the two surfaces,

friction is introduced in the contact. The coefficient chosen for the contact between the gears' teeth is of $\mu = 0.3$, as indicated by most references [12][141][148]. The same value has been kept constant for the teeth contact of all following steps of the model.

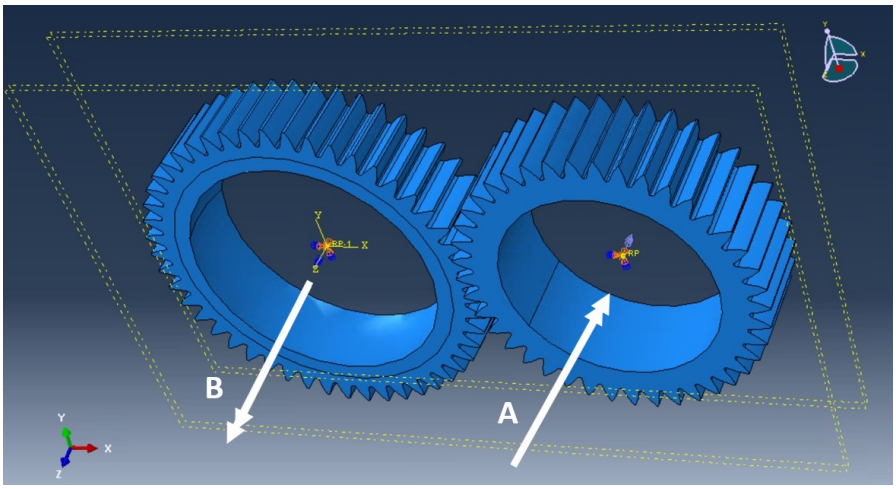


Figure 4.17: *Torque and rotation application on the sun and planet gears model*

The load is applied as a torque in the reference point coupled with the internal race of the sun gear, indicated with letter 'A' om Fig. 4.17. In the first step instead, the planet gear is kept fixed to allow the transmission of the load through the whole system.

The rotation is then applied with the same mechanism to the planet gear, transforming its boundary condition from an encastre to a pin with active rotation on the free direction (Fig. 4.17(B)). Also the sun has been pinned in this step and will rotate, dragged by the planet and its torque.

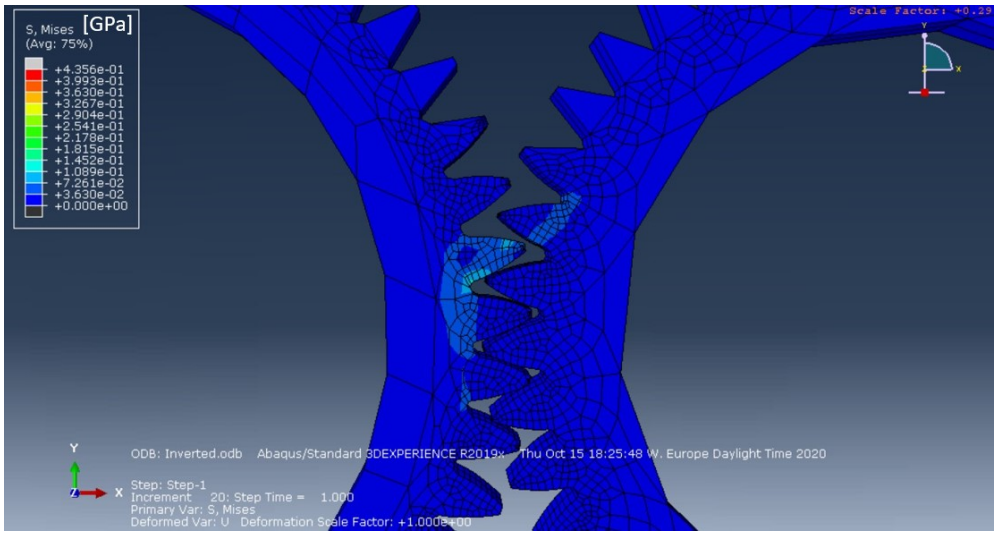
The resulting stresses are consequence of the contact interaction at the teeth (Fig. 4.18), whereas the ovalization deformation produces a relatively small contribution in a distant section of the gear (Fig. 4.19(b)). At this stage the values of the resulting stresses are purely arbitrary and not relevant, because of the lack of part of the complete system, which will take an important contribution in load sharing.

The material properties assigned for the analysis, from this point on, will be only the Young's modulus ($E = 2.06\ GPa$) and the density ($\rho = 9.8\ kg/dm^3$), corresponding to an $AISI$ 52100 alloy [163]. The stresses have then been weighted on this specific material, which corresponds with the most resistant alloy analyzed in the analytical-numerical model (Chap. 3).

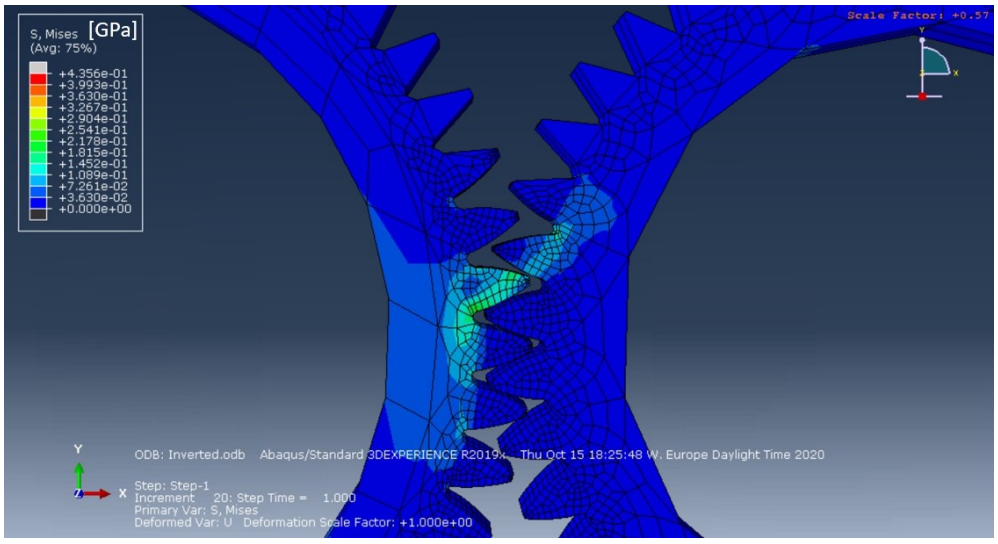
It can be noticed in Fig. 4.18 and 4.19 that the mesh is finer in the region of teeth contact, to allow a good analysis in the critical zone for the model and reduce the overall time of analysis with a coarser mesh in the peripheral regions.

Peaks reached at the socket angle between the loaded teeth, expressed in GPa, will exceed the material's elastic limit in this step. This should be due to the lack of components (as the outer ring and the bearings) from the model, which causes the load to be discharged entirely onto the teeth where contact occurs.

Anyway, rotation and load transmission are verified and the contact configuration results to be acceptable, which was the aim of the current analysis.

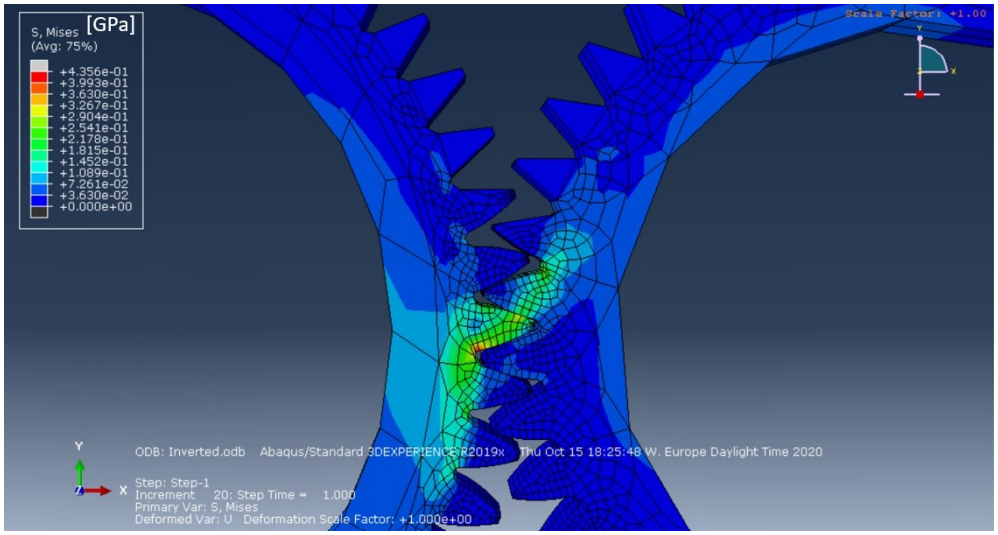


(a)

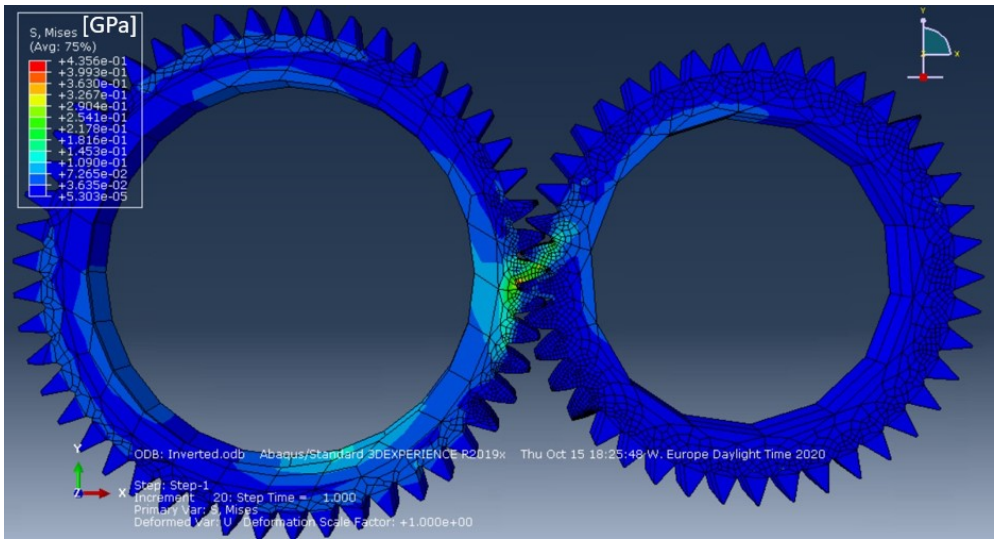


(b)

Figure 4.18: *Resulting stresses for sun and planet gears model at different steps of the rotation*



(a)



(b)

Figure 4.19: *Resulting stresses for sun and planet gears model at the final step: (a) detail and (b) general visual*

4.5.3 Rollers and Rolling Contact

The next component to be added at the model has been the rollers, inside the planet gear and in contact with its internal race. Through this rolling contact, torque and rotation will be transmitted, defining the state of stress on the planet gear, which is object of the analysis.

Before introducing the components in the general model, it has been realized a simulation model to verify the rolling contact feasibility and the interactions needed to realize it.

This simple model has been assembled with the original roller's geometry shown in Fig. 4.9, put into contact with a flat surface. Whereas here the component is taken singularly, in the complete system the rollers are multiple and inserted in a cage that holds them in position (Fig. 4.8). The beam has been modeled as the section of the bearing cage holding the specific roller in the complete model. The average section of the cage has been evaluated from the original model (Fig. 4.8) and a square section, corresponding to the average of the cage, has been assigned to the beam. The dimension of this section is of 27.5 mm. This approximation is assumed to guarantee a stiffness similar to the one of the cage.

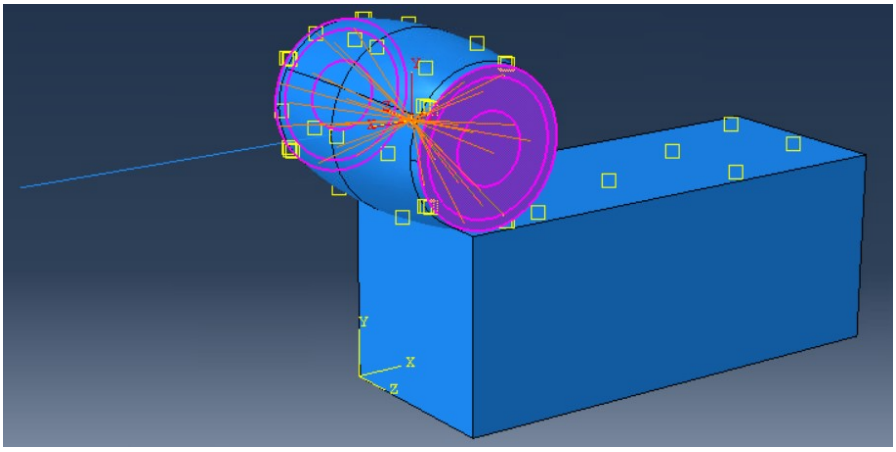


Figure 4.20: *Verification model with highlight on the interactions between beam and roller components*

For the sake of simplicity in the FEM analysis, due to the minor relevance of the cage component, this setup has been represented by a system of beam elements. These beams deputise for the more complex cage, keeping the rollers in position and allowing only the rotation on their central axis.

This setup will allow to verify the proper functioning of the constrains between the beam and roller elements, allowing a correct rolling contact.

In the verification model, a single beam will be introduced, pushing the element in its rolling motion over the race surface. The end of the beam has been coupled to the roller through a reference point, which is itself coupled to the faces of the roller, leaving free only the rotation, as shown in Fig. 4.20 and left visible in Fig. 4.23. This setting of fixed coupling between beam's end and reference point and instead free motion left on the coupling between point and roller's faces is the only one allowing the rolling

movement. If the coupling interactions are inverted, only sliding will occur.

In order to generate a rolling contact, a displacement has been applied at the free end of the beam, parallel to the race's surface. No vertical loads were added at the roller, leaving weight and friction as the only forces at the contact.

Different friction coefficients were analyzed to check the sensibility of the system. The friction regime chosen in this specific analysis was the low friction one. In fact, the rolling contact between rollers and gear's race in bearing gears occurs in lubricated regime, reducing the friction to very low values. Either in initial considerations on the complete model no pre-load has been introduced on rollers and beams, so the simple rolling assumption is legitimized.

It has been found that the Abaqus analysis for this setup cannot be completed for a friction coefficient lower than $\mu = 0.05$. Different analysis at lower values ($\mu = 0.01, 0.02, 0.03$) were not converging and automatically aborted by the system after several attempts. The results shown in Fig. 4.22 and 4.21 has been obtained with the minimum possible friction coefficient, of 0.05.

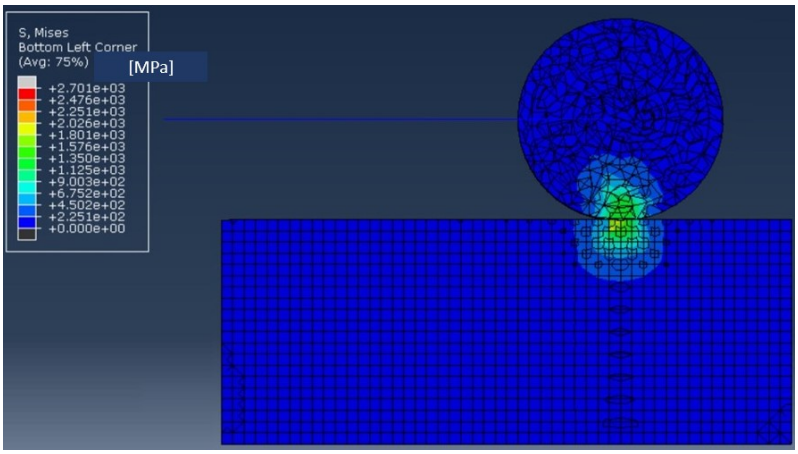


Figure 4.21: *Resulting stresses for rolling contact of a roller onto a surface: final step, increment* $t = 1$

It is evident how the stress state is formed gradually, when the roller is beginning to roll over the surface, due to the weight as unique source of load present (Fig. 4.22(a)). When the roller is instead completely over the races section the stress state generated is constant and moves with the roller (Fig. 4.22(b) and 4.21). In this way the contact and the constrains necessary between the parts are considered verified and ready to be implemented in the general model.

The absolute values of this results are purely demonstrative and exploited only to demonstrate the validity of the contact and its effects. The presence of an interference between the roller and the surface has been introduced to guarantee the rolling contact, but it affects the correct evaluation of the stresses. Moreover, the non conformity of the two parts acts as stress raising factor.

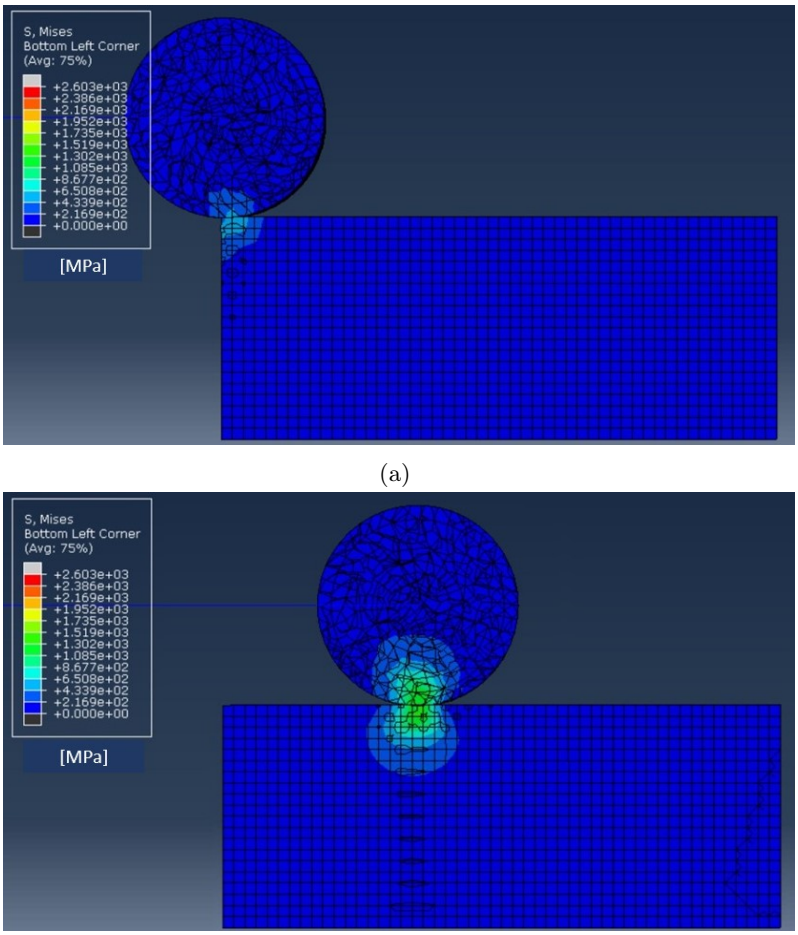


Figure 4.22: *Resulting stresses for rolling contact of a roller onto a surface: (a) increment* $t = 0.1$ *(b) increment* $t = 0.5$

4.5.4 Complete Model

To complete the model introduce in section 4.5.2, the internal system of the bearings and a section of the outer ring, should be added.

The bearings section had been modeled in the beam-roller system introduced in section 4.5.3. These parts will be replied 14 times in a circular, evenly spaced, distribution replying the original bearings configuration (Fig. 4.8). The resulting components to be added in the cavity of the planer gear are shown in Fig. 4.23, where the reference points, where couplings are applied, are left visible on one roller as demonstration.

The reference point at the center of the beams is assumed to be the center of the bearing cage, connecting all the beams as a unique structure. This central point is the coupled with another reference point, placed on the rotation axis of the sun gear, slightly offset from the sun's center. This second point represents the output shaft, from which rotation is given, transmitting it to the bearing cage with the setup shown

in Fig. 4.4(b).

Hence, the rotation will no more be directly applied to the planet gear, but at the output shaft. Rotation is then transmitted to the gears through the bearing cage and the contact between rollers and planet gear.

The final component added to the assembly is a section of the outer ring (Fig. 4.7). This gear acts as case and guide for planet gears' rotation and is assumed to be fixed in this model. Due to the symmetry considered since the beginning of the model, only a small sector of the ring is modeled in the assembly. This allowed to show a rotation which is still meaningful for the stress analysis, without modeling all the structure.

The same kind of surface-to-surface contact at teeth used for the interaction between sun and planet is used for the contact between planet and ring gears. This last component is then fixed at its external surface with an encastre boundary condition. Clearly on forces are applied to this component, which acts as bounding element to stabilize the system and allow a realistic share of the loads.

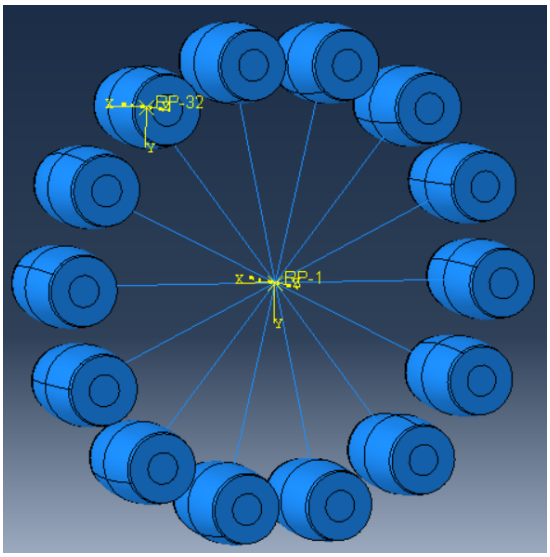


Figure 4.23: *Beams and rollers representing bearings and cage from the original model*

The complete structure assembly is shown in Fig. 4.24. Once the model has been completed in all its components and the interactions has been defined as long as the loads and boundary conditions, the analysis can be performed, after having meshed all the parts properly.

4.6 Mesh

The assembly has been meshed using standard elements, to avoid an excessive weight in the analysis, due to the already complex geometries and interactions (Fig. 4.25).

The beams has been meshed with B31 seeds, corresponding to 2-node linear beam elements with standard settings.

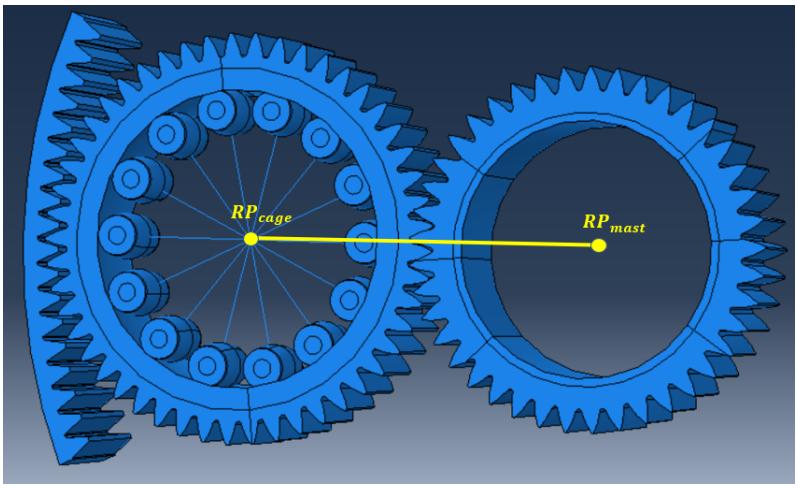


Figure 4.24: *Assembly of the complete model comprehending beams and rollers and ring, planet and sun gears*

The gears and the rollers has instead been meshed with C3D8R elements, generating 8-node linear bricks with enhanced hourglass control. The reduced integration setting has been chosen in the cells where the analysis was not required to be extremely refined. So, the planet gear's internal section on the half in contact with the sun gear has been chosen as the defined analysis region. Here, the reduced integration is removed to allow a more precise evaluation of the stresses. A further refinement of the mesh would be recommended to obtain even more precise results, but due to the complexity of the model and of the geometries involved, this approach is initially discarded.

The dimension of the elements has been tuned to achieve a refined analysis, avoiding the formation of distorted elements in regions with complex geometry.

Choosing a seed dimension of 2 mm in the planet gear's mesh, the formation of distorted elements is avoided, which instead arises with both coarser and finer seeding, due to geometries and compatibility issues with neighbouring regions.

4.7 Complete model results

The most critical values in the results of the analysis are shown at the interface between the teeth between of the gears, as correctly expected being this the region most subject to wear and fatigue phenomena. These values oscillates during the rotation due to the different angles and regions of contact. Anyway is shown (Fig. 4.26) that the stresses generated never exceed the elastic limit of the material, of around 2 GPa, as mentioned in Tab. 3.1. It has been assumed then, that this region will not fail due to plastic deformation and considered as not relevant. Fig. 4.26 is indicated as reference for the scale of the stresses, more detailed results in critical regions are reported in the next paragraphs.

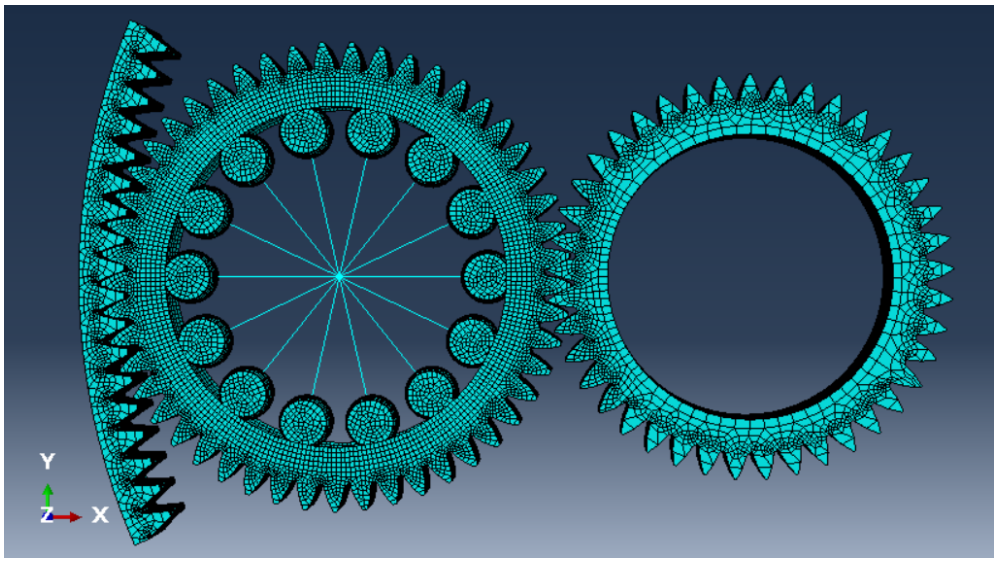


Figure 4.25: *Meshed assembly of the complete model.*

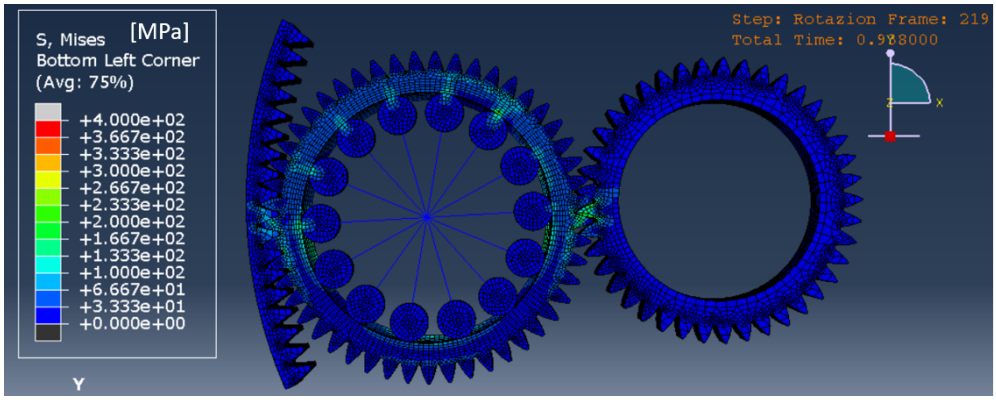


Figure 4.26: *General view of the resulting stresses*

The region under analysis in this model is instead the inner race of the planet gear, where rolling contact between the rollers and the race is the main source of stress. Here, contact stresses are correctly generated where rolling contact is verified (Fig.4.27).

As introduced by Depouhon et al. [7], ovalization of the planet gear is caused by the contact between the gears. This slight deformation on the sides of the component defines an upper zone, where contact stresses are generated in a more marked way, and a lower zone where the deformation of the gear reduces the contact interaction between the rollers and the gear's race. The stresses generated from this phenomenon are shown in Fig. and compared with reference predictions from Depohuon [7]. The absolute value of the stresses caused by ovalization of the gear is around 120-170 MPa as shown in Fig. 4.30, three times lower than the contact stresses generated between rollers and race.

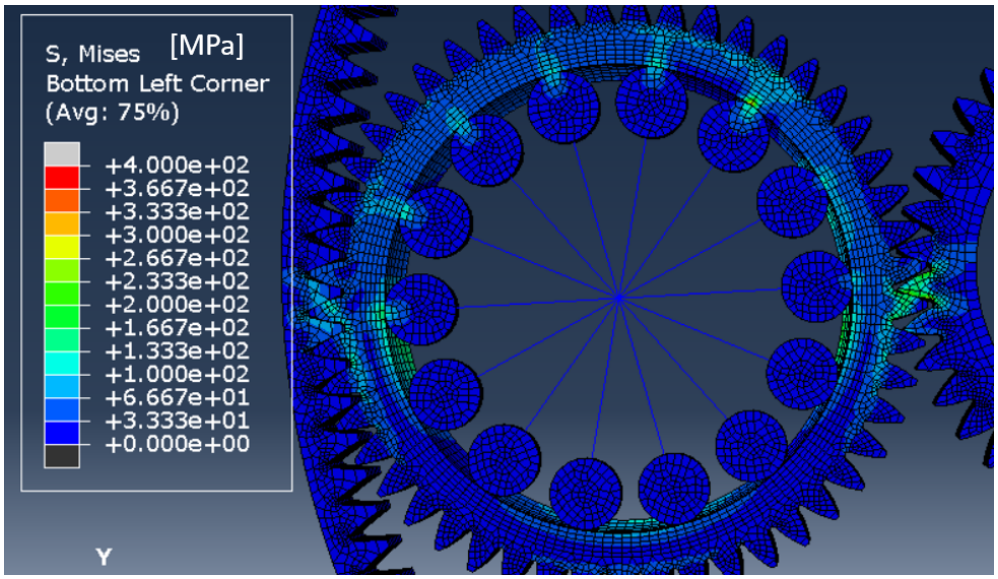


Figure 4.27: *Resulting stresses on the planet gear*

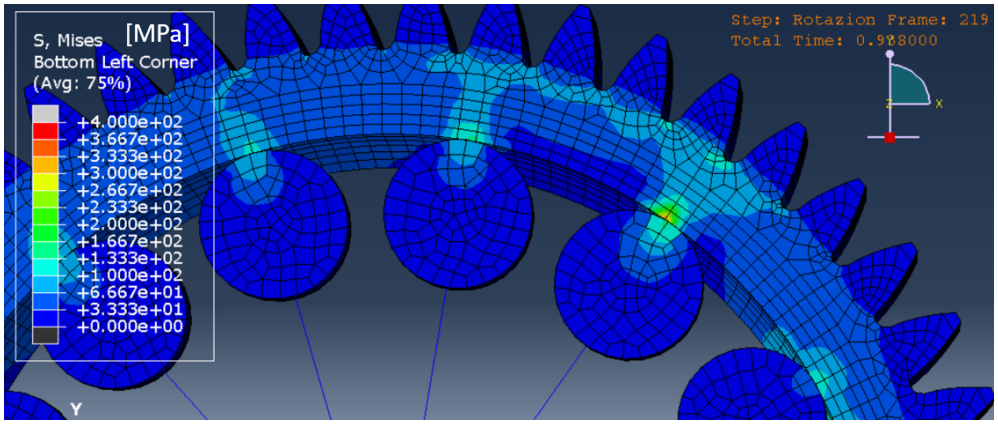


Figure 4.28: *Resulting contact stresses on the upper part of the planet gear*

In the first step of the analysis, torque is applied at the sun, introducing load through the structure. During this step all the components are fixed through their reference points or surfaces and only the torque is introduced into the system (Sec.4.5.2).

In the second step instead, the planet and sun gears are free to rotate along their central axis, in the relative reference systems, and rotation is applied at the opposite end of the system. The reference point identifying the output shaft, which is coupled to the beams' center (Fig. 4.24), then simulates the transmission of the rotation through the bearings cage and the rollers themselves. The rotation applied is of 0.18 *rad*.

In this way, recalling the tryouts performed on the cylinders shown in Sec. 4.4, torque and rotation are applied at the opposite ends of the system, allowing the analysis of the stresses inside the model.

The rotation of all the components is correctly verified, from the rotation of the sun and planet gears to the rolling motion of the rollers over the planet gear's race.

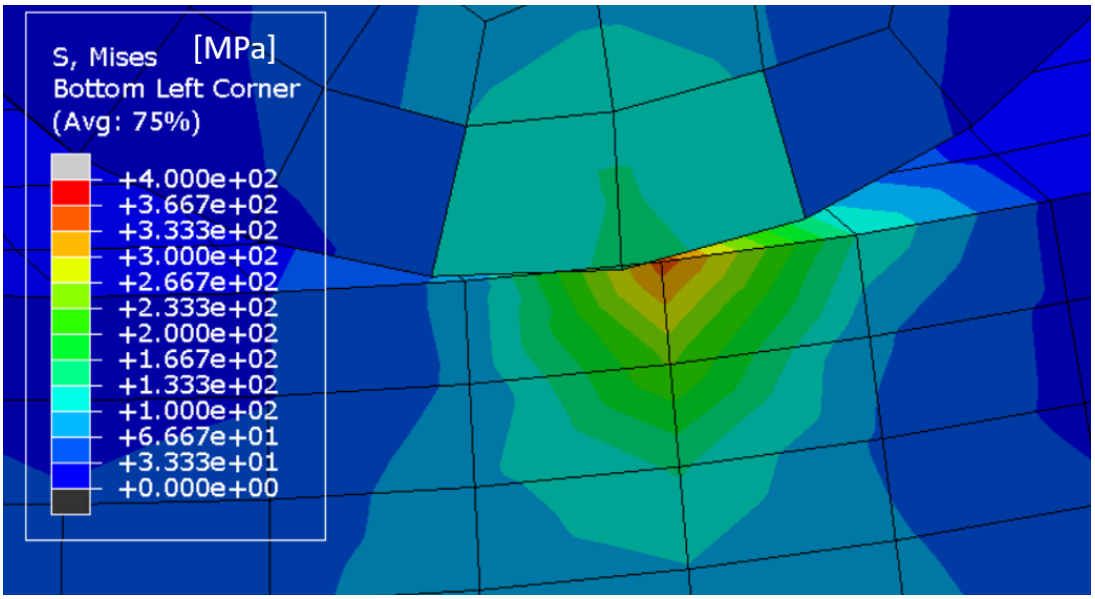


Figure 4.29: *Detail of the resulting contact stresses at the contact between a roller and the upper part of the planet gear*

In the second step, rolling contact stresses are correctly generated on the inner race of the planet gear, moving along this surface according to the movement of the rollers over it, in particular in the upper part of the planet gear (Fig. 4.27).

The scale of the stresses is correctly around values in the hundreds of MPa, where are predicted to be in literature and also in the analytical-numerical model presented in this thesis work (Sec. 2 and 3).

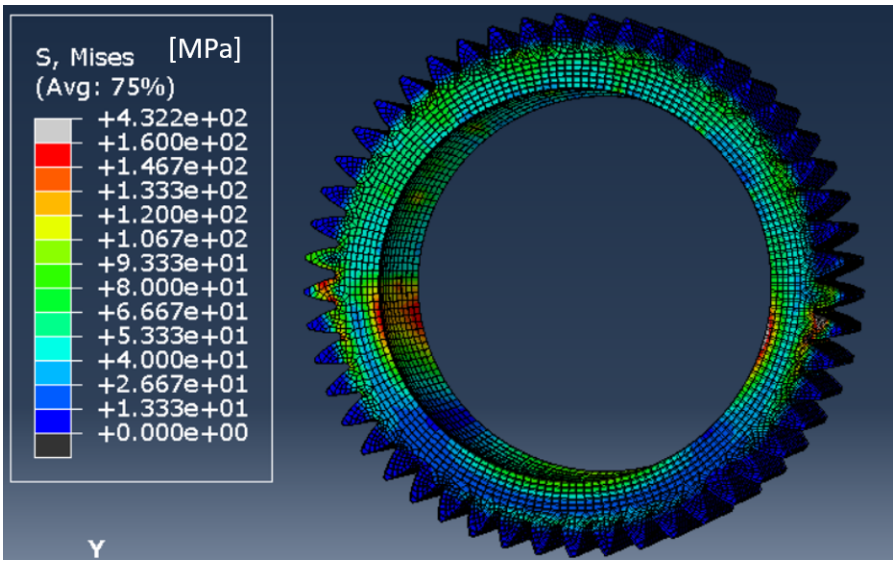
In detail, in the region considered as object of the current analysis, the maximum stresses reached during the rotation process are in the range of 300-400 MPa, as shown in Fig. 4.29. The region highlighted in res, indicating higher stresses, is located at a depth around 1/10 of the element. The dimension of these elements is fixed to be of 2 mm, so the most stresses zone is correctly in the range of around 200 μm, where rolling contact fatigue is predicted also from theoretical models (Sec. 2 and 3).

As already introduced in the previous section (Sec. 4.5.3), the results shown in standard mode by Abaqus represent the Von Mises equivalent stress. This gives an approximation to the order of magnitude of the equivalent maximum shear stress, chosen as critical in the fatigue analysis (Chap. 3) [16].

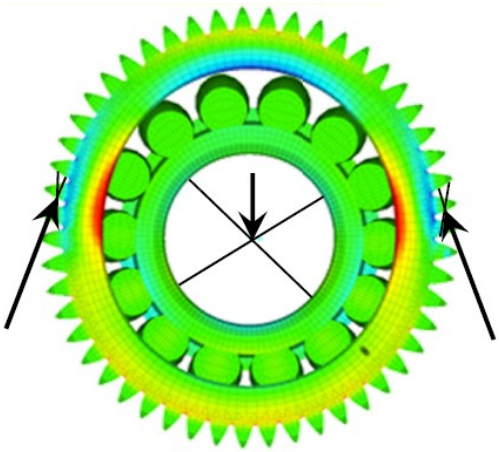
It can be seen in Fig. 4.29 that the maximum value of the stress generated in the gear's race is around 300-400 MPa, comparable with the stresses generated in the analytical-numerical model (Chap. 3) and able to generate damage in the material.

Anyways, if investigating the stress tensor generated in the complete model analysis the resulting principal stresses (Fig. 4.31) lack the precision required to have a validation with literature [5].

This lack of correspondence can be attributed to the complexity of the model and the consequent difficulty in refining extremely the mesh in the rolling contact region. From this considerations, the idea of exploiting the single roller model in conjunction with the complete model has raised, to evaluate more precisely the contact stress tensor.



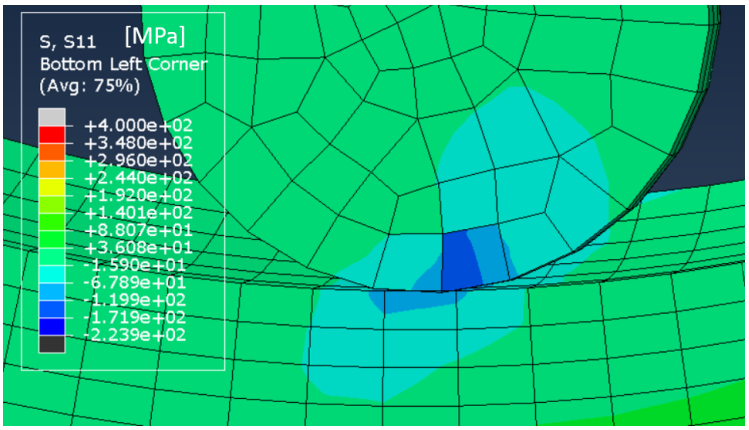
(a)



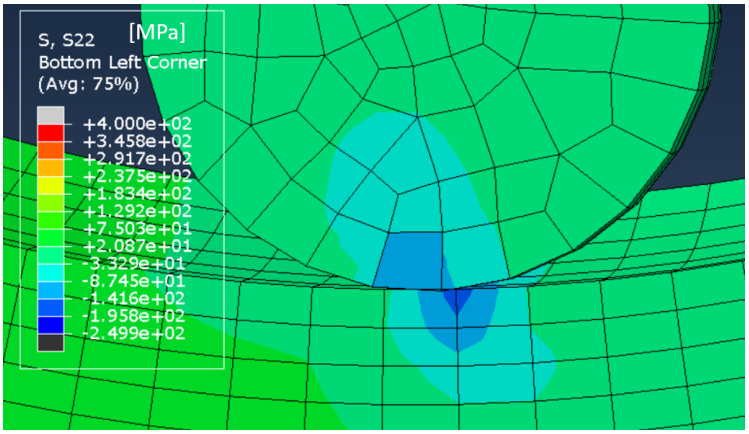
(b)

Figure 4.30: *Rim ovalization under global loads from gears' contact. (a) FEM results and (b) reference image from Depohuon [7], with equivalent hoop stresses indicated by the arrows*

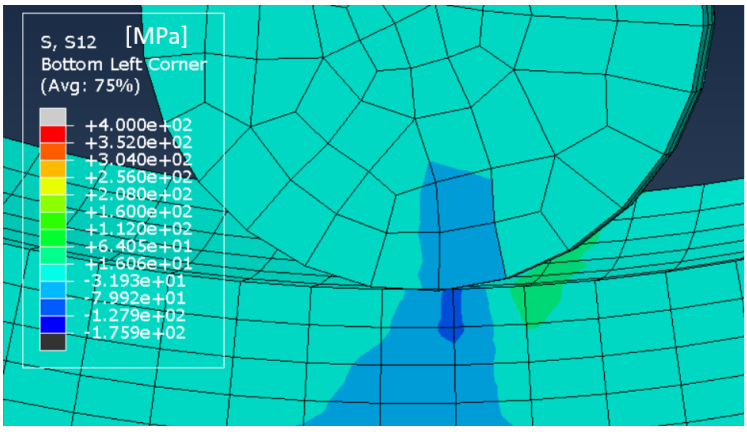
The stresses along the beams, simulating the bearing cage, has been evaluated and used to define the force acting on the rollers, pushing them onto the gear's race. The idea is to multiply the stress, felt by the beam, by its section to obtain the pushing force. Once obtained this normal force, it has been introduced in the single contact model to push the roller onto the surface with the value given from the transmission of loads in the complete gears model.



(a)



(b)



(c)

Figure 4.31: *Principal stresses of the contact stress tensor generated in the complete model FEM analysis. (a)* σ_x*, (b)* σ_z*, (c)* τ_{xz}

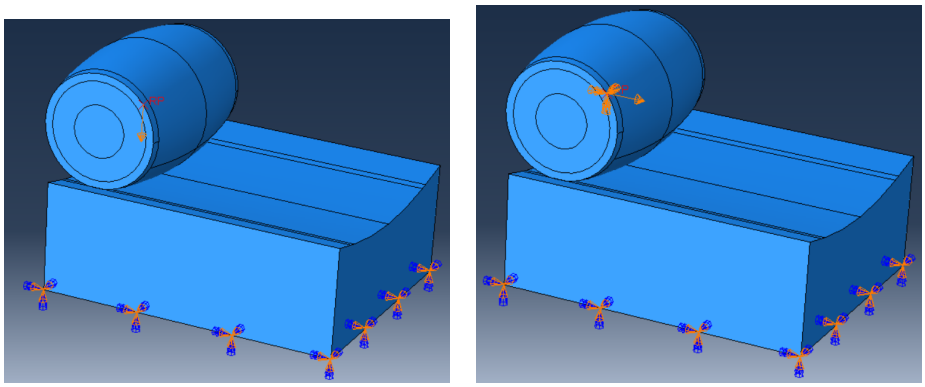
4.8 Further development: single roller model

Recalling the model used to validate the contact interaction, the only parts present in the analysis to simulate the contact interaction were a roller and a flat surface.

The beam element present in the previous single roller model has been eliminated in this last analysis, because the force acting normal to the surface has been evaluated in the previous section (Sec. 4.7), already considering the section of the beam.

$$\begin{aligned} \sigma_{beam} &= 18.014 \quad [MPa] \\ S_{beam} &= 756.25 \quad [mm^2] \\ \vec{F_n} = \sigma_{beam} * S_{beam} &= 13.62 * 10^3 \quad [N] \end{aligned} \tag{4.11}$$

In Eq. 4.11 are represented the stress felt from the beam (σ_{beam}), its section (S_{beam}) and the resulting normal force which is applied to the roller through a reference point placed at its center (Fig. 4.32(a)).



(a) (b)

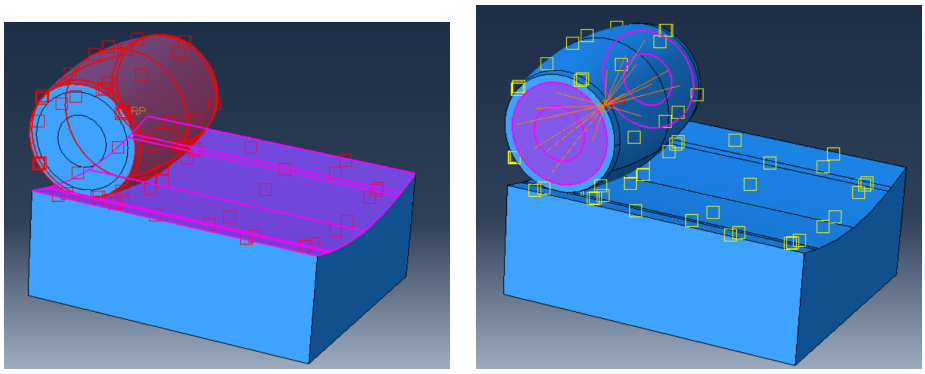
Figure 4.32: *Application of the load (a) and the translation (b) to the roller in the model*

The force has been applied vertically to the roller, transmitting then the pushing action to the whole component onto the race's surface. Once the load has been introduced and the contact established, a translation of the bearing over the race is applied to generate the rolling contact, with the same working procedure explained for the application of the force (Fig. 4.32(b)).

The one-dimensional conformity of the roller and the gears' race has been achieved to replicate the geometry in the complete model. In this way, the contact is distributed over a line, as assumed also in the previous sections (Sec. 3.2) for the analytical model and satisfies the line contact approximations assumed through the analytical model.

The boundary condition for contact is the surface-to-surface contact assumed also in the complete model (Fig. 4.33(a)), and the reference point is coupled to the roller's side surfaces, also like in the previous model (Fig. 4.33(b)).

The mesh of the race is refined at a dimension of 0.5 mm per cell, with complete integration, to allow more reliable analysis of the stresses generated from the rolling contact. This refinement increases the time needed for a complete analysis.



(a) (b)

Figure 4.33: *Interaction conditions of (a) contact and (b) coupling, in the single roller model*

The final result obtained from this model is going to be a tensor of stresses, showing the principal normal stresses and the shear generated inside the material. This kind of output is similar with the one obtained from the first part of the analytical-numerical model, for stress evaluation.

This operation has not been completed due to time issues, but is under study and is due to be ready by the day of the presentation.

5 Sinergy between analytical and FEM models

In Fig. 4.34 is shown the combined work scheme of the complementary functioning of the two models.

The FEM model starts from the CAD drawings of real components from a complex gearbox, analyzing the transmission of torque and rotation through a planetary stage.

Generated from the application of torque and rotation on the complete system, the transmission of loads through the contacts between the gears and the bearings results in a stress value in the beams, which simulates the bearing cage and its stiffness. This value is then transformed in a normal load by integration over the section of the beams, obtained as the average section of the bearing cage corresponding to a single roller.

The force is then going to be applied in the single roller model, which with its reduced complexity is able to simulate the rolling contact condition and is predicted to give a satisfying output. The stress tensor generated inside the gear race is going to be evaluated, which is the same kind of input used to perform the damage evolution analysis with the analytical-numerical model.

A fatigue initiation analysis has previously been performed with the analytical-numerical method presented in Chapter 3, able to evaluate the damage damage evolution inside a material subject to rolling contact exploiting continuum damage mechanics approach. The model has been validated with reference to literature [16] and has demonstrated the ability to correctly give a first approximation prediction of the fatigue initiation life as a function of the load applied and of the material properties. This model can be used as a tool to evaluate the rolling fatigue initiation behavior.

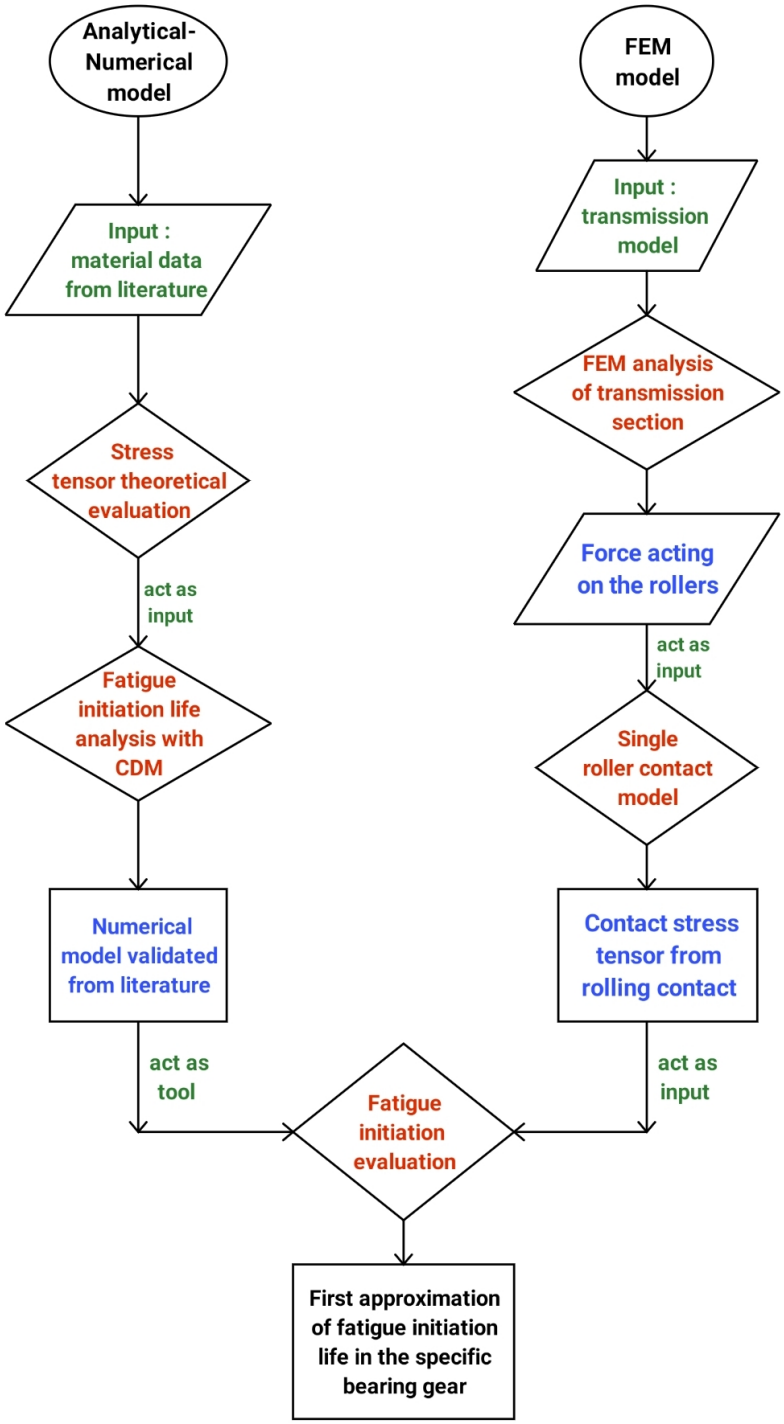


Figure 4.34: *Flowchart representation of the models interaction and working scheme*

The output of the FEM model, being a stress tensor, resembles the one obtained from the first part of the numerical model, where stresses were calculated following Hertzian theory. Here, the stress tensor from the single roller model is taken as input for the damage evolution analysis, and the fatigue initiation life is evaluated.

An analysis of the tensor generated in the single roller FEM model can be performed to obtain the values of the principal stresses necessary to evaluate the maximum shear stress. Once this passage is completed, the fatigue initiation prediction, based on continuum damage mechanics is going to be performed.

The final result is going to be the number of cycles at fatigue initiation, related to the specific planet gear considered in the FEM model. This first approximation is considered in the regime of near maximum torque described in the previous section (Sec. 4.3).

6 Conclusions

Starting from a real event where rolling contact fatigue led to catastrophic consequences, this thesis work took its cue toward the purpose of studying the fatigue behavior of bearing gears alloys and components. The final aim has been to develop support for the design of long-lasting components, predicting fatigue behavior.

Bearing gears are widely used in mechanical applications and, despite the more and more strict requirements and high-quality production standards, are subject to rolling contact fatigue. This mechanism results to be the main failure cause for bearings. The repeated interaction between a rolling element and a surface leads to the formation of micro-cracks, which can propagate to form spalling or even penetrate toward the bulk in critical cases.

Even if these components are designed to withstand the stresses to which they are subjected, and work in a safe range, the presence of inner defects or peaks in the stress history can initiate the formation of damage inside the material. Many resulting mechanisms such as pitting and spalling or case crushing are possible, even the total failure of the component if the cracks can penetrate the bulk.

The complex nature of rolling contact, involving pressure and traction combined in a moving load give rise to a multi-axial state of stress inside the material [5][6][7]. This behavior is described by the Hertzian theory of contact [5], allowing a theoretical evaluation of the principal stresses generated, in order to have a general overview of all the factors contributing to the tensor.

This is the basis for the formulation of analytical models, which can simulate some specific physical phenomenon based on theoretical or experimental assumptions. In this specific thesis work, Matlab has been chosen as software to develop an analytical model starting from the Hertzian theory of rolling contact.

After the evaluation of the multi-axial stress tensor, the approach proposed by Beheshti et al. [16] has been followed, in order to develop a model able to predict the fatigue initiation life of rolling elements. The definition of a critical stress factor has been assessed, designating the maximum shear as the main driving factor for rolling contact fatigue crack formation [7]. Shear stress has been found to give the main contribution to Mode II crack formation and propagation in rolling contact fatigue and then chosen as the critical factor for numerical analyses.

To describe the evolution of damage inside the material, fundamental for fatigue initiation, the continuum damage mechanics theory has been proposed [147][14][140]. This method allows to describe the effect of stress and strain on the component's integrity, evaluating the damage formation in the structure as a function of the load applied and of the properties of the material.

The material parameters needed for this kind of analysis are easily obtainable with simple mechanical tests such as tension tests [141][158]. Once the data are collected, the analytical model can be developed and used to predict the fatigue initiation life and the critical region at which damage occurs.

This kind of model can also be coupled with finite element modeling, in order to obtain the stress data necessary as inputs. More precisely, by calculating the stress tensor generated from rolling contact between specific geometries in a FEM model, it can be possible then to apply the damage evolution analysis on stresses directly associated with real geometry.

A combination of an analytical-numerical model, based on theoretical considerations, and a FEM model based on the geometry of real components is then proposed

in this thesis work. The aim of the synergistic functioning of these two models is to obtain a first approximation of rolling fatigue initiation behavior in bearing gears.

The objectives reached consisted then in the realization of:

- An analytical-numerical model realized in Matlab. Starting from Hertzian theory of rolling contact, the ideal stress tensor can be evaluated, and applying Continuum Damage Mechanics is possible to perform a damage evolution analysis, able to predict fatigue initiation life
- A FEM model of the section of a real planetary gearbox, able to simulate the transmission of torque and rotation along the system and analyze the stresses generated inside the complex structure. The output of this complete model can be used to refine the analysis and obtain the stress tensor resulting from the rolling contact of the real geometries

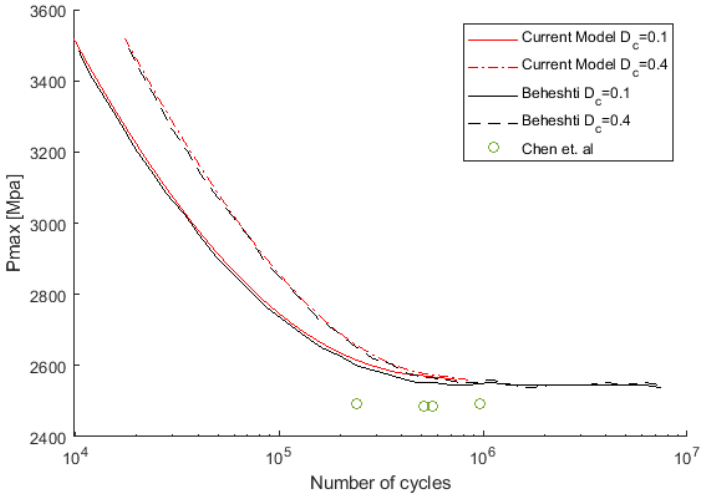


Figure 6.1: *Comparison of number of cycles to initiation for different maximum Hertzian pressure, on AISI* 52100*, between Beheshti and Chen references [16], [86] and model's results*

The analytical-numerical model has been validated with reference to literature, and has proven to be able to predict with good approximation the fatigue initiation life of bearing alloys in low friction regime (Fig. 6.1). A lower limit in the maximum pressure is shown, at which the generated maximum shear stress does not exceed the fatigue threshold specific of the material, preventing the formation of damage according to CDM theory.

The stress tensor generated is in good accordance with theory in all its components and gives a solid basis for the evaluation of the critical shear stress, indicated as the main driving factor for rolling contact fatigue. From the critical stress tensor, both the first initiation in a critical region (Fig. 6.1), and an in-depth fatigue initiation analysis (Fig. 6.2) has been run, with good results at low friction coefficients also for the latter.

These results shown in Fig. 6.2 are interesting as a basis for the study of crack initiation and propagation in deeper regions, which can result in catastrophic events.

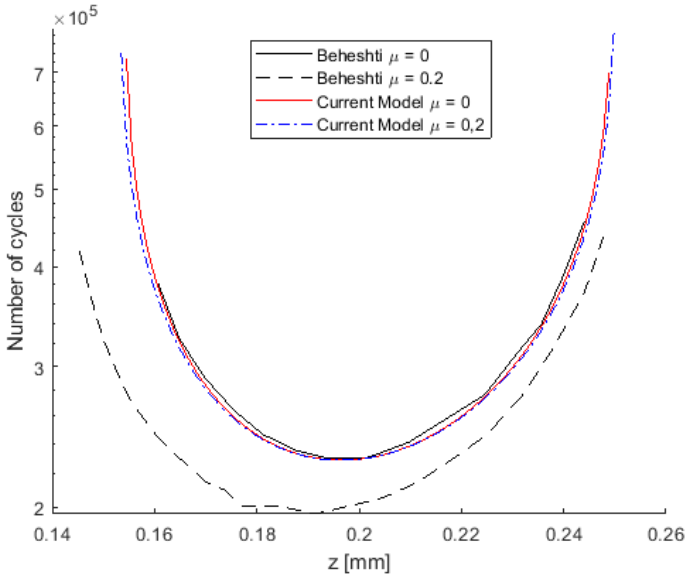


Figure 6.2: *Comparison of number of cycles to initiation at different depths between Beheshti [16] reference curve and the current model*

Moreover, the possibility to add a residual stress field in the stress tensor evaluation is proposed. It has been shown that the presence of a compressive stress field, achieved with a carburizing treatment in the case of the considered alloys, has a beneficial effect. The residual stresses are able to mitigate the amplitude of the resulting critical shear stress, improving the fatigue resistance of the material. This effect can be explained by the reduction of the principal stresses, which takes part in the formulation of the maximum shear stress, taken as the critical stress factor. As a consequence, higher external pressure is required to generate a shear able to exceed the endurance limit over which damage is predicted to form.

The FEM model instead has been developed using Abaqus, based on its CAD engine to create the parts. The components studied were imported from the model of a real helicopter's planetary gearbox. From the general model, the single components were extracted and refined to eliminate the parts which would have generated poor meshing.

The analysis has been assessed on to a section of the general system, which consisted of five planetary bearing gears acting around a sun gear. One of the planet gears has been isolated with its bearings as the subject of the FEM model.

Starting from considerations on torque and rotation transmission through shafts and gears, a model has been gradually built, component by component, to check the correct interaction feasibility. Once the sun and the planet gears have been placed in contact, the rotation and torque transmission through the teeth have been checked. The stresses generated were not able to cause the failure of the material, allowing a first safety check.

Subsequently, the implementation of the rolling contact between the rollers and the planet gear has been the subject of the model. A demonstrative local model has been analyzed to verify the effective rolling of the elements and the formation of contact

stresses during the process. This validation allowed to build the final model, with the addition of the bearings and of the section of the outer ring which acted as guide and boundary to the planet's rotation.

The cage which contained the bearings in the original drawings has been approximated with a beam structure, in order to reduce the complexity of the model, which already was elevated. These beams were given an ideal section equal to the average section of the original bearing cage.

Once all the components were placed into position, the analysis was performed applying torque and rotation at the opposite ends of the system. The values of the torque were derived from the specifications of the actual engines mounted on the helicopter from which the gearbox has been taken.

The resulting Von Mises equivalent stress resulting from the analysis on the planet gear's inner race is shown in Fig. 6.3.

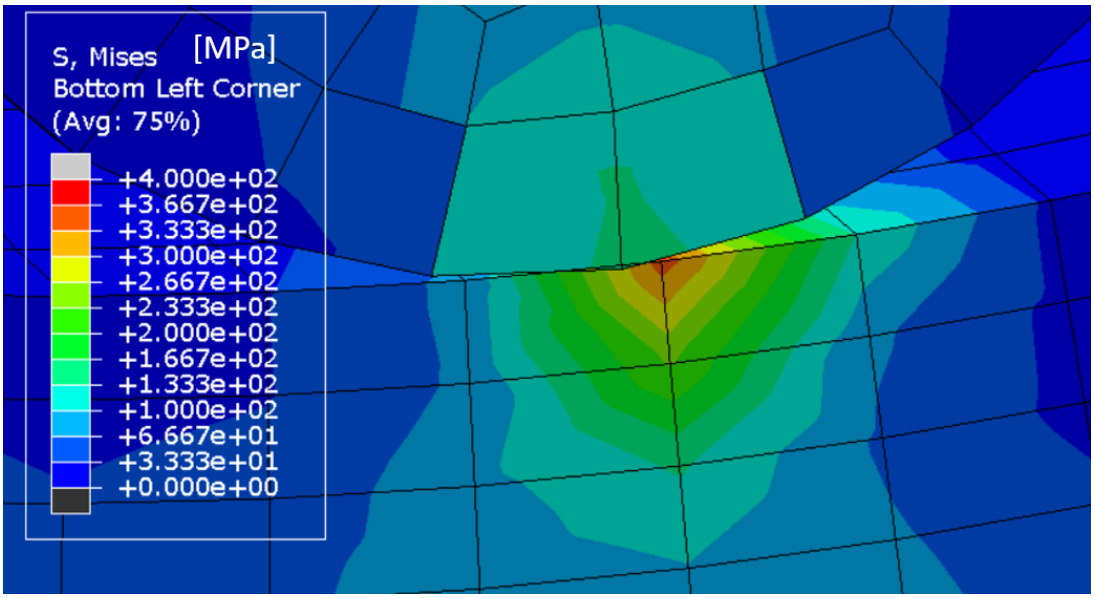


Figure 6.3: *Detail of the resulting contact stresses at the contact between a roller and the upper part of the planet gear*

The distribution of the stresses has been discussed and its values, if compared with the endurance limits introduced in the analytical model (Sec. 3.9), falls inside a safe range for the bearing alloys proposed in the literature. This comparison can be taken as good for a first approximation but, if looking at the principal stresses generated at the same contact, the behavior is not satisfactory.

This can be a consequence of the complexity of the model analyzed, which prevents refining the structure at desired levels (μm), and loses the accuracy in the rolling contact due to the elaborate transmission of the loads across the system.

A further model to be implemented to achieve a more complete analysis has been proposed and is under realization. A local single roller structure has been isolated from the complete model in order to have a simple system allowing to refine the mesh at an order of magnitude comparable with the critical depths indicated from theory and the analytical model.

The input for this supplementary analysis will be a force derived from the stresses

generate inside the beams simulating the bearing gear in the complete model. Due to the virtual one-dimensional nature of these elements, the stresses obtained will be axial. By multiplying this stress value for the section of the beam, assigned to simulate a stiffness comparable with the one of the bearing cage, a normal force can be obtained. This force is then applied vertically to a single roller, placed in conforming contact with a surface reproducing the geometry of the planet gear's inner race. The rolling movement of this element generates the contact that is expected to give a more refined stress tensor which can be introduced in the analytical model in order to perform an approximated fatigue life prediction, relative to the specific component and load conditions which has been chosen.

The general structure of this model has been presented and the results have not been presented because of time problems, but are expected in the next days before the presentation.

It can be finally concluded that the simultaneous utilization of analytical and numerical models can be a good way to overcome the respective limitations of the two approaches. Despite the complexity of the phenomenon, rolling contact fatigue can be faced from different perspectives, allowing to tackle the problem in more simple parts. In this way, starting from a theoretical basis integrated with drawings and properties derived experimentally, a more complete prediction of fatigue initiation life for rolling contacts, and in particular bearing gears, can be achieved.

References

[1] H. T.A., *Rolling bearing analysis*. New York (US): Wiley, 2001.

[2] J. Super. (). "Super puma crash investigation norway," [Online]. Available: https://www.youtube.com/watch?v=0ZbTN6pM2SE&list=PLgBqwf2A_K0qW5qXGSBVJwfZHflJpbNbT&index=1.

[3] W. Littmann, "The mechanism of contact fatigue," *NASA special report*, vol. SP-237, 1969.

[4] R. N. Ding Y., "Spalling formation mechanism for gears," *Wear*, vol. 254, no. 12, pp. 1307–1317, 2003.

[5] K.L. Johnson, *Contact mechanics*. Cambridge (UK): Cambridge University Press, 1985, ISBN: 0521255767.

[6] B. H.K., "Steels for bearings," *Progress in Materials Science*, no. 57, pp. 268–435, 2012.

[7] D. B. Depouhon P. Sola C., "A stress based critical-plane approach for study of rolling contact fatigue crack propagation in planet gears," *Airbus - Department of Research*, Airport International Marseille Provence - France.

[8] P. A. Lundberg G., "Dynamic capacity of rolling bearings," *Acta Polytechnica 7, Mechanical Engineering Series*, vol. 1, no. 3, pp. 1–52, 1947.

[9] H. T. Ioannides E., "A new fatigue life model for roller bearings," *Journal of Tribology*, vol. 107, pp. 367–378, 1985.

[10] M. J. Chiu Y.P. Tallian T.E., "An engineering model of spalling fatigue failure in rolling contact—the subsurface model," *Wear*, vol. 17, pp. 433–446, 1971.

[11] T. T.E., "Simplified contact fatigue life prediction model. Part I; review of published models. Part II: new model," *Journal of Tribology*, vol. 114, pp. 207–220, 1992.

[12] P. I. Sraml M. Flasker J., "Numerical procedure for predicting the rolling contact fatigue crack initiation," *International Journal of Fatigue*, vol. 25, pp. 585–595, 2003.

[13] S. F. Slack T., "Explicit finite element modeling of subsurface initiated spalling in rolling contacts," *Tribology International*, vol. 43, pp. 1693–1702, 2010.

[14] L. J., "How to use damage mechanics," *Nuclear Engineering and Design*, vol. 80, pp. 233–245, 1984.

[15] M. S., *Continuum damage mechanics. A continuum mechanics approach to the analysis of damage and fracture*. Ontario (CAN): Springer, Solids Mechanics and its applications, vol. 185, 2012, ISBN: 978-94-007-2665-9.

[16] K. M. Beheshti A., "On the prediction of fatigue crack initiation in rolling/sliding contacts with provision for loading sequence effect," *Tribology International*, vol. 44, pp. 1620–1628, 2011.

[17] T. modal shop an MTS company. (). "Bearings ntd quality inspection," [Online]. Available: https://www.modalshop.com/ndt-ram-bearing-testing?ID=1338.

[18] L. O.P., "Design of machine elements," in *Springer Handbook of Mechanical Engineering*, A. E. Grote K.H., Ed. Springer, 2009, ch. 6.

[19] e-Krishi Shiksha. (). "Lesson 27 introduction to bearings," [Online]. Available: http://ecoursesonline.iasri.res.in/mod/page/view.php?id=125536.

[20] ISO. (). "Standards by iso/tc4, rolling bearings," [Online]. Available: https://www.iso.org/committee/45544/x/catalogue/.

[21] Z. E.V., "Selection of rolling-element bearing steels for long-life applications," in *Effect of steel manifacturing processes on the quality of bearing steels*, H. J.J.C., Ed. Philadeplhia (US): ASTM, 1988, pp. 5–43.

[22] ——, "Rolling bearing steels - a technical and historical perspective," *Materials Science Technology*, no. 27, 2011.

[23] B. E.N., "Materials for rolling element bearings," in *Bearing design – historical aspects, present technology and future problems.* A. W.J., Ed. San Antonio (US): ASME, 1980, pp. 1–46.

[24] F. R. Cappel J. Weinberger M., "The metallurgy of roller-bearing steels," *Steel Grips*, no. 2, pp. 261–268, 2004.

[25] H. J. Hollox G.E. Hobbs R.A., "Lower bainite bearings for adverse environments," *Wear*, no. 68, pp. 229–240, 1981.

[26] H. F., "The hisotry of SKF3," *Ball Bearings Journal*, no. 231, pp. 231–232, 1987.

[27] K. Y. Tsubota K. Sato T., "Bearing steels in the 21st century," in *Bearing steels: into the 21st century*, G. W. Hoo J.J.C., Ed. Philadelphia (US): ASTM, 1998, pp. 202–215.

[28] L. S. Lund T., "Determination of oxygen and oxidic non-metallic inclusion contents in rolling bearing steels," *Ball Bearings Journal*, no. 231, pp. 36–47, 1987.

[29] O. L. Lund T., "Improving production, control and properties of bearing steels intended for demanding applications," in *Advances in the production and use of steel with improved internal cleanliness*, M. J. J.K., Ed. Philadelphia (US): ASTM, 1999, pp. 32–48.

[30] A. T. Furumura K. Murakami Y., "Development of long life bearing steel for full film lubrication and for poor and contaminated lubrication," *Motion Control*, no. 1, pp. 30–36, 1996.

[31] T. B. Zhang L., "Inclusions in continuous casting of steel," presented at the XXIV national steelmaking symposium, Morelia, Mexico, 2003, pp. 138–183.

[32] S. T. Sadeghi F. Jalalahamdi B., "A review of rolling contact fatigue," *Journal of Tribology*, no. 131, 2009.

[33] F. T. Yoshioka T., "Measurement of initiation and propagation time of rolling contact fatigue cracks by observation of acoustic emission and vibration," in *Interface Dynamics*, G. M. Dowson C. Taler C.M., Ed. Amsterdam: Elsevier, 1988, pp. 29–33.

[34] Z. E.V., "STLE life factors for rolling bearings," *STLE Publication*, vol. SP-34, 1992.

[35] B. M. Keer L.M., "A pitting model for rolling contact fatigue," *Journal of Lubrication Technology*, vol. 105, pp. 198–205, 1983.

[36] O. P. Benedetti M. Fontanari V., "Influence of residual stresses on fatigue behaviour of surface treated gears," *Proceedings of the XXX AIAS Conference*, pp. 263–272, 2001.

[37] V. M., "Failure of metals," 2017.

[38] S. N. Fleming J.F., "Mechanics of crack propagation in delamination wear," *Wear*, no. 44, pp. 39–56, 1977.

[39] G. A.A., "The phenomena of rupture and flow in solids," *Philosophical Transactions of the Royal Society of London*, vol. A, no. 221, pp. 582–593, 1921.

[40] I. G., "Analysis of stresses and strains near the end of a crack traversing a plate," *Journal of Applied Mechanics*, no. 24, pp. 361–364, 1957.

[41] A. W. Paris P.C. Gomez M.P., "A rational analytic theory of fatigue," *The Trend in Engineering*, no. 13, pp. 9–14, 1961.

[42] E. F. Paris P.C., "A critical analysis of crack propagation laws," *Journal of Basic Engineering*, pp. 528–534, 1963.

[43] D.-J. R. Roylance B.J. Williams J.A., "Wear debris and associated wear phenomena," *Proceedings of the Institution of Mechanical Engineers*, vol. Part J, no. 214, 2000.

[44] V. M., "Failure of metals," 2017.

[45] Hyde R.S., "Contact fatigue of hardened steel," *ASM Handbook*, p. 19, 1996.

[46] Santus C., Beghini M., Bartilotta I., "Surface and subsurface rolling contact fatigue characteristic depths and proposal of stress indexes," *International Journal of Fatigue*, no. 45, pp. 71–81, 2012.

[47] Tallian T.E., *Failure Atlas for Hertz Contact Machine Elements*. New York (US): ASME Press, 1992, ISBN: 978-0791800089.

[48] Glodez S., Ren Z., Fajdiga G., "Computational modelling of the surface fatigue crack growth on gear teeth flanks," *Communications in Numerical Methods in Engineering*, vol. 8, no. 17, pp. 529–541, 2001.

[49] Sraml G., Fajdiga M., "Fatigue crack initiation and propagation under cyclic contact loading," *Engineering Fracture Mechanics*, vol. 9, no. 76, pp. 1320–1335, 2009.

[50] Bormetti E., Donzella G., Mazzù A., "Surface and subsurface cracks in rolling contact fatigue of hardened components," *Tribology Transactions*, vol. 3, no. 45, pp. 274–283, 2002.

[51] Moyer C.A., "Fatigue and Life Prediction of Bearings," in *ASM Handbook*. ASM, 1996.

[52] Nélias D., Dumont M.L., Champiot F., "Role of inclusions, surface roughness and operating conditions on rolling contact fatigue," *Transactions of the ASME*, no. 121, pp. 240–251, 1999.

[53] Way S., "Pitting due to rolling contact," *Journal of Applied Mechanics*, vol. A49, no. 3, 1935.

[54] O. T. Kaneta K. Murakami Y., "Growth mechanism of subsurface crack due to Hertzian contact," *Journal of Tribology*, vol. 108, pp. 134–139, 1986.

[55] Dallago M., Benedetti M., Ancellotti S., "The role of lubricating fluid pressurization and entrapment on the path of inclined edge cracks originated under rolling–sliding contact fatigue: Numerical analyses vs. experimental evidences," *International Journal of Fatigue*, no. 92, pp. 517–530, 2016.

[56] Bower A.F., "The influence of crack face friction and trapped fluid on surface initiated rolling contact fatigue cracks.," *Journal of Tribology*, vol. 4, no. 110, pp. 704–711, 1988.

[57] Kaneta K., Murakami Y., Yatsuzuka H., "Mechanism of crack-growth in lubricated rolling sliding contact," *ASLE Transactions*, no. 28, pp. 407–414, 1985.

[58] Jin X., Keer L.M., Wang Q., "Behavior of a fluid filled subsurface crack under moving Hertzian loading," *Proceedings of ASME/STLE International Joint Tribology Conference*,

[59] Kaneta K., Murakami Y., "Effects of oil hydraulic pressure on surface crack growth in rolling/sliding contact," *Tribology International*, vol. 20, pp. 210–217, 1987.

[60] Murakami Y., Nemat-Nasser S., "Growth and stability of interacting surface flaws of arbitrary shapes," *Engineering Fracture Mechanics*, vol. 12, pp. 193–210, 1983.

[61] Johnson K.L., "The strength of surfaces in rolling contact," *Proceedings of the Institution of Mechanical Engineers*, vol. 203, pp. 151–163, 1989.

[62] Datsyshyn O.P., Panasyuk V.V., Pryshlyak R.E., "Paths of edge cracks in rolling bodies under the conditions of boundary lubrication," *Materials Science*, vol. 37, pp. 363–373, 2001.

[63] Datsyshyn O.P., Levus A.B., "Propagation of an edge crack under the pressure of liquid in the vicinity of the crack tip," *Materials Science*, vol. 39, pp. 754–757, 2003.

[64] Murakami Y., Kaneta K., Yatsuzuka H., "Analysis of surface crack propagation in lubricated rolling contact"," *ASLE Transactions*, no. 28, pp. 60–68, 1985.

[65] Benuzzi D., Bormetti E., Donzella G., "Modelli numerici per lo studio della propagazione di cricche superficiali da rolling contact fatigue in presenza di fluido," *Proceedings of the XXX AIAS Conference*, pp. 293–303, 2001.

[66] Dubourg M.C., Lamacq V., "A predictive rolling contact fatigue crack growth model: onset of branching, direction, and growth – Role of dry and lubricated conditions on crack patterns," *Transactions of the ASME*, vol. 124, pp. 680–688, 2002.

[67] Tal1ian T.E., Chiu Y.P., Van Amerongen E., "Prediction of traction and microgeometry effects on rolling contact fatigue life," *Transactions of the ASME*, vol. 100, pp. 156–166, 1978.

[68] Miller G.R., Keer L.M., Cheng H.S., "On the mechanics of fatigue crack growth due to contact loading," *Proceedings of the Royal Society of London*, vol. A, no. 387, pp. 197–289, 1985.

[69] Bucher E., Knothe K., Thelier A., "Normal and tangential contact problem of surfaces with measured roughness," *Proceedings of the 5th International Conference in Contact Mechanics and Wear of Rail/Wheel Systems*, pp. 96–103, 2000.

[70] El Refaie M., Halling J., "An experimental study of the apparent area of contact under nominally Hertzian contact of rough surfaces," *Proceedings of Tribology Convention 1969, Institution of Mechanical Engineers*, vol. 183, pp. 116–124, 1969.

[71] Sebara J., Benhe D., "Influence of surface waviness and roughness on the normal pressure distribution in the Hertzian contact," *ASME Transactions*, vol. 109, pp. 462–480, 1987.

[72] Cheng W.W., Cheng H.S., "Semi-analytical modeling of crack initiation dominant contact fatigue life for roller bearings," *Journal of Tribology*, vol. 119, pp. 233–240, 1997.

[73] Adamini R., Donzella G., La Vecchia M.G., "Fatica da contatto di camme cementate e nitrurate," *Proceedings of 18th Convegno sui Trattamenti Termici*, 2001.

[74] Cheng H.S., Keer L.M., Mura T., "Analytical modelling of surface pitting in simulated gear-teeth contacts," *SAE Technical Paper*, no. 841086, pp. 4987–4995, 1984.

[75] Clarke T.M., Miller G.R., Keer L.M., "The role of near-surface inclusions in the pitting of gears," *ASLE Transactions*, vol. 28, no. 1, pp. 111–116, 1985.

[76] Dubourg M.C., Villechaise B., "Analysis of multiple fatigue cracks, Part I: Theory," *Journal of Tribology*, vol. 114, pp. 455–461, 1992.

[77] Dubourg M.C., Villechaise B., Godet M., "Analysis of multiple fatigue cracks, Part II: Results," *Journal of Tribology*, vol. 114, pp. 462–468, 1992.

[78] Akama M., Kiuchi A., "Fatigue crack growth under non-proportional mixed loading in rail and wheel steel part 2: sequential mode I and mode II loading," *Applied Sciences*, vol. 9, no. 2866, 2019.

[79] O'regan S.D., Hahn G.T., Rubin C.A., "The driving force for mode II crack growth under rolling contact," *Wear*, vol. 101, no. 4, pp. 333–346, 1985.

[80] Blake J.W., Cheng H.S., "A surface pitting life model for spur gears: part I – life prediction," *Journal of Tribology*, vol. 113, no. 4, pp. 712–718, 1991.

[81] Choi Y., Liu C.R., "Rolling contact fatigue life of finish hard machined surfaces: part 1, model development," *Wear*, vol. 261, no. 5-6, pp. 485–491, 2006.

[82] Choi Y., "Spall progression life model for rolling contact verified by finish hard machined surfaces," *International Journal of Fatigue*, vol. 262, no. 1-2, pp. 24–35, 2010.

[83] Ding Y., Gear J.A., "Spalling depth prediction model," *Wear*, vol. 267, no. 5-8, pp. 1181–1190, 2009.

[84] Ding Y., Jones R., Kuhnell B., "Numerical analysis of subsurface crack failure beneath the pitch line of a gear tooth during engagement," *Wear*, vol. 185, no. 1-2, pp. 141–149, 1995.

[85] ——, "Elastic–plastic finite element analysis of spall formation in gears," *Wear*, vol. 197, no. 1-2, pp. 197–205, 1996.

[86] Chen L., Chen Q., Shao E., "Study on initiation and propagation angles of subsurface cracks in GCr15 bearing steel under rolling contact," *Wear*, vol. 133, no. 2, pp. 205–218, 1989.

[87] Murakami Y., Endo M., "Effects of defects, inclusions and inhomogeneities on fatigue strength," *Fatigue*, vol. 16, pp. 163–182, 1994.

[88] Donzella G., Mazzù A., Pola A., "Una procedura per il calcolo della pressione limite di componenti induriti superficialmente soggetti a fatica per contatto," *Proceedings of the XXIX AIAS Conference*, pp. 407–416, 2000.

[89] Lunden R., "Cracks in railway wheels under rolling contact load," *International Wheelset Congress*, pp. 163–167, 1992.

[90] Hellier A.K., McGirr M.B., Corderoy D.J.H., "A finite element and fatigue threshold study of shelling in heavy haul rails," *Wear*, vol. 144, pp. 289–306, 1991.

[91] Sakae C., Ohkoniori Y., Murakami Y., "Mode II Stress Intensity Factors for spalling cracks in backup roll," *Internal Report*, 1999.

[92] Murakami Y., Takahashi K., Kusumoto R., "Threshold and growth mechanism of fatigue cracks under mode II and III loadings," *Fatigue Fracture of Engineering Materials Structures*, vol. 26, pp. 523–531, 2003.

[93] Murakami Y., Sakae C., Hamada S., "Mechanism of rolling contact fatigue and measurement of ΔK_{IIth} for steel," in *Engineering Against Fatigue*, Baynon J.H., Brown M.W., Lindley T.C. et al., Ed. Rotterdam (NL): A.A. Balkema, 1999, pp. 473–485.

[94] Murakami Y., Hamada S., "A new method for the measurement of mode II fatigue threshold stress intensity factor range $\Delta K_{\tau th}$," *Fatigue Fracture of Engineering Materials Structures*, vol. 20, pp. 863–870, 1999.

[95] Rosenfield A.R., "A dislocation theory approach to wear," *Wear*, vol. 72, pp. 97–103, 1981.

[96] Sin H.C., Suh N.P., "Subsurface crack propagation due to surface traction in sliding wear," *Journal of Applied Mechanics*, vol. 51, pp. 317–323, 1984.

[97] Yoshimura H., Rubin C.A., Hahn G.T., "A technique for studying crack growth under repeated rolling contact," *Wear*, vol. 95, pp. 29–34, 1984.

[98] Kaneta M., Murakami Y., "Propagation of semi-elliptical surface cracks in lubricated rolling/sliding elliptical contacts," *Journal of Tribology*, vol. 113, pp. 270–275, 1991.

[99] Doquet V., Bertolino G., "Local approach to fatigue cracks bifurcation," *Journal of Fatigue*, vol. 30, pp. 942–950, 2008.

[100] Wong S.L., Bold P.E., Brown M.W., "Fatigue crack growth rates under sequential mixed-mode I and II loading cycles," *Fatigue Fracture of Engineering Materials Structures*, vol. 23, pp. 667–674, 2000.

[101] Doquet V., Pommier S., "Fatigue crack growth under non-proportional mixed-mode loading in ferritic–pearlitic steel," *Fatigue Fracture of Engineering Materials Structures*, vol. 27, no. 11, pp. 1051–1060, 2004.

[102] Pook L.P., Sharples J.K., "The mode III fatigue crack growth threshold for mild steel," *International Journal of Fracture*, vol. 15, R223–R226, 1979.

[103] Harada S., Kobayashi Y., Kuroshima Y., "A mechanics condition governing incipient fatigue crack gorwth under cyclic torsion," in *Proceedings of the Sixth International Conference on Biaxial/Multiaxial Fatigue and Fracture*, Freitas M.M., Ed. ESIS, 2001, pp. 639–646.

[104] Tanaka K., Akiniwa Y., Yu H., "The propagation of a circumferential crack in medium-carbon steel bars under combined torsional and axial loadings," in *Mixed-Mode Crack Behavior*, Miller K.J., McDowell D.L., Ed. West Conshohocken, PA (US): ASTM, 1979, pp. 295–311.

[105] Ritchie R.O., McClintock F.A., Nayeb-Hashemi H., "Mode III fatigue crack propagation in low alloy steel," *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, vol. 13, pp. 101–110, 1982.

[106] Nayeb-Hashemi H., McClintock F.A., Ritchie R.O., "Effects of friction and high torque on fatigue crack propagation in mode III.," *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, vol. 13, pp. 2197–2204, 1982.

[107] Houlier F., Pineau A., "Propagation of fatigue cracks under polymodal loading.," *Fatigue Fracture of Engineering Materials Structures*, vol. 5, pp. 287–302, 1982.

[108] Akama M., "Fatigue crack growth under non-proportional mixed loading in rail and wheel steel part 1: sequential mode I and mode III loading," *Applied Sciences*, vol. 9, no. 2006, 2019.

[109] Reichard D.W., Parker R.J., Zaretsky E.V., "Residual stress and subsurface hardness changes induced during rolling contact," *NASA technical note TN D-4456*, 1968.

[110] Reti T., "Residual stresses in carburised, carbonitrided and case-hardened components (Part 1)," *Heat Treatment of Metals*, vol. 4, pp. 83–96, 2003.

[111] Hizli H., Kaleli T., Gur H., "Measurement of residual stresses in the carburized steels by non-destructive techniques," *Proceedings of 18th International Metallurgy Materials Congress, UCTEA Chamber of Metallurgical Materials Engineers*, pp. 439–442, 2016.

[112] Zaretsky E.V., Parker R.J., Anderson W.J., "A study of residual stress induced during rolling," *Journal of Lubrication Technology*, vol. 91, pp. 314–319, 1969.

[113] Zhou R.S., "Surface topography and fatigue life of rolling contact bearings," *Tribology Transactions*, vol. 36, pp. 329–340, 1993.

[114] Palmgren A., *Ball and Roller Bearing Engineering*. Philadelphia,PA (US): SKF Industries, 1945.

[115] P. A. Lundberg G., "Dynamic capacity of rolling bearings," *Acta Polytechnica 7, Mechanical Engineering Series*, vol. 2, no. 4, pp. 96–127, 1952.

[116] Schlicht H., Schreiber E., Zwirlein O., "Fatigue and failure mechanism of bearings," *I Mechanical Engineering Conference Publications*, vol. 1, pp. 85–90, 1986.

[117] Shao E., Huang X., Wang C., "A method of detecting rolling contact crack initiation and the establishment of crack propagation curves," *Tribology Transactions*, vol. 31, no. 1, pp. 6–11, 1987.

[118] Leng X., Chen Q., Shao E., "Initiation and propagation of case crushing cracks in rolling contact fatigue," *Wear*, vol. 122, pp. 33–43, 1988.

[119] ISO, "Rolling bearings - Dynamic load ratings and rating life," *Draft International Standard ISO/DIS 281*, 1989.

[120] Zaretsky E.V., "Design for life, plan for death," *Machine Design*, vol. 66, no. 15, pp. 55–59, 1994.

[121] Harris T.A., McCool J., "On the accuracy of rolling bearing fatigue life prediction," *Journal of Tribology*, vol. 118, pp. 297–310, 1996.

[122] Otsuka A., Sugawara H., and Shomura M., "A test method for mode II fatigue crack growth relating to a model for rolling contact fatigue," *Fatigue Fracture of Engineering Materials Structures*, vol. 19, no. 10, pp. 1265–1275, 1996.

[123] Ioannides E., Bergling G., Gabelli A., "An analytical formulation for the life of rolling bearings," *Acta Polytechnica 7, Mechanical Engineering Series*, vol. 137, pp. 58–60, 1999.

[124] Kudish I.I., Burris K.W., "Modern state of experimentation and modeling in contact fatigue phenomenon: Part II—Analysis of the existing statistical mathematical models of bearing and gear fatigue life. New statistical model of contact fatigue.," *Tribology Transactions*, vol. 43, no. 2, pp. 293–301, 2000.

[125] Simizu S., "Fatigue limit concept and life prediction model for rolling contact machine wlements," *Tribology Transactions*, vol. 45, no. 1, pp. 39–46, 2002.

[126] Miyashita Y., Yoshimura Y., Xu J.Q., "Subsurface crack propagation in rolling contact fatigue of sintered alloy," *JSME International Journal, Series A*, vol. 46, no. 3, pp. 341–347, 2003.

[127] Kotzalas M.N., "Statistical distribution of tapered roller bearing fatigue lives at high levels of reliability," *Tribology Transactions*, vol. 127, no. 4, pp. 865–870, 2005.

[128] Bhargava V., Hahn G.T., Rubin C.A., "Rolling contact deformation, etching effects and failure of high strength steels," *Metallurgical and Materials Transactions A*, vol. 21, pp. 1921–1931, 1990.

[129] Cheng W., Cheng H.S., Mura T. Keer L.M., "Micromechanics modeling of crack initiation under contact fatigue," *Journal of Tribology*, vol. 116, pp. 2–8, 1994.

[130] Vincent A., Lormand G., Lamagnere P., "From white etching area formed around inclusions to crack nucleation in bearing steels under rolling contact," in *Bearing Steels: Into the 2st Century*, Hoo J., Green W., Ed. West Conshohocken, PA (US): ASTM Special Technical Publication, 1998, pp. 109–123.

[131] Zhou R.S., Cheng H.S., Mura T., "Micropitting in rolling and sliding contact under mixed lubrication," *Journal of Tribology*, vol. 111, pp. 605–613, 1989.

[132] Sehitoglu H., Jiang Y, "Fatigue and stress analyses of rolling contact," *College of Engineering, University of Illinois at Urbana-Champaign, Technical Report*, no. 161, 1992.

[133] Melander A., "A finite element study of short cracks with different inclusion types under rolling contact fatigue load," *International Journal of Fatigue*, vol. 19, no. 1, pp. 13–24, 1997.

[134] Lormand G., Piot D., Vincent A. et al., "Application of a new physically bsed model to determine the influence of inclusion population and loading conditions on the distribution of bearing lives," in *Proceedings of the ASTM Symposium Bearing Steel Technology*, STP1419. ASTM Publication, 2002, pp. 493–508.

[135] Harris T.A., Yu W.K., "Lundberg-Palmgren fatigue theory: considerations of failure stress and stresses volume," *Journal of Tribology*, vol. 121, pp. 85–89, 1999.

[136] Jiang Y., Sehitoglu H., "A model for rolling contact failure," *Wear*, vol. 224, pp. 38–49, 1999.

[137] Ringsberg J.W., "Life prediction of rolling contact fatigue crack initiation," *International Journal of Fatigue*, vol. 23, no. 7, pp. 575–586, 2001.

[138] Liu Y., Stratman B., Mahadevan S., "Fatigue crack initiation life prediction of railroad wheels," *International Journal of Fatigue*, vol. 23, pp. 747–756, 2006.

[139] Liu Y., Mahadevan S., "A unified multiaxial fatigue damage model for isotropic and anisotropic materials," *International Journal of Fatigue*, vol. 29, pp. 347–359, 2007.

[140] Kachanov L.M., "On the creep fracture time," *Izvestiya Akademii Nauk USSR Otd Tekh*, vol. 8, pp. 26–31, 1958.

[141] Bhattacharya B., Ellingwood B., "Continuum damage mechanics analysis of fatigue crack initiation," *International Journal of Fatigue*, vol. 20, no. 9, pp. 631–639, 1998.

[142] Krajcinovic D., "Continuum damage mechanics," *Applied Mechanics Reviews*, vol. 37, no. 1, pp. 1–6, 1984.

[143] Hult J., "Introduction and general overview," in *Continuum Damage Mechanics Theory and Applications*, D. Krajcinovic and J. Lemaitre, Ed. New York (US): Springer-Verlag, 1987.

[144] Lemaitre J., "Continuous damage mechanics model for ductile fracture," *Journal of Engineering Materials - ASME Transactions*, vol. 107, pp. 83–89, 1985.

[145] Bhattacharya B., Ellingwood B., "A new CDM-based approach to structural deterioration," *International Journal of Solids and Structures*, vol. 36, pp. 1757–1759, 1999.

[146] Bhattacharya B., "A damage mechanics-based approach to structural deterioration and reliability. PhD Thesis," Baltimore,MD (US), 1997.

[147] Lemaitre J., *A Course on Damage Mechanics*. Dordrecht (NL): Springer-Verlag, 1992.

[148] Ancellotti S., Benedetti M., Dallago M. Fontanari V., "Fluid pressurization and entrapment effects on the SIFs of cracks produced under lubricated rolling-sliding contact fatigue," *21st European Conference on Fracture, Procedia Structural Integrity*, vol. 2, pp. 3098–3108, 2016.

[149] M'Ewen E., "Stresses in elastic cylinders in contact along a generatrix (including the effect of tangential friction)," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 40:303, pp. 454–459, 1949.

[150] Poritsky H., "Stresses and deflections of cylindrical bodies in contact," *Journal of Applied Mechanics*, vol. 17, pp. 191–201, 1950.

[151] Smith J.O., Liu C.K., "Stresses due to tangential and normal loads on an elastic solid," *Journal of Applied Mechanics*, vol. 20, pp. 157–166, 1953.

[152] Sackfield A., Hills D.A., "Some useful results in the classical Hertz contact problem," *Journal of Strain Analysis*, vol. 18, pp. 101–108, 1983.

[153] ——, "A note on the Hertz contact problem: correlation of standard formulae," *Journal of Strain Analysis*, vol. 18, pp. 195–201, 1983.

[154] Williams J.A., Dwyer-Joyce R.S., "Contact between solid surfaces," in *Engineering Tribology*, Williams J.A., Ed. University of Cambridge (UK): Cambridge University Press, 2005, ch. 3.

[155] Lemaitre J., Desmorat R., *Engineering damage mechanics: ductile, creep, fatigue and brittle failures*. Berlin (GER), New York (US): Springer, 2005.

[156] Devree J.H.P., Brekelmans W.A.M., Vangils M.A.J., "Comparison of non local approaches in continuum damage mechanics," *Computers Structures*, vol. 55, pp. 581–588, 1995.

[157] Chow C.L., Wei Y.A., "Damage mechanics model of fatigue crack initiation in notched plates," *Theoretical and Applied Fracture Mechanics*, vol. 16, pp. 123–133, 1991.

[158] Beheshti A., Khonsari M.M., "A thermodynamic approach for prediction of wear coefficient under unlubricated sliding condition," *Tribology Letters*, vol. 38, pp. 347–354, 2010.

[159] Paas M.H.., Schreurs P.J., Brekelmans W.A., "A continuum approach to brittle and fatigue damage: theory and numerical procedures," *International Journal of Solids and Structures*, vol. 30, no. 4, pp. 579–599, 1993.

[160] Pasic H., "A unified approach of fracture and damage mechanics to fatigue damage problems," *International Journal of Solids and Structures*, vol. 29, no. 14/15, pp. 1957–1968, 1992.

[161] Tiejun W., Zhiwen L., "A continuum damage model for weld heat affected zone under low cycle fatigue loading," *Engineering Fracture Mechanics*, vol. 37, no. 4, pp. 825–829, 1990.

[162] All Metals Forge Group. (). "AISI/SAE 4340 Alloy Steel," [Online]. Available: %7Bhttps://www.steelforge.com/alloy-steel-4340/%7D.

[163] Matweb - Material property data. (). "AISI E 52100 Steel (100Cr6, SUJ2, UNS G52986)," [Online]. Available: %7Bhttp://www.matweb.com/search/DataSheet.aspx?MatGUID=d0b0a51bff894778a97f5b72e7317d85&ckck=1%7D.

[164] Voyiadjis G.Z., Kattan P.I., "A comparative study of damage variables in continuum damage mechanics," *International Journal of Damage Mechanics*, vol. 18, pp. 315–340, 2009.

[165] Harris T.A., Barnsby R.M., "Life ratings for ball and roller bearings," *Proceedings of the Institution of Mechanical Engineering*, vol. 215, pp. 577–595, 2001.

[166] Korsunsky A.M., James K.E., Aylott C. Shaw B.A., "Residual stresses in induction-hardened gear teeth mapped by neutron diffraction," *Journal of Strain Analysis*, vol. 37, no. 4, pp. 337–344, 2002.

[167] Leonardo Helicopters Division, *AW169 Emergecy Medical Services (EMS)*. Leonardo S.p.a., 2020.

[168] LEONARDO. (). "AW169 Next Generation Versatility," [Online]. Available: %7Bhttps://www.leonardocompany.com/en/products/aw-169%7D.

[169] Pratt Whitney Canada. (). “PW210 - Shaping a new generation of helicopters,” [Online]. Available: %7Bhttps://www.pwc.ca/en/products-and-services/products/helicopter-engines/pw210%7D.