

Further Understanding of Rolling Contact Fatigue in Rolling Element Bearings - A Review

Mostafa El Laithy^a, Ling Wang^a, Terry J. Harvey^a, Bernd Vierneusel^b, Martin Correns^b,
Toni Blass^b

^a*National Centre for Advanced Tribology at Southampton (nCATS), University of Southampton, University Road, Southampton SO17 1BJ, UK*

^b*Schaeffler Technologies GmbH & Co. KG, Georg-Schäfer-Straße 30, 97421 Schweinfurt, Germany*

Abstract

Rolling bearings are one of the most widely used components in industrial machinery. If suitably mounted, loaded, lubricated and isolated from contamination, rolling contact fatigue (RCF) is believed to be the most probable mode of failure resulting in subsurface originated failures. Over the past several decades many researchers have studied the failure and microstructural alterations in bearings such as dark etching regions and white etching bands and how operating conditions such as pressure, temperature and running time impact them during RCF. This paper aims to provide an overview of such alterations, their properties, formation mechanisms and impact on bearing failure.

Keywords: Rolling Contact Fatigue; Rolling Element Bearings; Dark Etching Regions; White Etching Bands; Subsurface Shear Stress; Inclusions; Carbides; Shakedown

1. Introduction

Rolling bearings are crucial components in tribomachinery for supporting and transmitting power while maintaining relative rotatory motion between two structures while minimizing friction losses. Bearings have been used for thousands of years but it was not until the industrial revolution where metallurgical innovations lead to the development of metal bearings to meet the growing demands of the advanced machinery. The evolving necessity of modern bearings particularly since the 1960s has warranted a better understanding of bearing operations to develop superior steels for rolling element bearing with extremely accurate geometry to enhance their performance and prolong bearing life under harsh environments [1].

If a ball or rolling element bearing is properly mounted, loaded, lubricated and well isolated from foreign contamination, rolling contact fatigue (RCF) becomes the main mode of failure [2–4]. Failure due to rolling contact fatigue can be divided into two categories, subsurface initiated and surface-initiated failures which can contribute to pitting and spalling.

Email address: mel2g13@soton.ac.uk (Mostafa El Laithy)

Surface-initiated failures initiate as a consequence of surface stress risers such as dents, scratches, surface contamination, surface roughness, textures and insufficient lubrication [5]. However, if under optimum surface and lubrication conditions, sub-surface initiated fatigue becomes the ultimate failure mode where cracks initiate at subsurface stress risers such as microstructural defects (dislocations, grain boundaries, carbides etc.), inclusions, residual stress build-up and secondary phases [5, 6] and propagate to the surface. Statistical analysis by Harris et al. [7] show that improper mounting, sealing and inadequate lubrication contributes to 90% of bearing failure.

Throughout the bearing life the failure mechanisms for surface and subsurface failure are competing with each other. Shifting the failure probability from surface-initiated to subsurface-initiated damage could pro-long the bearing life but this depends on the design constraints in the operating conditions. Some applications have fixed constraints such as assembly space, temperature or lubrication leading to severe environments such as mixed friction conditions making it challenging to avoid surface-initiated failures in certain applications. Nonetheless, if the probability of damage can be shifted towards subsurface-initiated damage, the bearing life is enhanced and the material properties become an important factor in extending the bearing life by impeding subsurface-initiated RCF. Modern manufacturing techniques such as reducing steel impurities, modern sealing and advanced filtration methods aim to reduce the occurrence of rolling contact fatigue [1].

One of the most distinctive features observed during rolling contact fatigue is microstructural alterations due to cyclic loading [3]. In 1947, Jones [8] was the first to notice that the stress induced during cyclic loading resulted in the formation of irreversible subsurface microstructural alterations within the inner raceway of the bearing commonly known today as dark etching regions (DERs) and white etching bands (WEBs). These have also been reported under different names such as troostite (tempered martensite) [8], low-temperature bainite [2, 6] and dark needle [9] and are believed to form beyond a certain threshold pressure [10]. While DERs and WEBs have been analysed and reported for decades, there still lacks a universal understanding of these altered regions, their formation mechanisms and influence of operating conditions on their development. Bearings have become essential for countless machinery used today, therefore there is a growing need to understand the effects of conditions such as load, speed, lubrication and material on the microstructural alterations to better predict bearing life.

Rolling elements such as balls or rollers in rolling bearings are typically held in an angularly spaced relationship by a cage [1]. The purpose of a bearing unit is to join rotating components in a machine. The simple cross-section of a ball bearing can be seen in Figure 1 consisting of the inner and outer rings and rolling elements in-between. Rotating components from machinery can be connected to the inner and outer rings and henceforth rotate independently or one of the rings may remain stationary while the other revolves. In rolling bearings, both rings are kept in contact with rolling elements (ball, rollers) where a radial, axial or combined load can be applied depending on the bearing type [1]. A common term referred to in the bearing industry is the over-rolling direction in which the contact point

between the ball and raceway moves relative to the inner ring [11, 12].

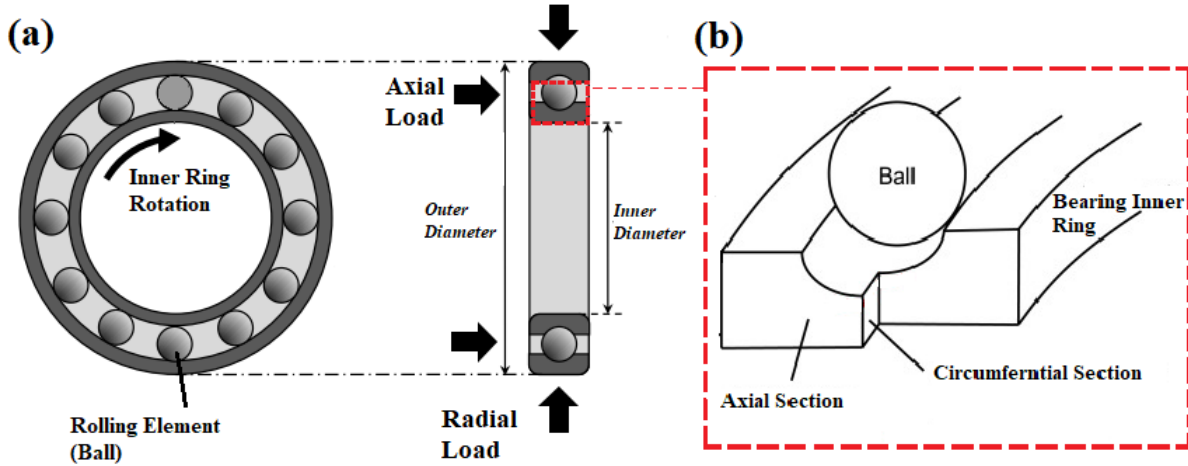


Figure 1: a) A deep groove ball bearing cross section where the rotation of each ring depends on the operation, b) Common terminologies used for sections of the inner ring. Adapted from [4].

2. Bearing Steel

Rolling bearings are usually manufactured from steels that are hardened to a high degree at least on the surface [1]. The ball bearing industry typically use SAE 52100 (100Cr6) which demonstrates sufficient hardness, wear resistance and fatigue life. Typical hardness specifications for commercial bearings may vary from 58-62 Rockwell C scale hardness equivalent to 670-840 Vickers hardness. Other steel alloys are used by manufacturers depending on the bearing application and requirements such as SAE 3310, 4620, and 8620 [1]. The typical compositions of various bearing steels are shown in Table 1. A typical 100Cr6 steel bearing consists of a microstructure with homogeneously distributed primary spherical $(Fe,Cr)_3C$ carbides in a matrix of martensite [2, 13], having a high concentration (beyond 0.7wt%) of both carbon and chromium and typically 6-12% retained austenite (RA) although there are also steel bearings with a lower RA content used commercially [6].

The use of low carbon steels (0.1-0.3wt%) are utilised for case-hardened bearings consisting of a hard surface and a relatively soft but tough core. Bearing steels with a higher carbon content such as SAE 52100 are typically used for through-hardened bearings where the microstructure is hardened throughout the cross section to provide uniform toughness and hardness [3]. The heat treatment process for the hardened steel microstructure begins with austenisation where the steel is heated to around 850 °C for \approx 15 minutes to form the face-centred cubic structure of austenite. Martensitic and bainitic structured steels are both commonly used in the bearing industry. A bainitic microstructure can be obtained later by quenching the microstructure in a salt bath at temperatures of 200-450 °C and martensite can be obtained by rapid quenching in a water/oil bath at around 60 °C [11]. Tempering is later required to improve the material toughness for practical application which consequently

Table 1: Typical composition of various steel (wt%) used in the bearing industry [14].

	C	Mn	P	S	Si	Ni	Cr	Mo
SAE 52100	0.95-1.10	0.25-0.45	≤ 0.025	≤ 0.025	0.20-0.35	-	1.30-1.60	-
SAE 3310	0.08-0.13	0.45-0.60	≤ 0.025	≤ 0.026	0.20-0.35	3.25-3.75	1.40-1.75	-
SAE 4620	0.17-0.22	0.45-0.65	0.035	0.04	0.15-0.35	1.65-2.00	-	0.20-0.30
SAE 8620	0.18-0.23	0.75-0.90	≤ 0.035	≤ 0.04	0.15-0.35	0.40-0.70	0.40-0.60	0.15-0.25

results in the decomposition of a proportion of the RA and the release of excessive residual stress. After low temperature tempering (160-220 °C), the steel softens due to the precipitation of transition carbides and cementite [15–17]. A typical martensitic through hardened bearing steel such as AISI/SAE 52100 has a complex microstructure after heat treatment consisting of tempered martensite, retained austenite, ferrite and cementite, undissolved carbides from austenisation (primary carbides) and metastable ϵ -carbides (transitional) [18]. In the bearing industry, the prior austenite grain size is accepted to have an ASTM number of 8 or larger and individual grains should not have an ASTM number smaller than 5 where the higher the ASTM number the finer the grain size [19, 20]. Baughman [21] has reported that the RCF life of bearings may be enhanced with finer grain sizes in agreement with Zaretsky [22] and Ooki et al. [23].

2.1. Retained Austenite

Austenite is considered a metastable phase at room temperatures because it exists outside its typical temperature range [24]. Large amounts of austenite (<15%) can typically be detected through optical microscopy while more accurate techniques such as x-ray diffraction methods are needed to measure lower amounts of retained austenite (<5%) accurately [25]. In the 1960s, a major U.S. aircraft company had to discard unused AISI M-50 rolling element bearings because the bore diameter expanded from that of the engine shaft diameter [20]. This was later explained as a consequence of the high content of retained austenite existing in the material decaying during operation. As a result, bearings in aerospace applications have a limited amount of retained austenite ranging from 2-5% even though other applications accept higher contents. Investigations by Abareshi and Emadoddin [26] and Shen [27] indicate that bearings with a higher RA content (up to 11.5%) exhibits improved bearing life. However, it is highlighted by Hu et al. [28] that austenite can enhance fatigue propagation due to strain-induced martensitic transformation. Nonetheless, to avoid significant changes in volume, the RA content should be <10% [25]. Volume expansion (typically 4-5%) resulting from the transformation of austenite induces significant internal stresses, contributing to dimensional instability in the bearing and under extreme cases, crack formation. The microstructure during bearing operation experiences a decay in retained austenite, causing compressive internal stresses in the region of decomposition. Martensite decay is also ex-

perenced by the microstructure leading to the development of ferrite and cementite. This decay is associated with a decrease in specific volume, generating tensile internal stresses in the decay region [29].

Zaretsky [30] reported the reliability of bearing operations to improve with a lower amount of retained austenite. Johnson et al. [31] examined the development of compressive residual stress during bearing operation and corresponding decomposition of retained austenite, concluding the magnitude of residual stress to be proportional to the decomposition of retained austenite. However, Muro et al. [32] suggested both phenomena to be independent while other factors such as shear stress, contact pressure and initial material hardening to be more potent in residual stress development. While high levels of retained austenite contribute to dimensional instability, retained austenite has also been reported to improve rolling contact fatigue life and impact strength of the material, resisting fracture when subjected to a sudden high-load. The ductility of retained austenite can assist in impeding crack growth by blunting the tip of newly formed cracks [33]. The developed compressive residual stresses arising from the decay of retained austenite during bearing operation can retard crack growth by clamping cracks closed [33]. Retained austenite can improve the steel life when the lubricant contains wear debris as it contributes to the bearing durability to resist dents on the surfaces of the raceway from the wear debris [34].

It is therefore essential for a manufacturer to acquire a level of retained austenite that provides a good balance between the improved fatigue life, impact strength and the dimensional stability of bearing components to optimize performance. This however is highly dependant on the microstructure and operating conditions such as contact pressure and operating temperature, where higher loads and temperatures intensify the rate of retained austenite transformation during operation [35]. Paladugu et al. [36] stated that retained austenite transforms more slowly in fine martensitic microstructures compared to coarser microstructures as the plastic deformation inducing the retained austenite transformation are more defined in coarser grains.

2.2. Steel Inclusions

Steel bearing reliability has significantly been enhanced since the 1970s primarily due to the availability of cleaner steels given that impurities can impact the mechanical properties of the steel based on their size, type and distribution [37]. It has been well established that cleaner steels that contain smaller and less non-metallic inclusions have longer fatigue life as the fatigue life is inversely proportional to the square-root of inclusion size [38]. General inclusions in steels consist of alumina inclusions that form during de-oxidation, sulphides and large metallic primary carbides that form on cooling during casting [37]. The oxygen concentration in steel, appearing as oxides have been proposed to correspond to RCF damage. Therefore, in modern steel bearings, the oxygen concentration is kept below 10 ppmw [11]. However, for every ton of steel with inclusions such as sulphides and oxides as low as 1 ppm, this supplies approximately $10^9 - 10^{12}$ inclusions which are $10 - 10^4$ inclusions per cubic millimetre [39]. Historically, it is believed that oxides are the most harmful inclusions shortening the RCF life of bearings [38, 40–43]. Nonetheless, awareness of the harmfulness

of sulphides has been increasing over the past decades [44, 45] leading to attempts to reduce more than just oxides in manufacturing. It is possible to reduce these inclusions to 1-10 ppm [39]. A balanced approach to reduce not only oxides but also sulphides can significantly enhance the performance compared to only reducing oxide inclusion and maintaining higher sulphur content for machinability purposes as in the past [44]. Nonetheless, the cost is the main factor determining how clean the steel components should be [46] where high cleanliness might not be cost effective in certain applications.

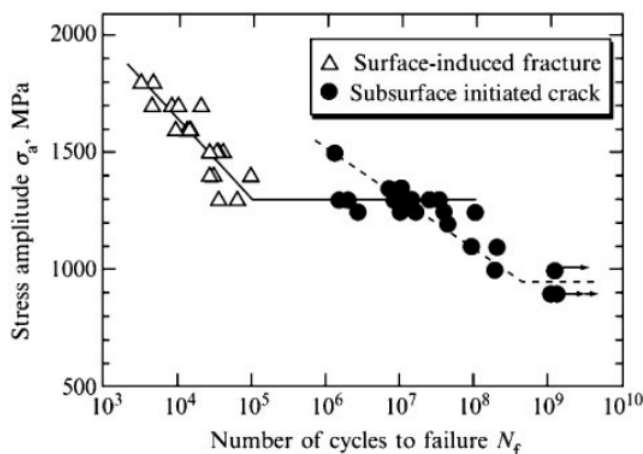


Figure 2: S-N curve from rotating bending fatigue testing high carbon bearing steel [4].

Typical bearing fatigue damages of the raceway surface arise due to smearing, plastic flow, pitting, spalling or fatigue. Small spalls can grow significantly if the bearing is not replaced due to the fact that spall edges act as stress concentrators [47]. The typical range where bearings fail due to subsurface-initiated RCF is between 10^9 - 10^{12} cycles which is commonly referred as very high cycle fatigue. An S-N curve shown in Figure 2 demonstrates different fatigue failures mechanisms for low and high cycle fatigue [48] which is more commonly linked with surface-initiated cracking. High stresses can enhance the probability of surface damage and hence failure at relatively lower cycles while higher ($>10^7$) cycle fatigue is influenced more by the subsurface microstructure where non-metallic inclusions can act as stress raisers [48] and initiation site for failure. Stress localisation between inclusions and the steel matrix interface could potentially lead to fatigue failure due to the variation of both elastic properties and thermal expansion between the matrix and inclusion during deformation and heating/cooling [49]. Leslie [50] has suggested that for high cycle fatigue ($>10^5$) almost all cracks initiate from inclusions whereas, for lower cycle fatigue ($<10^5$), slip band cracks dominate. However, no significant research has analysed how other stress raisers in the subsurface such as carbides or phase boundaries contribute to crack initiation and development.

There is a significant volume of metallurgical evidence for fatigue initiating from inclusions in bearing raceways [37]. The critical size of inclusions may vary from 5-500 μm and

decreases with increasing yield stress [51]. However, for high strength steel such as bearing steels, the bearing life is reduced if inclusions greater than 13 μm exist within the microstructure [52]. The critical size of inclusions for fatigue initiation depends on many factors such as the applied load [39] and depth from the surface [53]. However, Bhadeshia [11] has suggested that the morphology, composition and distribution of inclusions is a notably more important factor affecting fatigue which is similar to conclusions made from other reviews [39, 54–56]. Inclusions with different shapes and compositions will demonstrate different elastic and plastic properties. This coupled with the coefficient of thermal expansion of the inclusions will determine the residual stresses surrounding the inclusion which can be either tensile or compressive [37, 57].

In steel, the fatigue life is believed to be controlled by the crack propagation stage or the nucleation stage which occurs early from a debonded interface [58]. These interfaces have been previously investigated and confirmed that the interface condition between the matrix and the inclusion strongly influences the crack growth controlling RCF [37]. Rycerz et al. [59] and Longching et al. [60] state that $>80\%$ of the total life is consumed by the crack propagation stage while Yoshioka et al. [61] reported a propagation-life ratio of $< 2\%$. Nonetheless, classical RCF theories assume that the major part of the bearing life is consumed by the initiation stage [62, 63]. Enhancing the adherence of inclusions to the matrix and reducing porosity by hot isostatic pressing has been suggested to enhance the bearing life [64]. While the strength and coherency of the interface are as vital as inclusion size for determining the crack nucleation [37], leading to the possibility of inclusion engineering through processing to achieve superior bonding with the matrix. Figure 3 gives a representation of different inclusion morphologies prior and after deformation [65]. ISO 4967 and ASTM E45-05A methods use a restricted field of view analysis, assigning inclusions to different groups and rating them upon their size [37, 66]. A summary of the inclusion groups rated by the tests has been compared by Walker et al. [37] and are listed in Table 2. The inclusions are divided into groups depending on how harmful it could be for the steel. Titanium inclusions although typically exist in very small amounts can be very harmful due to sharp edges.

Representations of microscopic cleanliness levels are commonly analysed through the new standard SEP 1571 where method K determines an index of percentage area of non-metallic inclusions in the microstructure referred to an area of 1000 mm^2 [67] while an M method is used to determine the maximum size rating values for different inclusion types. Another common technique used is the ASTM E2293-08 extreme value analysis [68]. Four main types of inclusions which are typically analysed in steel and discussed in the literature are elongated sulphides (SS), elongated oxides (OS), dispersed oxides (OA) and globular oxides (OG). Oxide and sulphide inclusions can be distinguished through optical microscopy, oxide inclusions appear black in colour while sulphides appear dark grey. In the past, the ASTM E45 standard specified the magnification criteria where the inclusion rating will be determined with a compound microscope at 100x magnification (10x eyepiece and 10x objective lens). However, this standard has been updated to accommodate the growing trend in digital imaging, where a minimum resolution of 1 μm per pixel is required when

Table 2: Inclusions types found in bearing steels as rated by cleanliness standards ASTM [37, 66]

Type	Common Composition	Morphology	Comment
Type A	Sulphide	Elongated inclusions (plastically deformed)	The interface with the matrix is typically bonded sufficiently and relatively soft making it less harmful.
B type	Oxides	Hard inclusions that are broken up to form long stringer	Interfaces may be de-bonded leading to stress points and crack initiation while the orientation relative to shear stress can accelerate crack propagation.
C type	Silicates	Elongated inclusions (plastically deformed)	Rarely found in bearing steels and typically not rated by manufacturers and researchers for bearings.
D type	Typically oxides sometimes containing Ca	Hard globular inclusions <13 µm in diameter although larger sizes are possible.	Harmful due to their size, residual stresses and poor interfacial bonding with the matrix, leading to cavities surrounding the inclusions caused by an inability of the matrix to plastically flow around the inclusion. They are not deformed during hot rolling.
DS type	Typically oxides, spinels, Ca modified MnS, etc	Any oversized, globular particles >13 µm in diameter	harmful due to their large size leading to greater voids due to de-bonding of the interface, residual stresses accumulation and cracks within the brittle inclusion.
T type	Titanium carbonitrides	Small and angular	Very harmful as the inclusion is relatively brittle where cracks can initiate in the inclusions and propagate to the matrix and sharp edges act as stress concentration points.

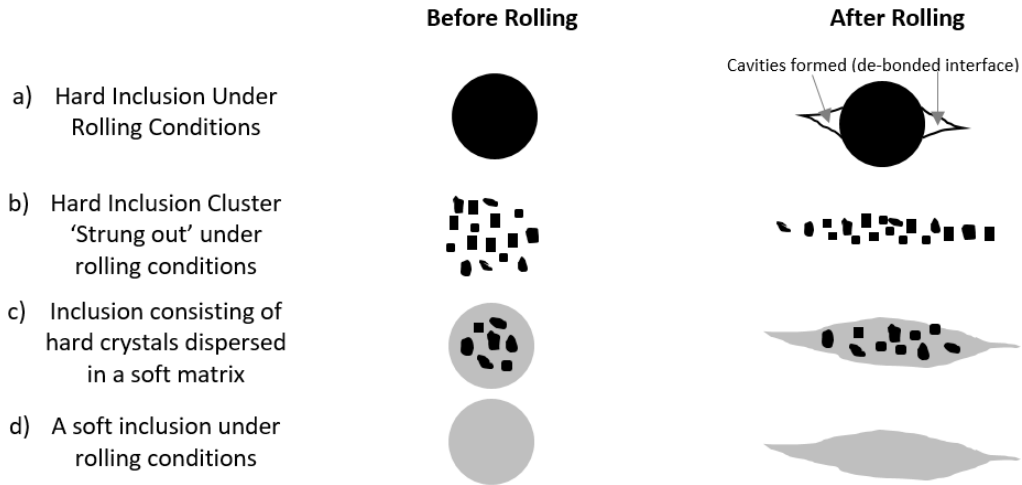


Figure 3: Schematic drawing demonstrating the morphologies of different inclusions before and after deformation. Adapted from [65]

using a 10x objective lens [69].

3. Rolling Contact Fatigue

Rolling Contact Fatigue (RCF) subsurface originated spalling is considered the ultimate failure in various components such as bearings, gears, camshaft and rail wheel. It is the most probable failure in bearings if properly aligned, mounted, not overloaded and lubricated sufficiently [5]. However, RCF is a complex phenomenon, which has made its modelling and analysis much more challenging compared to classical uniaxial fatigue. RCF is a multiaxial fatigue system where the loading history at a point within the subsurface is non-proportional and asynchronous (different stress components do not rise and fall with time in the same proportion as each other). Another feature of RCF is that the principal axis in the non-conformal contact continually change direction with stress cycles causing the plane of maximum shear stress to also vary. Therefore, it is difficult to identify the plane where maximum fatigue occurs [6]. RCF failure results in material degradation due to crack propagation arising from the alternating stress field within the material as defined by the Hertzian contact theory where the pressure experienced is in the order of GPa [11]. The typical operating condition for bearings in engineering applications can vary significantly with the maximum Hertzian contact pressure ranging from 1-3 GPa even though it can exceed this range under certain applications [70].

3.1. Hertzian Contact Theory

Rolling contact fatigue varies from classical fatigue. As the material is cyclically loaded, different regions in the material experience varying shear stresses during the load cycle [6]. RCF typically occurs within a small volume of the stressed material given the contact

stresses are highly localised. This eventually leads to subsurface or surface-initiated cracks as a consequence of the alternating contact stresses through over-rolling [71].

The resulting stress field due to two frictionless bodies pressed together into an elliptical contact whilst avoiding plasticity was initially examined by Hertz [72]. However, this analysis only involves frictionless elastic deformation which might not be fully representative of the operation of bearings since the raceway and rolling element surfaces in contact have a certain degree of roughness resulting in more complex stress fields due to asperity-asperity contacts. The pressures experienced could potentially be large enough to cause a certain amount of plasticity [11]. However, it is argued that the Hertzian model can be applied with reasonable accuracy if the oil film thickness used for lubrication is greater than the surface roughness [73]. Nonetheless, the contact area is relatively small compared to the two bodies in contact and surface curvature and is dependant on the shape of the two bodies in contact. In ball bearings, the contact geometry is usually a point contact whereas in roller bearings it is a line contact but under an applied load, the point and line contacts become elliptical and rectangular respectively as shown in Figure 4 and 5. The rectangular contact surface of roller bearings is not rectangular in reality, thus the rectangular shape is a simplified geometry for ideal hertz theory. Given that pressure accumulates within the edges of roller contacts, rollers are profiled with a crown shape at edges to make the stress distribution more uniform [74]. Therefore the contact surface can vary in shape depending on the profile and loading condition.

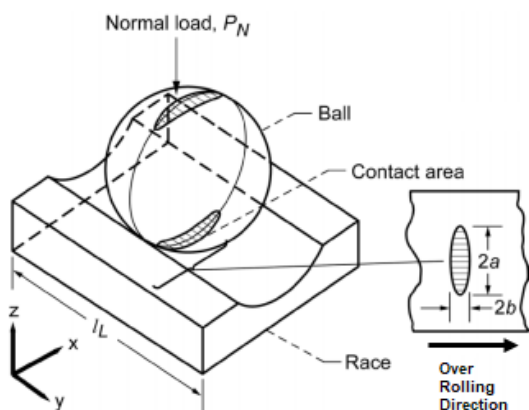


Figure 4: Ball on raceway model with elliptical contact area [75].

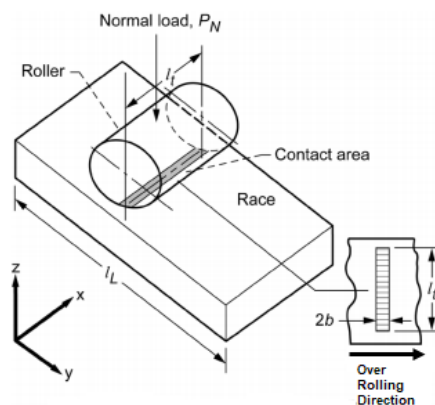


Figure 5: Cylindrical roller on raceway model with line contact [75].

The pressure normal to the contact area is not sufficient to cause significant plastic deformation even at a contact pressures of 3 GPa where the yield strength of SAE 52100 is approximately 2 GPa [37, 76]. The material does not yield as the stress states below the contact area are highly triaxial (Figures 6a & b) and the shear stress at the surface is only 10% of the surface contact pressure P_0 [37] (Figure 6c). Therefore, plastic deformation is more commonly linked to the von Mises yield criterion where the maximum von Mises

stress should be below the yield strength of the material to avoid plastic deformation [77]. The Hertzian model predicts the developed stresses initiated from elastic deformation of two curved surfaces statically pressed together. Half space models like Hertzian contact theory are suitable to calculate the stress field below the contact surface if the surfaces are sufficiently separated by a lubricant film [78]. It can be utilised to help explain subsurface microstructural changes that occur away from the contact point.

The stress state is determined by three normal stresses and six shear stress components in a Cartesian coordinate system. Given the symmetry of the tensor, the stress state can be represented by three normal stress components (σ_{xx} , σ_{yy} and σ_{zz}) and three shear stress components τ_{xy} , τ_{xz} and τ_{yz} [6]. These regularly vary in magnitude during a stress cycle which consequently changes the direction of the maximum shear stress. The maximum or principle shear stress τ_{45} (Figure 6b) varies in magnitude and direction depending on variations in the principal normal stress direction. τ_{45} acts at an angle of 45° to the principal stresses which are orientated at 0° and 90° at the center of the contact. Hence as the principal stress orientation varies away from the contact center, the maximum shear stress direction would vary under a fixed coordinate system. However, given that the maximum shear stress is commonly calculated at the centre of the contact, the notation τ_{45} is regularly used in literature given the principal shear stress is 45° to the surface at the center. The orthogonal shear stress τ_0 acts normal and parallel to the contact surface at three different planes τ_{xy} , τ_{xz} and τ_{yz} all of which reach their maximum below the surface. However, the plane with the significant τ_0 depends on the contact geometry and typically lies in the xz-plane since the short half-width of the contact is usually oriented in rolling direction and long half-width perpendicular to the rolling direction as shown in Figure 6a. The maximum shear stress (τ_{45}) due to the contact pressure P_0 occurs below the contact point as shown in Figure 6c at a depth of $0.76b$ for point contact with a value of approximately $0.32P_0$ [79] and $\tau_{0,max}$ (Figure 7) occurs at a depth of $0.5b$ at a value of $0.25P_0$. For line contacts, τ_{45} and $\tau_{0,max}$ have values of $0.3P_0$ and $0.25P_0$ respectively at depths of $0.79b$ and $0.5b$ [20]. Figure 7 demonstrates how τ_0 in the xz plane varies with the movement of the rolling element where maximum τ_0 is experienced at position of $\approx \pm 0.87 x/b$ relative to the contact area centre while the maximum amplitude of τ_{45} occurs at the centre of the contact [18].

There have been various hypothesis suggesting which stress component is the critical stress in rolling contact fatigue contributing to failure initiation and microstructural alterations. The fatigue life of rolling element bearings is heavily influenced on the contact geometry where fatigue tests have shown that for the same maximum Hertzian pressure, longer fatigue life is achieved when the contact area ratio $\frac{b}{a}$ increases [80–82]. Popinceanu et al. [83] observed that when $\frac{b}{a}$ increases, the orthogonal stress and von Mises stress decrease and therefore concluding τ_0 be the most influential of the stress components since the fatigue life can be enhanced by reducing the orthogonal stress. Increasing the $\frac{b}{a}$ ratio would reduce the stressed volume of the material which would increase the fatigue life based on the Lundberg-Palmgren equation [62]. However, for microstructural alterations during RCF, there have been various opinions on which stress component is most responsible, some have suggested the microstructural alterations to occur at the depth of the maximum orthogonal

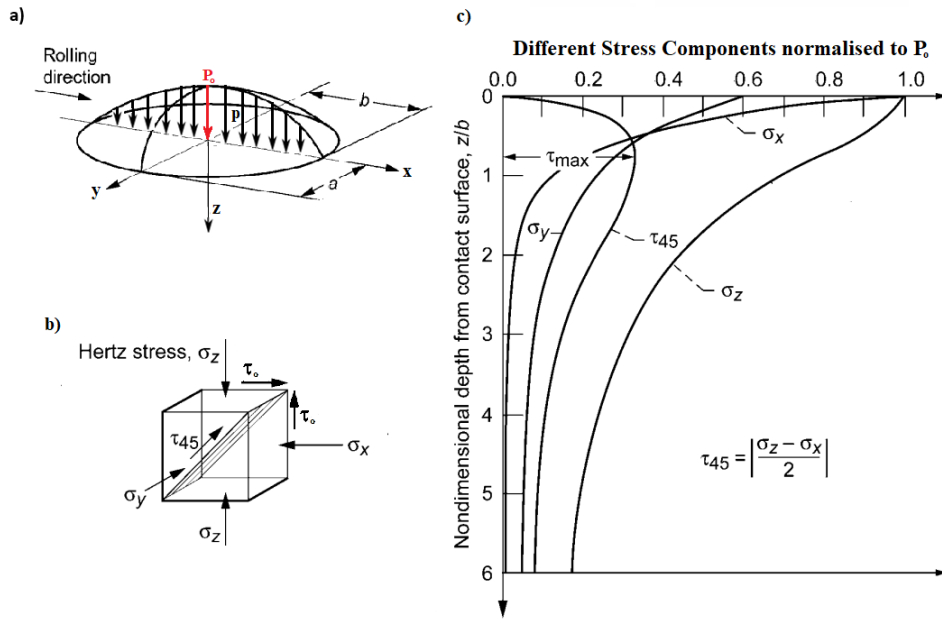


Figure 6: Parabolic distribution of contact pressure along elliptical contact with maximum contact pressure at the centre. b) Direction of principal stresses and shear stresses in the material subsurface. c) Variation of normalised principle stresses and maximum shear stress τ_{45} with normalised depth z/b in z direction at centre of contact. Adapted from [20].

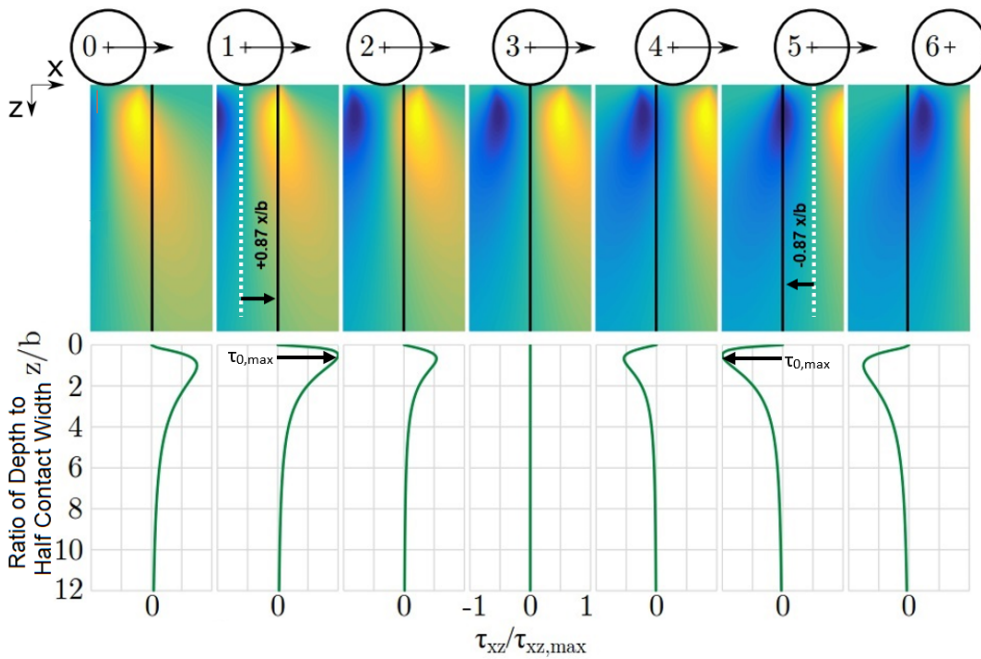


Figure 7: a) Variation in τ_{xz} with normalised depth at different positions within contact area between rolling element and raceway.

shear stress while others have linked it to the principal shear stress τ_{45} [9, 84–88], the von Mises stress [83, 89, 90] and Dang Van criterion [77, 91, 92].

3.2. Stages of Rolling Contact Fatigue

Given the bearing life is limited to the probability of subsurface initiated spalling, there have been various rolling contact fatigue empirical and analytical models to predict the performance of the rolling bearings [6]. Rolling contact fatigue is known to appear in two forms: local damage such as surface and subsurface initiated cracks and as gradual alterations in the bulk material. The material degradation due to RCF occurs in different rolling bearing components (inner ring, outer ring and rolling elements). Nonetheless, majority of literature has investigated the microstructural alterations in the inner rings. The material degradation process can be divided into three stages referred to as (i) shakedown, (ii) steady-state elastic response and (iii) instability [6]. Voskamp [29] suggested a change in material response during each of these stages during RCF. The microstructural alterations observed at approximately 0.2 mm below the inner raceway of a DGBB has been summarized by Voskamp [29] and can be seen in Figure 8 coupled with the approximate duration in running cycle in which these stages operate.

3.2.1. Shakedown

Indeed, the Hertzian model is only suitable for elastic deformation. However, it can still be used for a clean steel that has reached its shakedown point [37]. Shakedown is obtained when the bearing behaves elastically under further loading beyond the elastic limit [93]. When the elastic limit is exceeded in the initial cycle, a small degree of plastic deformation occurs which introduces residual stresses. This subsequently alters the material properties [29]. During shakedown, residual stress is built up while the material strength and micro-yield stress are both enhanced as an outcome of work hardening during the early bearing life [94]. This is believed to be due to the partial transformation of retained austenite in the microstructure to martensite [37, 95] where the build-up of compressive residual stress is a consequence of the local shape variation constrained by the surrounding matrix [29].

Consequently, in the second stage, the load experienced by the material is subjected to both contact stresses and residual stresses introduced previously. Plastic micro-deformation befalls in previously unaffected regions in the first cycle resulting in work hardening and residual stress build-up as earlier. A certain number of running cycles is required ($\approx 10^3$ cycles) for the whole subsurface volume of the material to be altered in the manner discussed. At this point, further plastic deformation is more difficult due to all the volume exposed to load is work hardened which marks the start of the steady-state stage [29, 72]. Experimental results from Arakere et al. [96] suggests the hardness of the subsurface volume of M50-Nil exposed to RCF gradually increases to reach a maximum saturation level between 13.5×10^6 - 240×10^6 stress cycles, indicating the start of the steady-state response.

Vegter et al. [98] reported the shakedown to occur before 10^4 - 10^5 revolutions which corresponds to around 0.1 - 1% of the total bearing life. The shakedown limit for a given loading history is considered as the greatest stress that can be applied without inducing

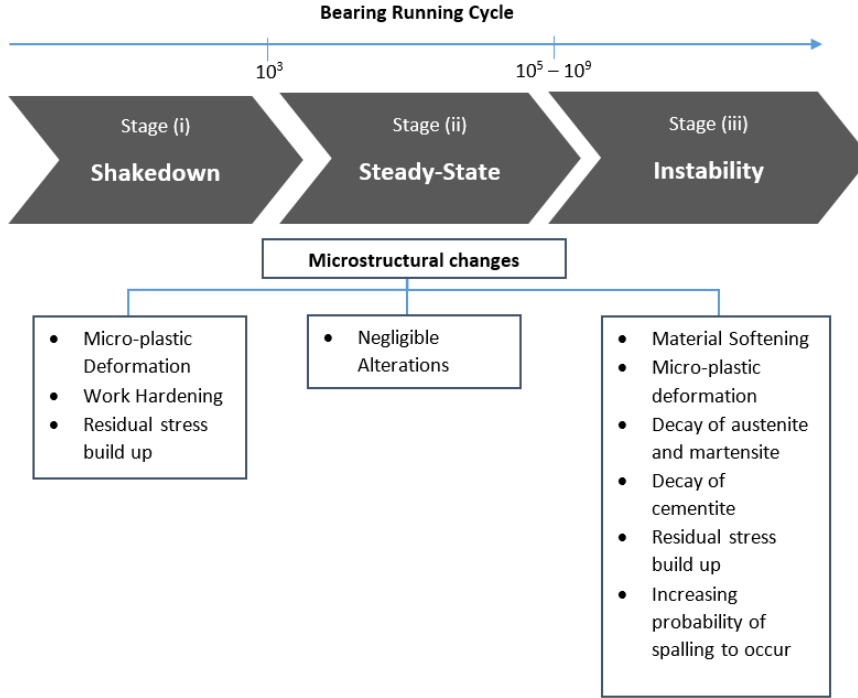


Figure 8: A representations of the material response at 0.2 mm below the inner raceway in the three different stages of bearing operation experiencing rolling contact fatigue. Adapted from [97].

plastic deformation [11] which is within the yield criterion. However, micro-plastic deformation which depends on local stress raisers exists even when the applied stresses lie within the yield envelope [99]. The steady-state condition will only be reached if the stress induced from loading is below the shakedown limit. If greater than the shakedown limit, the material will continue to plastically deform in a macroscopic manner [29]. Typical shakedown limits can be seen in Table 3.

Table 3: Shakedown limits for 52100 steel depending on the operating pressure and contact area [100]

	Shakedown Pressure (MPa)	Operating Pressure (MPa)
Point Contact	2400	2000-2800
Line Contact	2040	1400-1800

Two approaches can be utilized to obtain the shakedown limit. A numerical model consisting of finite element mesh where the internal stresses and deformation of each element of the mesh due to an applied cyclic load can be determined [101]. After numerous cycles, once the residual stress and deformation no longer vary from one cycle to the other, this indicates elastic shakedown has occurred. This behaviour can be numerically modelled to determine the shakedown limit for the contact [102, 103]. Nonetheless, this analysis requires significant computational effort since it involves elastic-plastic finite element calculations and

approaches steady state using an iterative technique. Given that this numerical method is non-analytic, it would be challenging to predict how variations in the material or the geometry would impact the shakedown. Indeed, there have been recent computational methods with enhanced efficiencies in approaching the steady state [104–108] but they remain non-analytic solutions.

Another approach utilizes shakedown theorems [109–111] of the plasticity theory to obtain a lower and upper boundaries which acts as a range for the shakedown limit. Even though this does not provide an exact value for the shakedown limit, it still provides a range of values where the limit lies within where the accuracy of the method is determined by how small the range of value i.e the difference between the upper and lower limit of the approximation to obtain a more precise estimation of the shakedown limit [101]. Given that this solution is analytic, the variables such as geometry become essential for the calculation which is beneficial. This method approaches the steady state directly rather than through iterations and henceforth, requires much less computational effort.

An overview of how the cyclic loading effects the material response at its early life is shown in Figure 9. If the load is beneath the material yield, this results in a perfectly elastic response of the material which can be observed in Figure 9a. Ideally, operating conditions would aim to keep the stress values within this limit but practically, commercial and industrial applications demand higher loadings. If the load exceeds the elastic limit, the material will experience shakedown. However, it is shown that there are two different types of shakedown which have been referred to by Williams [101] as 'elastic shakedown' and 'plastic shakedown' depending on the applied load as shown in Figure 9.

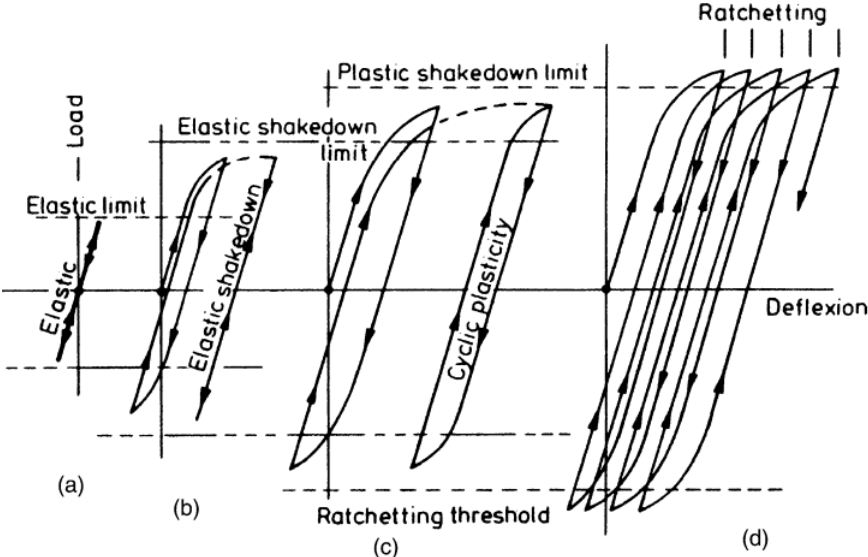


Figure 9: Various forms of structural response to cyclic loading showing a) perfectly elastic, b) elastic shakedown, c) plastic-shakedown and d) Ratchetting [101].

Under the condition where the load exceeds the structural elastic limit but under the

elastic shakedown limit, plastic flow will be observed within the material in at least the initial application of the load [101]. Due to the generation of residual stress during operation, the steady cyclic state may remain completely elastic after the initial plastic flow as shown in Figure 9b. This regime is referred to as elastic shakedown. At higher loads beyond the elastic shakedown limit, each loading cycle would consequently lead to a combination of both elastic and plastic deformation. However, if the load remains below the plastic shakedown limit as shown in Figure 9, the plastic deformation will disappear after a limited number of cycles [112]. Plastic shakedown is achieved when the response becomes one of cyclic plasticity where the steady-state strain cycle is closed (Figure 9c). Cyclic plasticity progresses when the load exceeds the plastic shakedown limit resulting in unidirectional strain accumulation [101, 113] and continuing plastic deformation. A contact with maximum contact pressure exceeding the plastic shakedown limit would cause incremental collapse or ratchetting (Figure 9d), making it liable to premature failure via wear or fracture due to the plastic deformation accompanying each load cycle. However, if the operating condition lies within the shakedown limit (elastic or plastic), a relatively prolonged bearing life would be expected.

3.2.2. Steady-State Response and Instability

During the second stage, it is believed fatigue damage to be non-existent since the cyclic response is elastic [114]. The microstructure can be assumed to be fairly similar to that developed in the shakedown stage. To improve the bearing life, maintaining the elastic response in this stage is critical. The steady-state stage is heavily dependant on multiple parameters such as applied stress, operating temperature, number of cycles and material characteristics such as heat treatments, alloy composition and inclusions, residual stress levels and hardening response during operation [35, 114]. Voskamp et al. [35] highlighted that the stability of finely dispersed carbides in tempered martensite can potentially extend the elastic response of the material significantly. Observations have suggested that the length of this stage can span to 10^9 cycles without signs of further plastic deformation [29]. The subsurface volume in this stage that is plastically deformed upon increasing revolutions is considered to be negligible up to a certain point across the bearing life [94, 95].

Under the first stage of operation (shakedown), transformed retained austenite and residual stress development results in cyclic hardening, preventing further deformation and microstructural alterations for a number of cycles (steady-state response). However, the inevitable development of microstructural alterations in the subsurface such as carbon diffusion within the microstructure, where local temperature peaks can activate underpinning and potential slip systems [6] causes cyclic softening. For elastic shakedown, the steady-state response is considered to be perfectly elastic. However, for plastic shakedown, the steady-state becomes a closed elastic-plastic loop as shown in Figure 9c globally but locally plastic deformation may accumulate near subsurface stress points such as inclusions or carbides which can contribute to microstructural alterations and cyclic softening where plastic strain start emerging globally. This results in a decline in yield stress [29], causing the cyclic strain amplitude to increase which marks the start of the 'instability' regime. During this stage,

microstructural deterioration accelerates which ultimately results in crack formation and spalling [100]. To extend the bearing life over a longer number of loading cycles, prolonging the steady-state response of the material is vital to delay microstructural alterations in the subsurface from initiating material softening and causing the material to enter the instability stage where the rate of material degradation is amplified [6, 97].

3.3. Friction

The position of the maximum shear stress typically depends on the contact geometry and friction coefficient. Under rolling conditions with sufficient lubrication, the maximum shear stress is typically found in the material subsurface as mentioned previously. However, under sliding conditions the maximum shear stress may be located at the surface due to the higher friction as shown in Figure 10a. Practically all bearings experience a certain degree of slip between contact elements during operation. The existence of a degree of slippage during operation consequently raises the maximum subsurface shear stress towards the surface of the contact element subject to friction [18]. If the friction level is significantly high, the maximum shear stress may be positioned at the surface as shown in Figure 10b. Nonetheless, the location of the maximum shear stress may yet exist below the surface in sliding conditions if the corresponding friction coefficient is kept low as shown in Figure 10c. At high friction, the probability of failure due to surface initiated cracks increases while subsurface initiated failure becomes more likely at lower friction given by the maximum shear stress position.

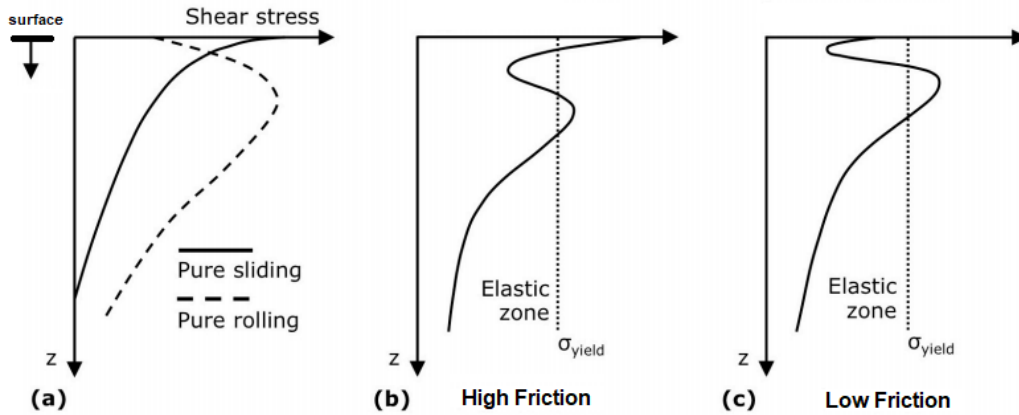


Figure 10: a) Shear stress distribution as a function of depth z for pure sliding and rolling conditions. b) stress distribution at high friction condition. c) stress distribution at low friction condition. z axis represents depth below the surface and x axis represents the stress magnitude [18].

The slide-roll ratio is given by the ratio between the sliding velocity and rolling velocity. While for lubricated tests in a two-roller machine where the axes of rotation are fixed, the rolling velocity is the same as the lubricant entrainment velocity which is the speed at which the lubricant is drawn towards the contact which determines the lubricant thickness. The sliding speed is the speed of the two surfaces relative to each other which controls heat and

friction generation. The slide-roll ratio of rolling elements is believed to determine the performance and can impact the resistance to micro-pitting and RCF. Increasing the slide-roll ratio tends to reduce the resistance of the material to RCF [115] and consequently increases surface friction. However, it has been observed that service life decreases with higher slide-roll ratio until a different surface failure mechanism (wear) becomes pre-dominant at a slide-roll ratio of 40-80% [116]. Wear and RCF are competing failure mechanisms which influence on another.

Experimental results from Sadeghi et al. [117] agree with previous findings where the location of the maximum shear stress moves closer to the surface as the slide-roll ratio increases. Therefore, if microstructural alterations are dependant on the shear stresses in the microstructure, this would suggest that these alterations will be positioned closer to the surface compared to pure rolling conditions. It was also shown in the same investigation that for high load applications, the effect of sliding is more noticeable where only a small amount of slide is required to draw the maximum shear stress location to the surface [117] which could be due to the greater friction generated at higher loads. Sliding in bearings can generate excess heat due to friction at the surface which can consequently reduce the oil viscosity and hence oil film thickness [118]. However, Nuruzzaman et al. [119] have argued that the film thickness reduction with increasing slide-roll ratio is negligible where the operating load and speed are more deterministic factors affecting the film thickness. Nonetheless, it has been shown that having a slide-roll combination shortens the fatigue life compared to pure rolling operations due to the friction on the surface [84].

4. Bearing Life Models

Many models have been established to estimate the bearing life under RCF and the standard bearing life model used in industry has enhanced over the past decades to accommodate the growing demand and better represent modern bearings. The basis of standardized bearing life calculations used today stems from the Lundberg and Palmgren model [62]. Their work proposed that a simultaneous occurrence of a weak point (randomly distributed within material microstructure) in the subsurface at a depth corresponding to the maximum orthogonal shear stress would result in the initiation of subsurface cracks. The probability of survival of the stressed volume due to subsurface initiated fatigue is obtained through the Weibull statistical strength theory, which is applied to the stressed volume in a pure Hertzian contact [6]. The failure in this model is crack initiation dominant. The formula developed through this model which is the basis of bearing life standards in industry today (ISO 281) [120] is shown in equation 1.

$$L_{10} = \left(\frac{C}{P} \right)^p \quad (1)$$

Where L_{10} is the life for 10% probability of failure given in millions of cycles, C is the bearing basic dynamic load rating which is the theoretical load a bearing can carry for a life of one million inner ring revolutions with 90% probability of survival (N). P is the

operating load on the bearing (N). The index p is the load-life exponent which is 3 for ball bearings with an elliptical contact area, $\frac{10}{3}$ for roller bearings with a line contact [6, 121]. Nonetheless, this original model has some limitations even though it is widely accepted. This model neglects the possibility of surface initiated failure and the presence of a lubricating film on the surface. The model also assumes the load to be normal to the contact area and hence ignores any surface shear traction, which could theoretically alter the location of the maximum orthogonal shear stress. Another limitation of the feature is its assumption of a perfectly smooth contact surface which would ignore possible stress concentration points such as asperity-asperity contacts. These limitations however are addressed through the modified bearing life model.

This bearing life standard was modified multiple times to accommodate the performance of modern rolling bearings. With the latest being ISO 281:2007, an integrated life adjustment factor a_{iso} is added as shown in equation 2 which includes four independent factors: contamination, load, lubrication and fatigue stress limit of the bearing material as shown in equation 3.

$$L_{10} = a_{iso} \left(\frac{C}{P} \right)^p \quad (2)$$

$$a_{iso} = f \left(\frac{e_c C_u}{P}, \kappa \right) \quad (3)$$

The factor e_c is used to quantify contamination, which is the ratio of the maximum internal stress in a clean contact to the stress in a contaminated contact. The higher the value of e_c , the cleaner the condition. This is due to contaminant particles embedded in the raceway which may cause indentations in the raceway and result in additional stress in the underlying material during operation as shown in Figure 11a. The e_c value essentially considers the level of solid particle contamination arising from the lubricant influence. This value is determined using ISO/TS 16281 through detailed graphs and tables based on pitch and general conditions [121]. The fatigue stress limit C_u is the load where the fatigue limit of the bearing is just reached which is calculated for ideal conditions using formulas from ISO 281 Annex B. The lubricant regime is determined through the κ value which is the ratio of the actual kinematic viscosity of the lubricant at operating temperature to a reference viscosity of the lubricant which can be determined graphically based on the bearing diameter and operating speed as shown in Figure 11b. The reference viscosity is that which would produce a lubricant film thickness equal to the composite surface roughness of the raceway and rolling element, i.e gives a λ value of one. The final factor P is the dynamic equivalent load. Figure 11c shows how the life adjustment factor is determined graphically relative to the mentioned parameters. As opposed to the original Lundberg and Palmgren model, the modified life calculation considers lubrication (lubricant film, roughness) and surface initiated failures due to the κ dependence of a_{iso} while friction is considered through simplified factors within the load rating calculations.

Early work on predicting the life of rolling element bearings was based on empirical

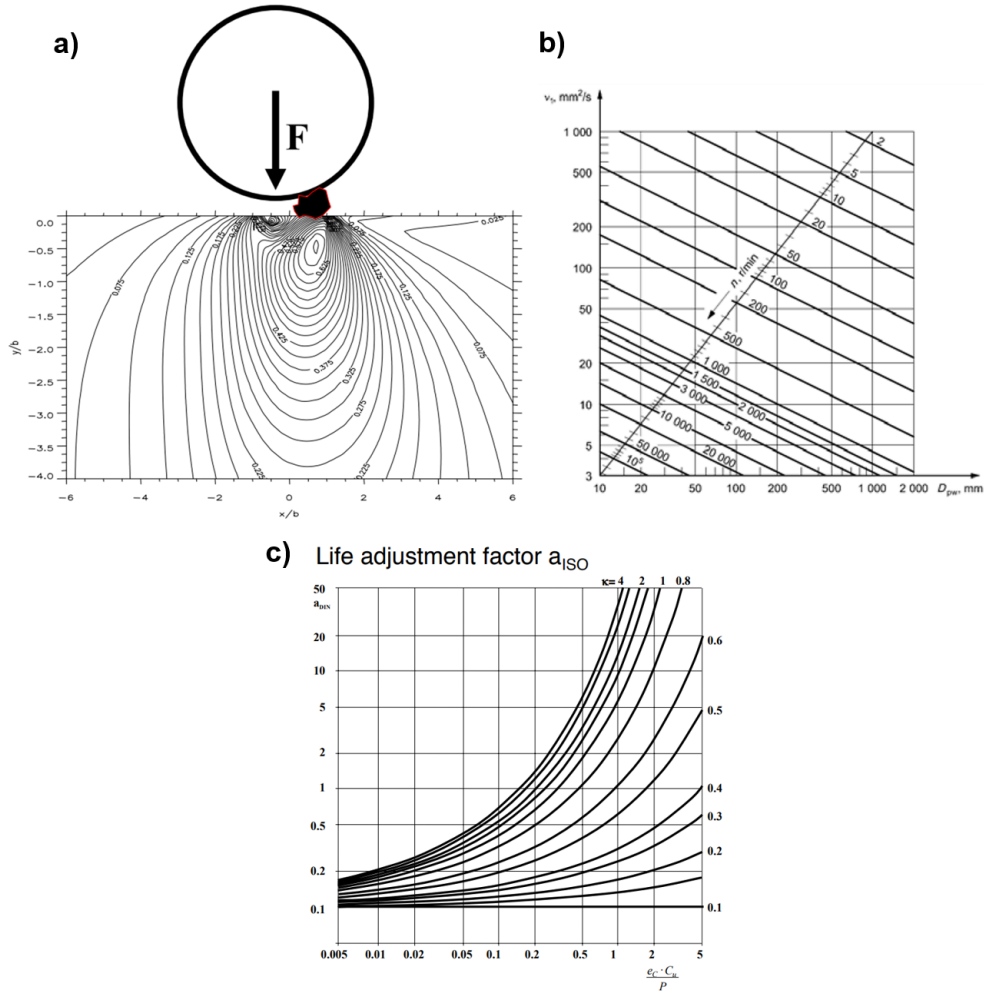


Figure 11: a) diagram showing additional stress induced in the subsurface due to contaminant particle caught between rolling element and raceway. b) Reference viscosity (y axis) can be determined through the pitch diameter (x axis) and bearing speed (diagonal lines) to calculate κ . c) The life adjustment factor a_{ISO} on the y axis and the x axis being the ratio $\frac{\sigma_c \cdot C_u}{P}$ and the curved lines representing different κ values [121].

results such as the Lundberg and Palmgren [62] due to the difficulty of relating RCF to classical fatigue. However, deterministic research models were later developed to take into consideration the mechanics of the failure mechanism based on either crack initiation [122–125] or crack propagation [126, 127] while some models account for both [128, 129]. The majority of proposed deterministic models in literature are based on the assumption of material homogeneity within the contact region to predict the RCF life [130]. Such assumptions negate the influence of certain inhomogeneities in the microstructure such as inclusions from the manufacturing process as discussed in section 2.2 which can act as nucleation sites for fatigue cracks. Inclusions also contribute to localised stress concentration within the microstructure due to the difference in properties compared to the surrounding matrix [131].

Nonetheless, other inhomogeneities in the microstructure such as carbides and secondary phases may also demonstrate similar consequences as inclusions. It has been reported that the stability of the finely dispersed carbides in steel bearings is an important factor for extending the steady-state response of the material during RCF [35]. Effects from these inhomogeneities are often only considered implicitly through material strength values and other material related parameters or factors in classical models. Nonetheless, basic models do not consider such inhomogeneities on life as these factors are typically kept constant. In addition to the mentioned inhomogeneities existing in steel bearing prior to operation, microstructural alterations have been known to develop in rolling element bearings during operation which assist in the fatigue process such as butterflies, white etching areas, dark etching regions and white etching bands which are discussed in more detail later in this paper. Paulson et al. [132] presented a model to compare the effects of grain topology and crystal orientation on a sample microstructure to previous isotropic models. It was shown that mismatch between crystal orientations created stress concentrations. The stress concentrations at grain boundaries resulted in higher critical shear stress than values predicted by typical isotropic Hertzian theory. A Voronoi finite element study by Vijay et al. [133] shows not only that the critical stresses within the material is greater than that of Hertzian theory, it is also demonstrated that there is an obvious scatter in the position of the critical stress within the domain. However, the average depth of the critical stress are comparable to results from Hertzian theory as demonstrated in Figure 12.

Limited number of models associate the micro-mechanical behaviour of the material to RCF which makes it difficult to analyse using classical and commercially available finite element methods. Given that fatigue cracks initially appear as micro-cracks at weak material points, the mechanism of crack initiation is influenced by the microscopic characteristics of the material. Hence, macroscopic based models would tend to negate such microscopic details. The RCF bearing lives show scatter due to the spatial distribution of material strength and scattering of inclusions. While the Weibull distribution represents bearing lives observed experimentally, the crack initiation mechanism is basically determining the weakest point in the material locally which is typically neglected in analytical research models in literature [6]. Empirical life models are practical life prediction tools, however such models are based on the Weibull strength theory which is not intended to be used to account for microstructural effects on life. However, the practical failure of bearings is assumed to be influenced by inhomogeneity within the microstructure such as inclusions, carbides, grain boundaries and secondary phases [5]. Such inhomogeneity in the microstructure could be one aspect why bearings under identical operating conditions and material batch demonstrate a range of scatter in terms of fatigue life.

There have been models proposed considering the inhomogeneity of the microstructure consisting of randomly sized and shaped grains [132–135]. Investigating both the topological randomness from geometric variability within the microstructure and the non-uniform distribution of material properties throughout the microstructure using a Voronoi finite element method has shown the L_{10} life and Weibull slope to be lower compared to a uniform material. This is believed to be due to the introduction of internal flaws in the material

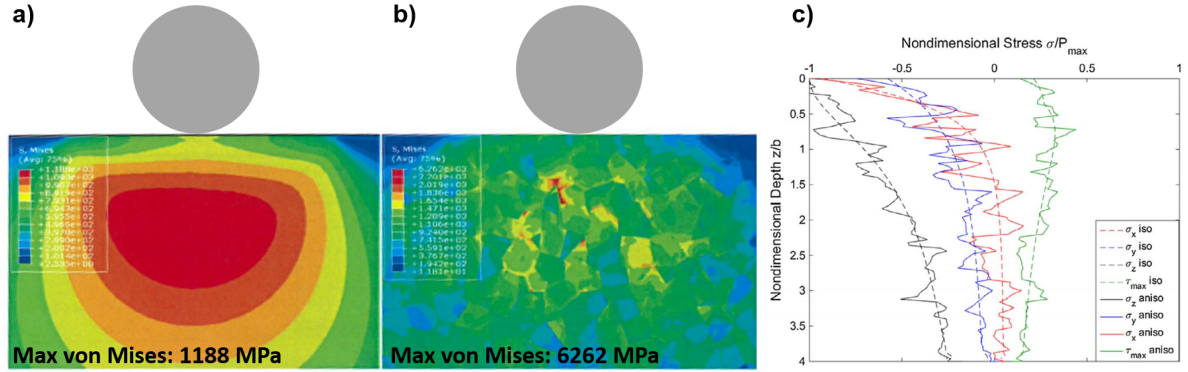


Figure 12: a) Von Mises stress contours with a mesh refinement of 150 elements/Voronoi grain subject to a maximum Hertzian pressure of 2000 MPa for a) isotropic domain b) anisotropic domain. c) centreline stress variation with depth for isotropic and anisotropic material [133].

and the non-uniform distribution of the elastic modulus [134]. However, it is interesting to note that the Weibull slope reaches a limit beyond a specified concentration of flaw which is similar to that predicted by the Lundberg-Palmgren theory.

Investigations by Zhang et al. [136] demonstrate the combined impact of surface roughness and material anisotropy on bearing life. Surface roughness contributes to an elevated stress distribution in the subsurface which reaches its maximum at a shallower depth compared to smoother surfaces [137]. The surface would also consist of higher stress appearing as multiple isolated contact islands arising from asperity-asperity contact [136]. The presence of shallow inclusions and other material inhomogeneities have a more negative impact on fatigue life compared to that deeper in the microstructure [138]. Hence, a rough surface would intensify the negative impact of such inhomogeneities near the surface. However, an interesting conclusion from the model shows the detrimental impact of inclusions on bearing life is less influential in an anisotropic model compared to isotropic models. Therefore, it is suggested that other material parameters such as grain size, carbides and other local anisotropy variation could be as detrimental as inclusions. Nonetheless, even though carbides are considered as stress points in the microstructure, spheroidized carbides can strengthen the material and reduces the possibility of surface initiated failure which improves the fatigue life [139]. Nonetheless, they may yet contribute to subsurface failure.

Modelling of different surface roughness textures have shown longitudinal textures (longer autocorrelation length parallel to rolling direction compared to that perpendicular) results in the highest pressure distribution in the subsurface which is believed to be due to the resulting reduction in the hydrodynamic effect [140, 141]. It is also concluded that such a texture increases the generated fatigue cracks in the microstructure which reduces the fatigue life compared to a transverse texture (longer autocorrelation length perpendicular to rolling direction compared to that parallel). Guan et al. [140] show that as the surface roughness mean square root increases, a transition occurs from subsurface initiated cracks to surface initiated cracks while significantly large mean square root values results in both

surface and subsurface initiated cracks which coalesce leading to failure.

A recent study by Allison et al. [142] compared various critical factors in modelling of the fatigue life of bearings to determine the most detrimental. In general the presence of inclusions is suggested to be deleterious to the fatigue life where the existence of porosity and debonded inclusions dominate the bearing fatigue life. When these are excluded, it is reported the variation in elastic modulus and yield strength of the material due to temperature variation is the most influential factor in fatigue life. The main limitation in this model is the assumption of a homogeneous isotropic steel matrix and defines the fatigue limit as the stress below which no microplasticity occurs and hence is a conservative estimate which negates crack initiation or growth. However, given the impact of the varying mechanical properties has on the fatigue life suggests that modelling inhomogeneities in the microstructure such as different phases with varying properties is essential to develop a more precise analysis of the fatigue performance of bearings. Indeed different phases exist in the microstructure prior to bearing operations but the phases in the microstructure varies significantly with stress cycles as well. Hence, the modelling of the development of different phases in the microstructure with their corresponding mechanical properties could improve the accuracy of current models.

5. Microstructural Alterations: Dark Etching Regions

Materials in rolling element bearings are stressed during cyclic loading which typically results in the manifestation of irreversible subsurface microstructural alterations such as Dark Etching Regions (DER) and White Etching Bands (WEBs). It is believed that the microstructural alterations in bearing steels occur at the steady-state response stage or even as early as the plastic shakedown stage of RCF where a steady state plastic flow is present [97, 143, 144]. DERs usually form under moderate to high contact stress in the area of maximum shear stress after a large number of rolling cycles (5-100 million cycles) [2, 13, 87, 88, 90, 97, 145–147]. In accordance with the contact geometry and load-dependent depth of maximum shear stresses, the depth of the region has been found to vary from approximately 0.1-0.65 mm beneath the contact surface. Nonetheless, it has been well established that their size and formation density increases with contact pressure and running time. The lower boundary of the DER expands quicker than the upper boundary close to the surface with running cycles, indicating that not only that DER thickness increases with running time, but also appears at the greater depth beneath the surface [6]. Bush et al. [87] have determined the following conclusions about DERs after analysing their microstructure which appears to be well accepted among researchers:

- The altered structure resembles tempered martensite.
- DERs only develop when the operating pressure exceeds a certain threshold level believed to be around 2.5 GPa.
- The transformation initiates in localised areas within the region of maximum orthogonal shear stress.

- The nucleation rate of altered regions is proportional to the contact stress.

The 2.5 GPa threshold value agrees with multiple reports believed to be due to the shake-down limit for 52100 steel point contact which is around 2.4 GPa [4, 10, 87, 143, 148]. The sequence of microstructural alterations in bearing steel during operation is seen through changes in etching response of the microstructure which typically starts with DER. The name given is due to the dark appearance of this region when observed under light optical microscopy. The dark contrast consists of scattered deformed and elongated patches due to stronger etchant concentration [4]. Optical micrographs of DER in martensitic 100Cr6 steel bearing inner ring can be seen in Figure 13 where DER is parallel to the surface in the circumferential section but appear as a sickle shape in the axial section corresponding to the size of Hertzian contact area.

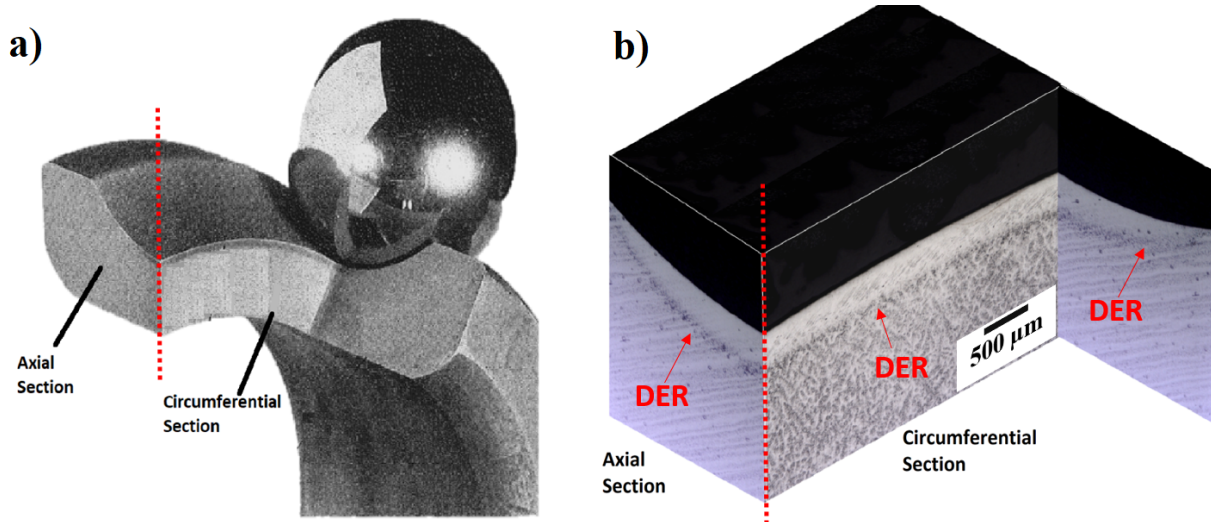


Figure 13: a) Axial and circumferential section of bearing inner ring b) Optical images of DER in axial and circumferential section corresponding to cuts shown in (a). Adapted from [97].

DER is a consequence of martensite decay under RCF due to stress-induced phase transformations [2, 3, 6, 8, 13, 24, 86, 90, 143, 147, 149–152]. The altered microstructure consists of a ferritic phase with inhomogeneously distributed excess carbon corresponding to the parent martensite [13, 143, 153, 154]. DERs are believed to be a mixture of this phase coupled with residual martensite. The boundary between the deformed DER microstructure and the unaltered matrix is gradual with no clear boundary [13]. However, it does appear that the lower boundary of the DER is more gradual while the upper limit near the surface is more definitive but no paper has proposed a theory for this variance. Throughout the DER, the contrast due to etching varies as shown in Figure 13b and Figure 14a and b. In other words, the DER is not uniformly dark but consists of dark and lighter patches which vary across the DER [4] causing the inhomogeneous distribution of dark areas shown. The structure of DER consists of long disc-like cementite particles embedded in a cell structure consisting

of deformed and misoriented martensite plates where fine hardening carbides dissolve during cyclic deformation and re-precipitate as ϵ -carbides and henceforth, over-tempering the microstructure [90, 145, 155].

Looking at Figure 14a and 14b, the inhomogeneous distribution of dark patches is coupled with small white primary spheroidised carbides. It is apparent that primary spheroidised carbides appear smooth and flat protrusions in scanning electron imaging (SEI) taken from a scanning electron microscope (SEM) but dark and spherical in back-scatter electron (BSE) images. Dark patches observed in Figure 14a and Figure 14b are suggested to be linked with the clusters of elongated dark features appearing as deep grooves in Figure 14c which appear as elongated grains (red arrows) in Figure 14d. Coupled with the elongated grains, small globular grains are observed in both SEM images as a result of the stronger etchant attack at these sites (yellow circles) [13]. Given that these globular and elongated grains are attacked more by nital suggests the grains are ferritic, giving the dark appearance observed in the region.

5.1. Causes of DER

There has been multiple debates and theories to determine the cause of DER formation in bearings under RCF. Multiple reports have linked the formation of DERs to tempering [13, 145, 156] and carbon migration from martensite to heavily dislocated regions [97]. During tempering, the transport of carbon atoms in the matrix causes the carbide particle size

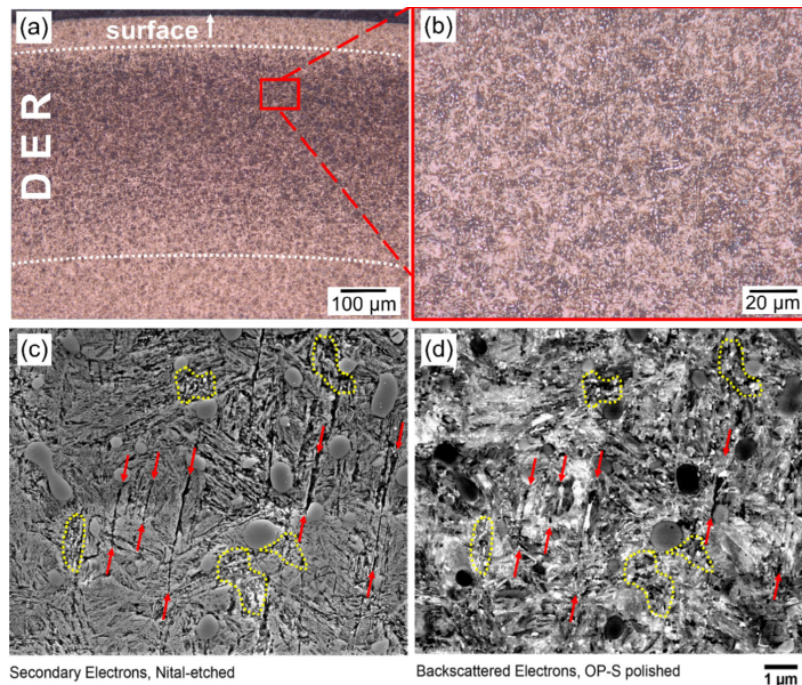


Figure 14: a),(b) Light optical micrographs of nital etched DERs c) same region observed with secondary electron SEM d) backscattered electron SEM (OP-S polished). Elongated and globular grains are observed with red and yellow respectively [13].

to expand which consequently increases the matrix/carbide interfacial area [4]. Kang et al. [157] suggested dislocation assisted tempering would generate dislocations glide during deformation, where carbon near the dislocations can be transported from the matrix and towards the carbides, making the matrix carbon depleted. Nital and Picral etchants are known to reveal carbide/ferrite interfaces. Therefore, it is believed that etching strength can be directly related to carbide size and hence carbon migration [158].

DER formation has been suggested to be caused by heat generated through elastic hysteresis due to cyclic stressing [156]. However, it is argued by Kang [4] that plastic flow is most likely behind DER formation as it is never observed at the surface and its threshold stress exists (shakedown limit ≈ 2.5 GPa). This suggests the essential role of stress and plastic deformation in DER formation. There have been attempts to relate the formation of DERs to the highest stress in the material subsurface but it is yet not clear which stress component is the biggest contributor to DER formation [3]. Warhadpande et al. [2] suggested via optical microscopy tests that the depth of DERs closely corresponds to the depth of the maximum shear stress. However, there still remains some disagreements towards which phenomenon contributes to the DER formation. While some authors have linked it to the maximum shear stress [2, 86, 153, 157, 159], others believe it is due to the orthogonal shear stress [87, 143, 150] or von Mises stress [83, 89, 90]. Nonetheless, given how similar in depth these stress criterions are, it is difficult to isolate the responsible stress.

Smelova et al. [13] reported that the elongated and globular grains in DERs are formed via dynamic recrystallization due to the low misorientation of the grains in the DER compared to the unaltered matrix and the existence of high angle grain boundaries between grains and gamma-fibre texture. However, no chemical element re-distribution occurs throughout the recrystallization process of the dark etching regions. Primary spheroidised particles remain unaltered both crystallographically and chemically whereas the retained austenite is transformed in the altered region to martensite [13, 160]. Electron diffraction patterns indicate that the original plates of martensite are deformed as each plate consists of significant misorientations in the form of a cell structure [159]. It has been reported that slip marks in DERs are orthogonal to each other and at an angle of 45 degrees from the material surface [3]. However, Sugino et al. [145] implied that the newly formed patches were directionally oriented towards undamaged martensite platelets even though other orientations have also been declared [90, 94].

Martensite decay is also accompanied by a decrease in specific volume [97], generating internal tensile stresses opposing the compressive stresses induced by the volume expansion due to the decay of austenite. The existence of residual stress within highly localised plastically-deformed regions in the subsurface results in different deformation textures observed in DERs. Voskamp et al. [94, 97] reported that above 45 °C and maximum contact pressure <3.2 GPa, two textures develop within subsurface at the ferrite matrix each with an associated spall as shown in Figure 15 corresponding to DER while samples without DER formation demonstrated no distinctive textures.

- A $\{100\}\langle 110 \rangle$ texture develops at relatively low temperatures (but greater than 45 °C)

with deformation being the dominant cause. (with $\{hkl\}$ being parallel to the surface and $\langle uvw \rangle$ parallel to the direction of over-rolling). This texture is associated with flat spall bottoms as shown in Figure 15a.

- A $\{111\} \langle 211 \rangle$ texture develops at relatively greater operating temperatures and is believed to be caused mainly by dynamic recrystallization. This texture is associated with irregular spall bottoms shown in Figure 15b.

A recent model has suggested that DER is strain induced, where the excess carbon in solid solution is incorporated to pre-existing carbide precipitates during the martensite-ferrite transition [143, 157]. The growing carbides may also explain the dark appearance of DERs, where nital etchant attacks the carbide/matrix interfaces leading to the dark contrast [161]. Experiments have shown that increasing both the contact pressure and operating temperature can accelerate the formation of DER earlier in the bearing life [159]. The initial heat treatment of the material is also a major factor impacting DER formation [86, 97]. Longer tempering treatment may lead to a reduction in microstrain and a loss of coherency of ϵ -carbides within the matrix which consequently diminish the dark etching response [97]. While other reports have also suggested that DER is not found in steels where the retained austenite is $<2\%$ [97, 162] indicating high levels of austenite which can be controlled through heat treatment, contribute to DER formation and material softening during RCF.

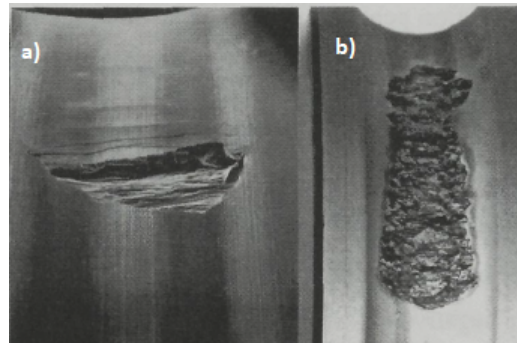


Figure 15: a) flat spall bottoms associated with a $100\langle 110 \rangle$ texture formation. b) irregular spall bottoms associated with a $111\langle 211 \rangle$ texture formation [97].

5.2. DER Features

EBSM data by Smelova et al. [13] suggests the distribution of primary spheroidised carbides remain homogeneously distributed and fairly similar to the unaltered region even though the retained austenite is considerably lower while a high chromium and carbon concentration exist in the primary spheroidised carbides [13]. Carbon and chromium distributions in the DERs have been reported to be similar to that of the typical unaltered 52100 steel microstructure [13, 154]. However, Fu et al. [143] have analysed the DER using atomic probe tomography at a higher resolution and suggested a much lower carbon and chromium

content in the ferrite patches in DER compared to the unaltered martensite matrix due to the carbon migration and the redistribution of Cr in the region of the Cr-enriched globular primary carbides.

Some have suggested that DERs have a slightly higher hardness compared to the unaltered microstructure due to the work hardening of the martensite or the transformation of retained austenite to martensite where an increase of up to 100HV has been reported [8, 9]. While others have suggested a slight reduction in hardness due to a reduction in carbon concentration in the region [4, 8, 13, 87, 90, 150]. From literature, a maximum reduction of 150HV has been recorded in the DER region. Hardness measurements in literature generally lack in detail and have caused some confusion. However, it is possible that softening (decay of martensite) and hardening (decay of austenite) are competing mechanisms during bearing operation and the prevailing mechanism depends on the material properties such as retained austenite content. DER consists of ferritic dark patches believed distributed in the parent martensite matrix with the former gradually consuming the latter with increasing load cycles. However given that the solubility of carbon in ferrite is only 0.02wt% [161], this does lead to uncertainties as to where the excess carbon goes during this martensite-ferrite transformations. Swahn et al. [90] explained the excess carbon resides at dislocations where the cyclic stress promotes the carbon migration. However, this requires a high dislocation density of $10^{17} /m^2$ which is unrealistic and contradicts theories stating hardness reduction in DERs. Kang et al. [157] suggested more recently that during DER formation, the excess carbons migrate to pre-existing carbide precipitates with the assistance gliding dislocations thickening them. Even though this model provided an explanation of softening in DERs, it has failed to clarify the impact of temperature, contact pressure and the rotational speed of the bearing. Fu et al. [163] used a dislocation-assisted carbon migration theory to describe the precipitation of lenticular carbides in bearings exposed to RCF which can also be utilised to describe the thickening of pre-existing carbide precipitates in DERs due to carbon migration.

Swahn et al. [90] have demonstrated how the material subsurface hardness varies along the depth and running cycles which can be demonstrated in Figure 16. It has been found that the hardness of the material increases up to approximately 10^6 cycles which is expected during the early stages of RCF. However, it has been found that increasing the running cycles further reduces the overall material hardness as an outcome of cyclic softening, which is the mechanical removal of excess dislocations [145, 164]. The delayed softening of the microstructure during RCF is caused by the decay of martensite to cell-structured ferrite while the redistribution of carbon cancels out the initial hardening due to cyclic stressing and decomposition of retained austenite during loading [11, 155]. The hardness differs throughout the subsurface and reaches its lowest value in the region known to develop DER and WEBs. Martin et al. [86] and Buchwald et al. [165] did not observe DER in bearings tempered initially to a hardness of 700 HV, supporting the argument that DER formation results from the over-tempering of martensite. Given that bearing steels with lower hardness is achieved by longer tempering time at higher temperature, there may not be a sufficient amount of carbon in the solid solution for dislocation assisted tempering [157].

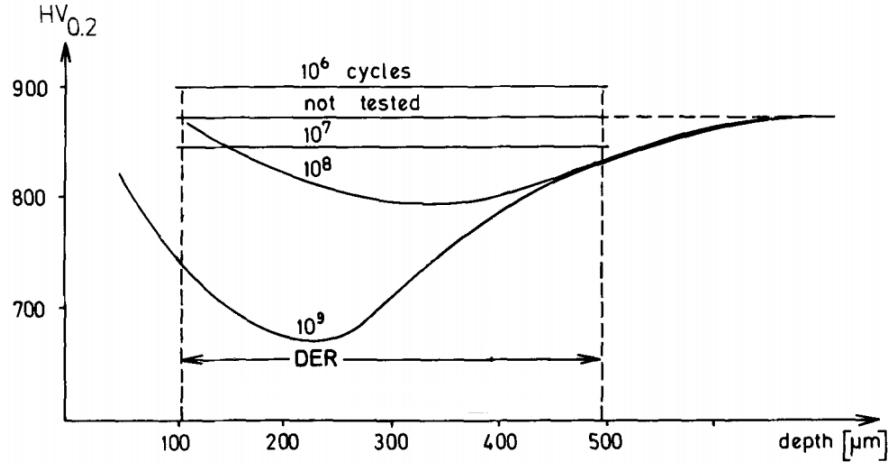


Figure 16: Micro-hardness as a function of depth under the raceway in inner rings before testing and at different running cycles along the bearing life compared to the typical DER location [90].

Fu et al. [143] modelled the extent of martensite decay and the formation of DER by quantifying the area fraction of dark patches (ferrite) termed as DER% to parent martensite. While given that ferrite has low carbon solubility of 0.02wt% [101], the DER% can also be an indication of carbon depletion in the DER area. Figure 17 shows the experimental results from the model showing the hardness reduction in accordance with the gain of dark ferrite patches in the altered region. It can be seen that a higher density of dark patches in the region (ferrite) contribute to the reduction in hardness given that these patches are depleted from carbon. The model also linked the region with the severest martensite decay to the depth with the maximum orthogonal shear stress in accordance to literature [87, 150, 160]. It has been suggested that DER forms when the orthogonal shear stress τ_{yz} exceeded 0.56 GPa [143]. While an investigation of the material response to many parameters suggested that temperature and pressure increase can accelerate martensitic decay to occur much earlier in the bearing life whereas rotational speed was shown to have a negligible effect.

6. Microstructural Alterations: White Etching Matter

Operating at cycles beyond the formation of DERs can result in the formation of white etching matter (WEM) even though certain WEM features can occur much earlier in the bearing life. This degradation of the microstructure may lead to premature failure due to white structure flaking (WSF) which is linked with the formation of WEM as flakes of the material separate from the raceway surface of the bearing [166]. WEM is a generic term applied to all damaged and altered microstructures which demonstrate a light contrast (white/grey) compared to the rest of the microstructure when etched. There are various features of WEM from literature but two main types of WEM can be classified as hard and soft WEM each associated with certain features which occur at different stages in the bearing life as shown in Table 4. However, hardness measurements in literature have caused confusion regarding the properties of different WEMs. Even though literature has com-

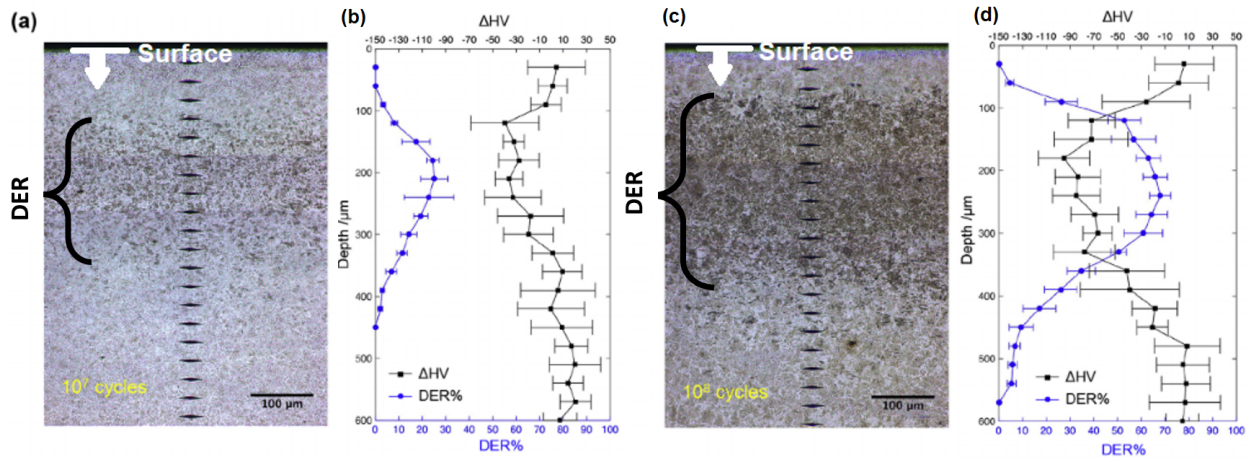


Figure 17: (a) and (c) LOM images of DERs developed in the circumferential sections of bearing inner rings run after 10^7 and 10^8 cycles, respectively. (b) and (d) compares the relationship of DER% and change in HV from the original value prior to testing with respect to the depth measured from the samples shown in (a) and (c) [143].

monly stated white etching bands (WEBs) to be soft relative to the parent microstructure, Smelova et al. [13] stated that WEBs consists of carbon-rich areas which possess higher hardness adjacent to the ferrite bands which are softer than the parent microstructure and therefore the hardness fluctuates throughout the region. However this raises the question on whether these carbon-rich areas are considered part of the WEBs or whether the term WEB just include the soft ferrite band. It is commonly thought that rolling contact fatigue cracks initiated at pores precede the formation of hard white etching matter which is associated with severe deformation [18, 167, 168]. In WEBs, the interstitial is rejected from the bands but dissolved in the butterfly (hard WEM) due to severe deformation. The chemical potential of carbon at defects is lower compared to when it is present in cementite [169]. This might explain the difference between both white etching matters wherein butterflies, the high dislocation density in the white region can trap carbon atoms causing supersaturation which suppresses the precipitation of carbides commonly reported with the soft WEB [11]. Similarities between WEBs and white etching cracks (WECs) have been reported in structure and formation mechanism [151, 168]. Both WEC and WEBs demonstrate globular and elongated grains. Elongated grains in WEB demonstrate a defined direction related to rolling direction whereas those in WEC are locally defined in clusters without a specific direction related to rolling direction. Figure 18 shows an image of reported hard white etching areas corresponding to WEC and butterflies which are commonly reported in bearings.

6.1. White Etching Bands

White etching bands (WEBs) which are commonly referred in literature as soft WEM typically form after DERs develop in the bearing and are positioned within DERs. Nonetheless, DER formation is not a prerequisite for WEB initiation as reported by Lund et al. [150]. It is suggested that WEBs could occur in a soft matrix tempered to 670HV in the absence of DER. WEBs are parallel three-dimensional thin plates parallel to the contact surface in the

Table 4: Comparison between hard and soft white etching matter

	Soft White Etching Matter	Hard White Etching Matter
Hardness	Typically exhibits a hardness reduction of up to 9 HRC (approx 15% reduction) compared to unaltered microstructure [170].	Exhibits a higher hardness (30 - 50%) compared to the unaltered microstructure [166].
Carbon Content	Depleted of carbon [153] due to the diffusion of carbon into the surrounding. Carbon content has been recorded to be as low as 0.06% while surrounding region has 1.3wt%	supersaturated in carbon as a result of carbon dissolution [166] with varying percentages greater than the unaltered matrix.
Structure	Structure can be transgranular [97, 171], composed of fine equiaxed ferrite grains (approx 20 nm) and contains regions with an amorphous structure and microvoids [171] while a high dislocation density is present at the edges of the bands [172].	Consists of a structure of fine material [171] or equiaxed fine nano-crystalline ferrite grains 10-100 nm in diameter and is free of large carbides [86, 90, 172, 173]. There also appears to be an amorphous phase with nano-martensite and austenite [174]. WEA exhibits no preferential crystallographic orientation [173].
Formation Mechanism	Carbon partitions from WEB into adjacent DER to become trapped at dislocations or precipitates as layers of cementite [165]. Dislocation breaks away from carbon atmospheres in the locally plasticized white region, leaving behind ferrite supersaturated with carbon which then partitions into the surrounding region to precipitate as cementite [12, 175].	Typically nucleate from abnormalities in the structure such as cracks in the form of debonded surfaces, voids and microcracks. The repeated impact of the cracked faces against each other causes severe deformation and intense hardening [18].
Cycle Duration	Occurs late in the bearing lifetime $>10^8$ cycles (earlier at higher pressure)	Occurs early in the bearing lifetime $<20\%$ of L_{10} [176]
Associated Phenomena	White Etching Bands (WEB)	White Etching Cracks (WEC), butterflies, white etching layer (common in rail tracks), white etching area (WEA), spheres

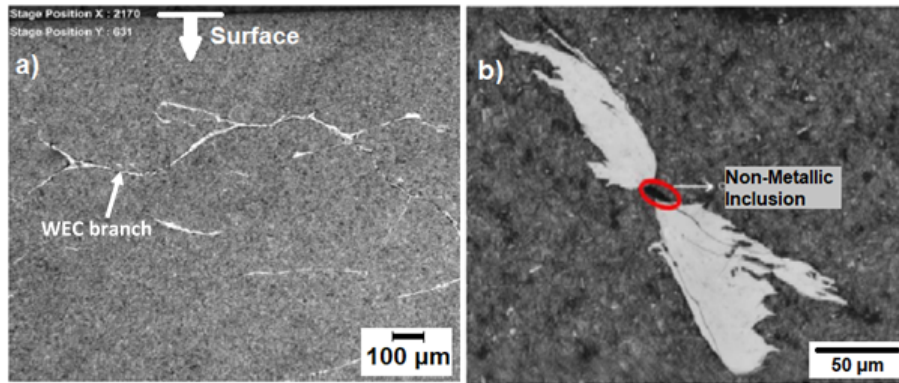


Figure 18: a) WEC crack branch observed in the inner ring of a spherical roller bearing b) Optical image of butterfly and initiating inclusion. Adapted from [177, 178].

axial direction, perpendicular to the over rolling direction and inclined to the contact surface in the circumferential cross-section as shown in Figures 19a and b. They are normally found

at a depth ranging from 200 μm to 760 μm beneath the surface and display a maximum density of around 230 μm to 400 μm from it [3]. Similar to DERs, this region is believed to correspond to the position with the highest shear stress which depends on the contact geometry and loading [6, 13, 179]. A distinctive feature of WEBs is their directionality which will be discussed later in more detail.

WEBs could be categorized into low angle bands (LABs) which are inclined at an angle varying with depth from surface, ranging from 35° closer to the surface to 20° deeper in the microstructure [97] and high angle bands (HABs) which are inclined to 65-85° to the surface in the over rolling direction towards the surface as shown in Figure 19b. The 3D shape of these features is shown in Figure 19c where they intersect each other based on axial and circumferential cuts from experiments. An automated 3D analysis of LABs by Ganti et al. [178] demonstrated a similar plate structure. LABs and HABs are commonly referred to as 30° and 80° bands respectively [11]. These bands are longer than typical austenite grains (50 μm) and their development does not parallel the strengthening of certain crystallographic textures forming due to rolling contact deformation [11, 94]. It is commonly believed that LABs forms initially and succeeded by HABs which form at >500 million load cycles as stated in Table 5. However, Mitamura et al. [159] suggested HABs could potentially form prior to LABs under high contact pressure ranging from 4.6-5.5 GPa. This condition appears to be overload since the generated contact stress of 5.5 GPa would result in significant plastic deformation based on the von Mises yield criterion. Therefore more experiments are needed to examine this observation.

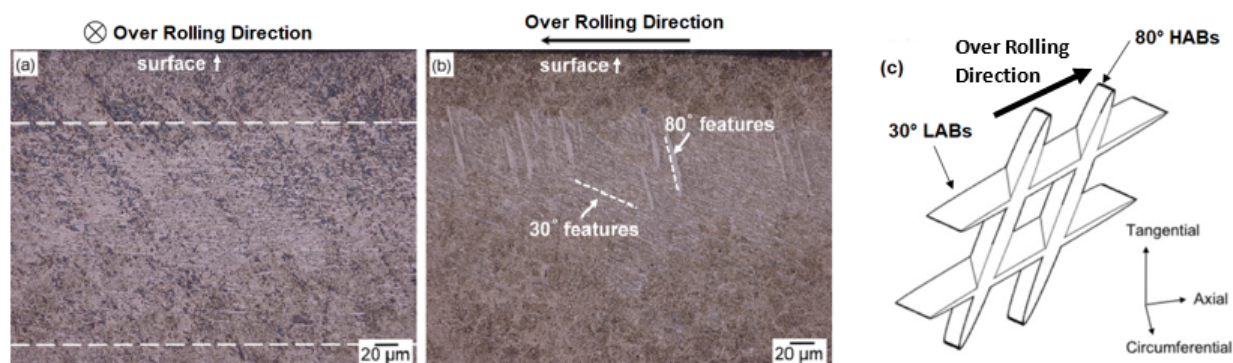


Figure 19: Light optical microscopy images of 30° and 80° features in nital etched solution. (a) axial cross section (b) circumferential cross section [13] (c) schematic showing the overall morphology of white etching bands [4].

Adjoining the white bands formed within the microstructure, high magnification microscopy revealed the existence of another altered structure that does not etch significantly like DER and WEB, it is narrower than WEB where the thickness can reach 1 μm and does not break during tempering [90, 180]. These deformation bands which form in cyclically stressed steel bearings are known as lenticular carbides and can be seen in Figure 20. It has been reported that lenticular carbides are due to the plastic deformation of the primary

spherical carbides [9]. However, Martin et al. [86] proved that the existence of lenticular carbides could not result from the deformation of pre-existing carbides by examining samples which were given a special double-austenitizing heat treatment resulting in extremely fine carbides. Where the lenticular carbides formed were significantly larger than any pre-existing carbide. However, no alternative explanation was offered. Borgese [180] has suggested that lenticular carbides are formed as carbon diffuses over some microns within the bands as a consequence of the stress from the rolling contacts [90]. The diffused carbon precipitates in the form of lenticular carbides where the boundary between WEB and lenticular carbides is suggested to be a plane of weakness [9, 86] which may initiate damage during loading given the variation in mechanical properties between the two structures. Smelova et al. [13] reported these lenticular carbides or carbon-rich areas to have a hardness of 9-11 GPa where the ferrite grains associated with WEBs show a hardness of 4.2-6 GPa. Both WEBs and lenticular carbide have been known to thicken with the number of cycles [165]. The formation mechanism of WEB is believed to be a process of carbon redistribution due to RCF. However, there have been multiple theories for decades on the driving force for carbon migration [163] which is discussed later.

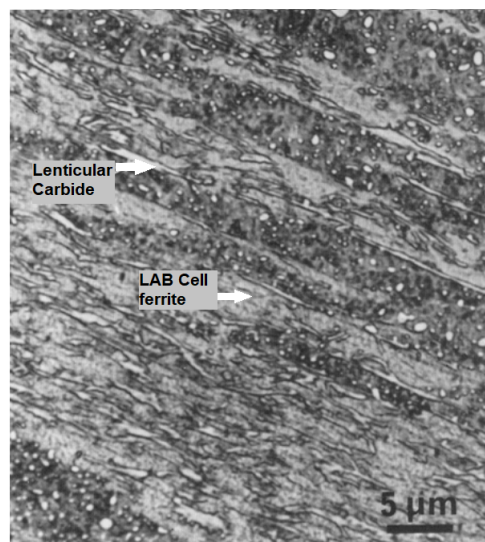


Figure 20: optical microscopy images of cell ferrite LABs (30° bands) and adjacent lenticular carbides. Adapted from [147].

LABs or 30° bands are characterized as disc-shaped regions [178] of ferrite bordered by a narrowly elongated layer of lenticular carbides. The spacing between these bands are typically $0.5\ \mu\text{m}$ to $10\ \mu\text{m}$ [90]. If tempered, LABs cannot be distinguished from tempered martensite, suggesting that LABs is a carbon supersaturated region that forms carbides during tempering [90]. LABs are occasionally seen to extend across the carbide/matrix boundaries. When HABs initiate formation, they merge together with various LABs [3] as shown in Figure 19b. Comparison between DER, LABs and HABs is shown in Table 5.

HABs are considerably thicker and longer compared to LABs as shown in Figure 19b.

Table 5: Comparison of properties of DER, LABs and HABs [4, 11, 13, 90, 149, 153]

	<i>Dark Etching Regions</i>	<i>Low Angle Band (30°)</i>	<i>High Angle Band (80°)</i>
Dimensions	100-650 μm depth	5-30 μm length	\approx 100 μm length
Orientation*	Parallel to raceway	20-35°	65-85°
Band Thickness	-	0.1-0.5 μm	\approx 10 μm
Formation (cycles)	5-100 million	>100 million	>500 million
Carbon Content	Unaltered (1wt% for 100Cr6)	0.2wt%	0.06wt%
Structure	Ferrite + retained martensite	Ferrite with some partially dissolved Carbides	Purely Ferrite

* Circumferential direction of inner ring.

The spacing between HABs can range from 5 μm to 50 μm [13]. Once the highest density of 30° bands is obtained, 80° bands form closer to the surface while penetrating the 30° WEBs. Along the cell structures within the bands, the edges of the cells are also elongated in the 80° direction similar to the bands [10]. It is well known that WEBs consist of cellular structures with dimensions ranging from 50 nm - 100 nm typically without carbides [13, 86], even though remains of partially dissolved carbides have been seen in 30° bands [149].

6.2. Formation of WEB

The precise mechanism leading to the formation of WEBs has been a debate for decades. There are multiple theories which aim to provide an insight towards the formation mechanism of WEBs listed below. It can be seen that the debate mainly consists of whether WEB formation is related to plastic deformation or carbon diffusion.

- Initial theories have suggested tempering to be the main cause of these alterations as a consequence of local temperature rise within the subsurface but no evidence has been shown [8, 86, 150]. However, given that threshold stresses have been found which must be surpassed for these alterations to occur suggests the mechanism is stress-induced [86, 93].
- It has been suggested that WEB formation is caused by an intrusion-extrusion mechanism. Where material is being carried from certain regions of the carbide particles into the neighbouring matrix and the shearing of spherical carbides is emphasized to explain the origin of lenticular carbides. Consequently creating an intrusion in the carbide and an associated carbide extrusion in the matrix, resulting in WEB formation [87].

- Martin et al. [86] suggested white etching regions are located in areas corresponding to the maximum shear stress and WEB formation is a direct result of plastic deformation of the material due to the cyclic, reversed plastic strain in selected volume elements. He also found that WEB did not form when the maximum shear stress < 0.72 GPa which is similar to the elastic limit of 52100 steel (0.62 GPa), suggesting plastic deformation plays a critical role in WEB formation. Martin et al. [86] determined the position of the maximum shear stress to closely correspond to the position of WEBs in the subsurface, whereas the orthogonal shear stress peaked at a lower depth closer to the surface.
- Another theory links the altered microstructure to recrystallization, which includes diffusion [12, 150]. It is possible for carbon to diffuse from the martensite matrix and towards adjacent DERs and consequently becoming trapped at dislocations or to precipitate as layers of lenticular carbides. It was also concluded that spherical carbides dissolve in white etching regions due to compressive strains arising from localized plastic deformation and the association of carbon with enhanced dislocation density [2]. Lenticular carbides grow surrounding WEBs due to carbon migration from the white etching regions [12, 150].
- Becker [175] reported a mechanism in which dislocations break away from their carbon atmosphere within the locally plasticized white region and consequently leaving behind a supersaturation of carbon in the ferrite which partitions later on towards adjacent regions to form lenticular carbides which was also confirmed by Kino et al. [181] via atomic emission spectroscopy.
- Ochi et al. [181] mentioned the WEB to be due to strain localization where carbide dissolution was observed within the white areas. Given the hardness is reduced within the white etching bands, this indicates that carbon depletion by precipitating lenticular carbides to the boundaries of the ferrite bands. Swahn et al. [90] reported the energy increase due to plastic deformation could act as the driving force for carbon flowing towards the surrounding matrix.
- Buchwald et al. [165] suggested the flow of carbon to the surrounding matrix from the ferrite bands stems from stress gradient and carbon concentration where a model was created to determine the rate of thickness increase of lenticular carbides as a reference to WEB formation. Nonetheless, the results suggested an unrealistic temperature rise of 200°C to achieve the reported growth rate. Polonsky et al. [12] postulated a diffusion-based mechanism where carbon injection into the solid solution during dislocation annihilation promotes carbon leaving the ferrite bands.
- Fu et al. [163] created a model to predict the thickening of lenticular carbides with bearing cycles to indicate WEB formation. The theory is based on stress-induced lenticular carbide formation during RCF. The growth of lenticular carbides during operation is achieved by consuming the carbon from adjacent ferrite bands which

consequently forms WEBs. However, this model has failed to differentiate between LABs and HABs.

Focusing on the rearrangement of dislocations in the ferrite bands could help develop a better understanding of the redistribution of carbon. Given that ferrite arises due to plastic deformation, it may contain significant dislocation density. The strain field surrounding dislocations can attract carbon atoms [182] which consequently gives a strong interaction between the two. Bearing operation results in significant amounts of dislocation gliding, which can henceforth impact the carbon migration within the microstructure and explain the growth of both WEB and lenticular carbides with the number of cycles [163]. It has been debated for decades whether the formation mechanism is due to carbon diffusion or plastic deformation. However, it is likely that the formation mechanism is combination of both components where carbon diffusion is induced by plastic deformation. Recent research has investigated strain induced carbon diffusion for the formation of DER and WEBs [143, 157, 163]. A recent theory has linked the formation of DER, WEB and hard WEAs such as WEC and butterflies to be due to strain-induced carbon redistribution [151, 168]. Hard WEA forms early in the bearing life followed by DER and WEBs. This theory suggests the carbon migration distance as a major factor impacting the formation of these alterations which is proportional to the number of cycles in which these alterations form. For WEA, carbon migration distance is tens of nanometers which is the average radius of dislocation cells. While for DERs, the carbon in solid solution travels hundreds of nanometres to reach pre-existing carbides. For WEBs, the carbon travels across the ferrite bands which are several micrometres wide [151]. The increase in carbon migration distance agrees with the sequence in which these alterations occur within the bearing life.

6.3. Orientation of WEB

From current theories and debates regarding the formation mechanism, it could be concluded that carbon diffusion or redistribution has a vital role in the development of WEB during RCF. However, there yet remains a universally accepted explanation on the reason behind the distinctive orientations of 30° and 80° bands.

Among the first theories explaining the orientation of WEBs was reported by Lyman [183] where it was proposed that the WEBs orientation was determined by the combination of shear stress acting along the WEBs and the maximum normal strain 45° to the bands. However, this theory was only considered for 30° bands. Zwirlein et al. [10] linked the orientation of 30° and 80° bands to plastic deformation where the orientation is determined by the state of stress at the location of the maximum von Mises stress. The proposed mechanism for 30° bands is that they grow perpendicular to the direction of the highest principle stress (relative tensile) while the 80° bands develop in the direction of the maximum shear stress. However, this theory has shown to be inconsistent with experimental observations when applied to bainitic steel bearings [3]. Zwirlein et al. [10] have also attempted to link the band orientation to the surface friction but this theory has been heavily disputed [2].

Johnson [179] proposed that the orientation of WEBs is determined by the direction and location of the maximum shear stress. It was assumed that when 30° bands are formed, the

compressive residual stresses are moderate where the ratio of the compressive residual stress to the applied Hertzian pressure is 0.06-0.08. However, this ratio increases to 0.2-0.25 later in the bearing life which leads to the formation of 80° bands. Nonetheless, Voskamp et al. [184] reported that enhanced amount of compressive residual stresses can develop early during the bearing life operation prior to 30° band formation, indicating potential deficiencies in Johnson’s theory.

Many of the theories currently suggested appear flawed and highly debatable and inconsistent. Polonsky et al. [12] provided some paradoxes listed below which must be answered to truly obtain a unified theory for the development of WEBs and their orientation.

1. The 30° and 80° angles observed with WEBs correspond to neither the direction of the maximum orthogonal shear stress (0 or 90° to the bearing raceway) nor the direction of the maximum shear stress (45° to the principle stresses).
2. Contact stress is symmetrical about the centreline of contact in the absence of surface traction. Therefore, if one particular direction is favoured for the formation of WEBs, another direction should exist for the growth of these bands via symmetry. Nonetheless, WEBs are never observed in a symmetric direction unless the rolling contact direction is reversed.
3. WEBs are not observed in conjugate directions. Since the stress tensor is symmetric, a certain shear stress acting at a certain direction must also be acting in the conjugate direction. Therefore for a shear stress linked to WEB formation, the conjugate direction should be equally as favourable for WEB formation. However, this conjugate direction is practically never favoured for the growth of WEBs throughout the microstructural alterations encountered by the steel [93]. The directions which should be favourable for LAB and HABs in the steel are demonstrated in Figure 21.

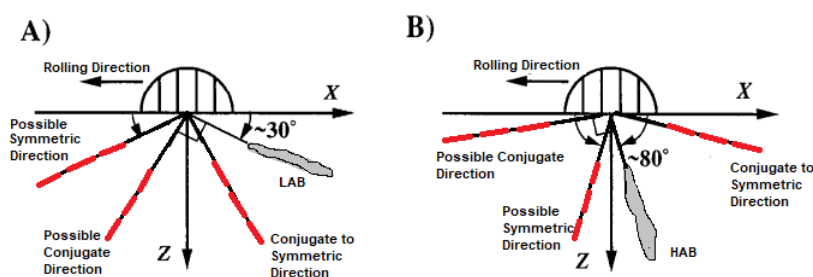


Figure 21: Potential orientations for A) LABs and B) HABs in steel subsurface. Adapted from [12]

6.4. WEB Features

WEBs are not homogeneous, consisting of distinct features compared to the surrounding microstructure. Some discrepancies have also been reported between 30° and 80° bands relating to features and structures of the bands. Unlike DERs, primary spheroidised carbides

are completely broken down in WEBs [13]. Figure 22 shows the altered microstructures consisting of 30° and 80° bands etched with nital under SE SEM (Figure 22a) in comparison to EDX maps (Figure 22b & c). Experiments carried out by Smelova et al. [13] suggests that chromium from the primary spheroidised carbides are homogenised within 30° and 80° bands (Figure 22b) while carbon is depleted homogeneously in 80° bands (Figure 22c). A similar carbon depleted region can be seen in 30° bands but are intermixed with carbon-rich areas believed to be lenticular carbides.

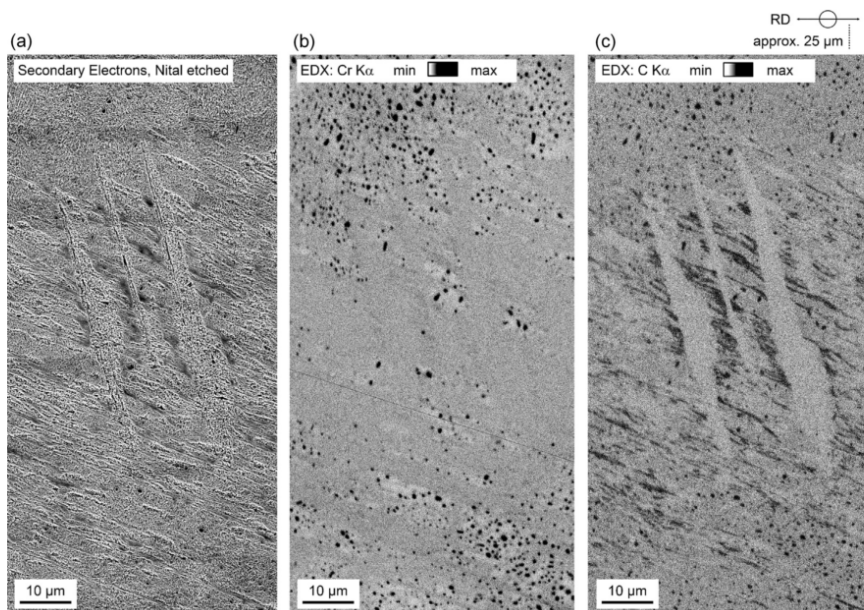


Figure 22: An image of 30° and 80° features a) Scattered electron SEM (Nital etched) b) EDX chromium map (not etched) and c) EDX carbon map (not etched) [13].

WEBs are broken down by Smelova et al [13] into 3 different regions existing within the 30° and 80° bands microstructure consisting of globular ferritic grains, elongated ferritic grains and carbon-rich areas. Dissolving primary spheroidised carbides have been observed next to the carbon-rich areas within the microstructure, indicating the source of carbon in these areas. While the hardness of the bands have been found to vary along the band microstructure where it reduces in regions with globular grains and increases in carbon-rich areas. The cause of these variations is believed to be related to the grain size variation, crystallography and chemical composition [13]. Globular and elongated ferritic grains are formed within the microstructure similar to DERs but with significantly larger grain sizes. This could indicate that microstructural alterations formed during RCF continue to grow with the number of cycles and form ferrite crystals [13, 151].

WEBs show a clear chemical redistribution compared to the unaltered microstructure, unlike DERs. The chromium released from the dissolved primary spheroidised carbides during RCF has been found to be compatible with the BCC lattice of the growing ferrite and has been incorporated into the newly formed crystal structures [13], causing a chromium

homogenisation in the altered regions independent of the orientation of the WEBs as shown in Figure 22b. However, the carbon released from the primary spheroidised carbides is incompatible with the BCC lattice and consequently pushed out of the regions where the recrystallized ferrite grains grow. Carbon has been found to accumulate between globular regions of 30° bands and near the corners in which the 30° and 80° bands intersect, leading to carbon-rich areas perhaps consisting of carbidic structures [13]. Many researches have reported these carbon-rich areas as 'disc-shaped carbides' [147, 185], 'rung-like carbides' [90] and 'lenticular carbides' [86, 87, 163, 165, 180, 186]. The driving force for recrystallization should be large if WEBs are generated through intensely localised deformation and the resulting defects can consequently aid the process of diffusion [11]. Crystallographic determinations suggest a $\{111\}\langle 211\rangle$ subsurface texture develops within the damaged materials under large rolling contact stresses (3.8 GPa) and temperatures exceeding 55 °C, a similar observation to the recrystallization texture observed in annealed cold-rolled ferritic steels [94]. A $\{100\}\langle 011\rangle$ texture is observed at comparatively low temperatures and contact stresses which is consistent with the deformation texture of cold-rolled ferritic steels [11].

7. Overview and Conclusion

Intensive research towards understanding RCF in bearings has been conducted over the past several decades yet there remains some uncertainties in many RCF aspects, particularly microstructural alterations. While there has been several debates regarding the formation mechanisms of DERs and WEBs on whether plastic deformation or carbon diffusion is the cause, a unified theory has been suggested where strain-induced carbon diffusion is believed to be the cause based on recent research. However, there are some uncertainties yet to be answered such as why do the ferrite patches in DER show no distinct direction whereas ferrite bands in WEBs are aligned at 30° and 80° and how does the dislocations within the ferrite band effect the carbon migration intensity and direction. A common confusion in literature is distinguishing between hard and soft WEM and the terms associated with each phenomenon where some papers treat them both as WEAs. However, as suggested in this paper hard WEM (WECs, WEA, butterflies) demonstrate variations in features and properties compared to soft WEM (WEBs) even though some have reported similarities in the formation mechanism and grain structure within the altered regions. An overview of the bearing cycles and corresponding contact pressure for DER, LABs and HABs formation reported in experiments from the literature is shown in Figure 23 which shows the relationship between pressure and stress cycles for the formation of each microstructural alteration (DER, LABs and HABs). Although it is usual to document bearing life in literature as number of revolutions as this is a application-driven number, analysis of stress cycles would provide a more accurate comparison between various bearing geometries and hence this might cause confusion when comparing cycles in literature.

Research over the past several decades has aimed towards analysing RCF in bearings and the associated microstructural alterations. A detailed review has been conducted on different stages of RCF and the different parameters that can affect the bearing life during operation.

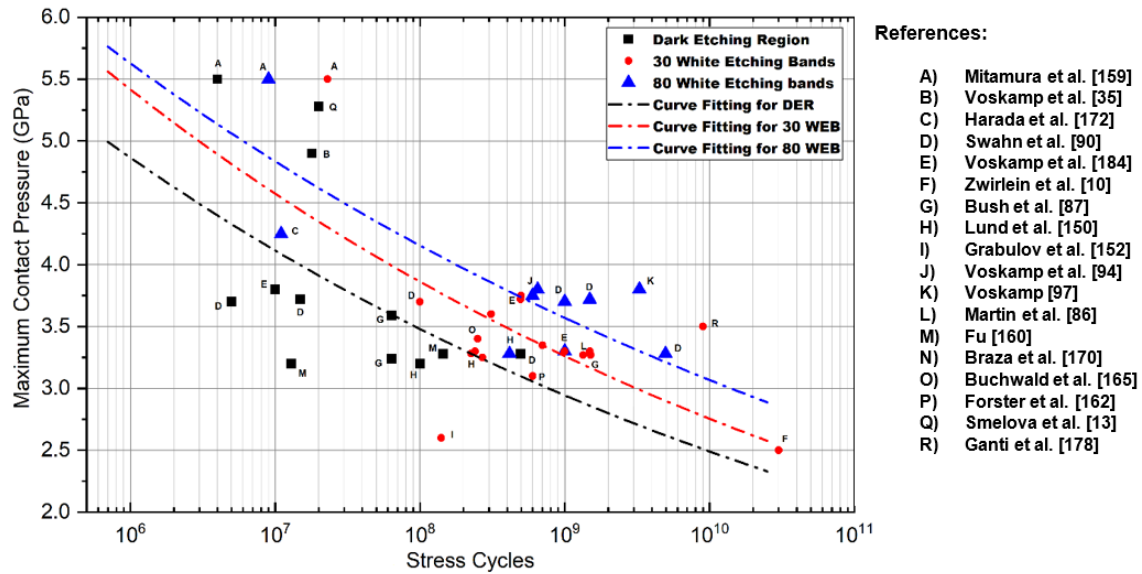


Figure 23: Overview of experimental parameters from literature reporting DER, 30° WEB and 80° WEB. Adapted from [2].

A comprehensive analysis of the associated microstructural alterations known as dark etching regions, 30° white etching bands and 80° white etching bands including their properties and possible formation mechanisms obtained over the past several decades of research has been reviewed. It has been clearly defined that DER and WEB formation accelerates with contact pressure and increases with stress cycles but there has been much less focus on analysing other operating factors such as temperature, friction and speed in relation with DER/WEB development. Nonetheless, there has indeed been significant research towards these features but there still remains a lack of understanding towards how these features contribute to RCF failure and crack initiation. Majority of RCF failure has been towards crack propagation and initiation from inclusions within the microstructure which many agree is the main source of crack initiation. However, no papers have investigated how subsurface phase changes and microstructural alterations such as DER/WEB during bearing operation affect crack initiation. It remains unclear if the boundaries of DER/WEB could cause crack initiation points or whether the feature development is related to inclusions similar to cracks. As an important step towards understanding DER/WEB formation, it is vital to determine at what RCF stage do these features occur as different papers have stated different stages but limited experimental work has analysed. RCF is a very complex phenomenon and indeed there has been multiple research over decades analysing its development in bearings including steel inclusions, heat treatment, stress development, crack initiation and propagation modelling and phase development through the bearing life but further investigation, particularly towards microstructural alterations in the subsurface, is necessary to develop a better understanding of RCF.

References

- [1] T. A. Harris, M. N. Kotzalas, Essential concepts of bearing technology, CRC press, 2006.
- [2] A. Warhadpande, F. Sadeghi, R. D. Evans, Microstructural alterations in bearing steels under rolling contact fatigue part 1: historical overview, *Tribology Transactions* 56 (3) (2013) 349–358.
- [3] J.-H. Kang, B. Hosseinkhani, P. E. Rivera-Díaz-del Castillo, Rolling contact fatigue in bearings: multiscale overview, *Materials Science and Technology* 28 (1) (2012) 44–49.
- [4] J. Kang, Mechanisms of microstructural damage during rolling contact fatigue of bearing steels, Ph.D. thesis, University of Cambridge (2014).
- [5] R. Ahmed, Rolling contact fatigue, *ASM handbook* 11 (2002) 941–956.
- [6] F. Sadeghi, B. Jalalahmadi, T. S. Slack, N. Raje, N. K. Arakere, A review of rolling contact fatigue, *Journal of tribology* 131 (4) (2009) 041403.
- [7] T. A. Harris, M. N. Kotzalas, Advanced concepts of bearing technology: rolling bearing analysis, CRC Press, 2006.
- [8] A. Jones, Metallographic Observations of Ball Bearing Fatigue Phenomenon, Philadelphia, 1947.
- [9] H. Muro, N. Tsushima, Microstructural, microhardness and residual stress changes due to rolling contact, *Wear* 15 (1970) 309–330.
- [10] O. Zwirlein, H. Schlicht, Rolling contact fatigue mechanisms: accelerated testing versus field performance, in: *Rolling contact fatigue testing of bearing steels*, ASTM International, 1982.
- [11] H. Bhadeshia, Steels for bearings, *Progress in materials Science* 57 (2) (2012) 268–435.
- [12] I. Polonsky, L. Keer, On white etching band formation in rolling bearings, *Journal of the Mechanics and Physics of Solids* 43 (4) (1995) 637–669.
- [13] V. Šmeļova, A. Schwedt, L. Wang, W. Holweger, J. Mayer, Electron microscopy investigations of microstructural alterations due to classical rolling contact fatigue (rcf) in martensitic aisi 52100 bearing steel, *International Journal of Fatigue* 98 (2017) 142–154.
- [14] Heat-treated steels, alloy steels and free-cutting steels - part 17: Ball and roller bearing steels, Tech. rep., International Standard ISO 683-17 (2014).
- [15] Y. Hirotsu, S. Nagakura, Crystal structure and morphology of the carbide precipitated from martensitic high carbon steel during the first stage of tempering, *Acta Metallurgica* 20 (4) (1972) 645–655.
- [16] C.-B. Ma, T. Ando, D. Williamson, G. Krauss, Chi-carbide in tempered high carbon martensite, *Metallurgical Transactions A* 14 (6) (1983) 1033–1045.
- [17] H. Borchers, K. Doffin, Kinetik der bildung des karbids fe₂c in stahl 100 cr 6, *steel research international* 40 (6) (1969) 493–498.
- [18] M.-H. Evans, White structure flaking failure in bearings under rolling contact fatigue, Ph.D. thesis, University of Southampton (2013).
- [19] A. S. Irwin, W. J. Anderson, W. J. Derner, Review and critical analysis: Rolling-element bearings for system life and reliability.
- [20] E. V. Zaretsky, Rolling bearing steels - a technical and historical perspective, *Materials Science and Technology* 28 (1) (2012) 58–69.
- [21] R. Baughman, Effect of hardness, surface finish, and grain size on rolling-contact fatigue life of m-50 bearing steel, *Journal of Basic Engineering* 82 (2) (1960) 287–294.
- [22] E. V. Zaretsky, Selection of rolling-element bearing steels for long-life applications, in: *Effect of Steel Manufacturing Processes on the Quality of Bearing Steels*, ASTM International, 1988.
- [23] C. Ooki, Improving rolling contact fatigue life of bearing steels through grain refinement, Tech. rep., SAE Technical Paper (2004).
- [24] M. Holzweissig, M. Uslu, H.-G. Lambers, D. Canadinc, H. Maier, A comparative analysis of austenite-to-martensite and austenite-to-bainite phase transformation kinetics in steels, *Materials Research Letters* 1 (3) (2013) 141–147.
- [25] D. H. Herring, A discussion of retained austenite, Tech. rep., *Industrial Heating* (2005).
- [26] M. Abareishi, E. Emadoddin, Effect of retained austenite characteristics on fatigue behavior and tensile properties of transformation induced plasticity steel, *Materials & Design* 32 (10) (2011) 5099–5105.
- [27] Y. Shen, S. M. Moghadam, F. Sadeghi, K. Paulson, R. W. Trice, Effect of retained austenite–

- compressive residual stresses on rolling contact fatigue life of carburized aisi 8620 steel, *International Journal of Fatigue* 75 (2015) 135–144.
- [28] Z. Hu, M. Ma, Y. Liu, J. Liu, The effect of austenite on low cycle fatigue in three-phase steel, *International journal of fatigue* 19 (8-9) (1997) 641–646.
- [29] A. P. Voskamp, Microstructural changes during rolling contract fatigue: Metal fatigue in the subsurface region of deep groove ball bearing inner rings.
- [30] E. Zaretsky, Rolling bearing and gear materials, *Tribology for aerospace applications* (1997) 325–451.
- [31] B. A. Voskamp, T. Andersson, Experience of element and full-bearing testing of materials over several years. in rolling contact fatigue testing of bearing steels, *ASTM STP 771* (1982) 190–205.
- [32] H. Muro, N. Tsushima, Microstructural, microhardness and residual stress changes due to rolling contact, *Wear* 15 (5) (1970) 309–330.
- [33] G. Krauss, *Steels: heat treatment and processing principles*, ASM International, 1990, (1990) 497.
- [34] H. Böhmer, R. Eberhard, Microstructural optimisation of bearing steels for operation under contaminated lubrication by using the experimental method of dented surfaces, in: *Bearing steel technology*, ASTM International, 2002.
- [35] A. Voskamp, E. Mittemeijer, The effect of the changing microstructure on the fatigue behaviour during cyclic rolling contact loading, *Zeitschrift für Metallkunde* 88 (4) (1997) 310–320.
- [36] M. Paladugu, R. S. Hyde, Influence of microstructure on retained austenite and residual stress changes under rolling contact fatigue in mixed lubrication conditions, *Wear* 406 (2018) 84–91.
- [37] P. Walker, Improving the reliability of highly loaded rolling bearings: the effect of upstream processing on inclusions, *Materials Science and Technology* 30 (4) (2014) 385–410.
- [38] E. S. Alley, R. W. Neu, Microstructure-sensitive modeling of rolling contact fatigue, *International Journal of Fatigue* 32 (5) (2010) 841–850.
- [39] S. Li, Effects of inclusions on very high cycle fatigue properties of high strength steels, *International Materials Reviews* 57 (2) (2012) 92–114.
- [40] R. Baum, K. Böhnke, T. Boeckers, H. Klemp, Properties of through hardening bearing steels produced by bof blowing metallurgy and by electric arc furnace with ladle metallurgy, *Steel Research* 58 (9) (1987) 432–438.
- [41] J. Akesson, Skf rolling bearing steels-properties and processes, *Ball Bearing Journal* 217 (1983) 32.
- [42] Y. Murakami, H. Usuki, Quantitative evaluation of effects of non-metallic inclusions on fatigue strength of high strength steels. ii: Fatigue limit evaluation based on statistics for extreme values of inclusion size, *International Journal of Fatigue* 11 (5) (1989) 299–307.
- [43] T. B. Lund, Sub-surface initiated rolling contact fatigue influence of non-metallic inclusions, processing history, and operating conditions, *Journal of ASTM International* 7 (5) (2010) 1–12.
- [44] M. Dinkel, W. Trojahn, Influence of sulfur inclusion content on rolling contact fatigue life, in: *Bearing Steel Technologies: 10th Volume, Advances in Steel Technologies for Rolling Bearings*, ASTM International, 2014.
- [45] K. Hashimoto, K. Hiraoka, K. Kida, E. Costa Santos, Effect of sulphide inclusions on rolling contact fatigue life of bearing steels, *Materials science and Technology* 28 (1) (2012) 39–43.
- [46] R. Kiessling, *I. of Metals, Non-metallic inclusions in steel. Part 5*, London : Institute of Metals, 1989, "Book number 411."–T.p. verso.
- [47] N. K. Arakere, N. Branch, G. Levesque, V. Svendsen, N. H. Forster, Rolling contact fatigue life and spall propagation of aisi m50, m50nil, and aisi 52100, part ii: stress modeling, *Tribology Transactions* 53 (1) (2009) 42–51.
- [48] T. Sakai, Y. Sato, N. Oguma, Characteristic s–n properties of high-carbon–chromium-bearing steel under axial loading in long-life fatigue, *Fatigue & Fracture of Engineering Materials & Structures* 25 (8-9) (2002) 765–773.
- [49] A. Evans, The role of inclusions in the fracture of ceramic materials, *Journal of Materials Science* 9 (7) (1974) 1145–1152.
- [50] W. Leslie, Inclusions and mechanical properties, *Mechanical Working and Steel Processing XX* (1982) 3–50.

- [51] A. Ghosh, Secondary steelmaking: Principles and applications, india institute of technology (2001).
- [52] G. Auclair, F. Ruby-Meyer, R. Meilland, P. Rocabois, Cleanliness assessment: a critical review and a real need to predict rolling contact fatigue behaviour, in: Bearing Steels: Into the 21st Century, ASTM International, 1998.
- [53] W. Duckworth, E. Ineson, The effects of externally introduced alumina particles on the fatigue life of en24 steel, *Clean steel* 77 (1963) 87–103.
- [54] J. Lankford, Effect of oxide inclusions on fatigue failure, *International Metals Reviews* 22 (1) (1977) 221–228.
- [55] P. Thornton, The influence of nonmetallic inclusions on the mechanical properties of steel: A review, *Journal of Materials Science* 6 (4) (1971) 347–356.
- [56] M. Atkinson, The influence of nonmetallic inclusions on the fatigue properties of ultra-high-tensile steels, *J. Iron Steel Inst* 195 (1960) 64.
- [57] N. Ånmark, A. Karasev, P. G. Jönsson, The effect of different non-metallic inclusions on the machinability of steels, *Materials* 8 (2) (2015) 751–783.
- [58] K. Tanaka, T. Mura, A theory of fatigue crack initiation at inclusions, *Metallurgical Transactions A* 13 (1) (1982) 117–123.
- [59] P. Rycerz, A. Olver, A. Kadiric, Propagation of surface initiated rolling contact fatigue cracks in bearing steel, *International Journal of Fatigue* 97 (2017) 29–38.
- [60] C. Longching, C. Qing, S. Eryu, Study on initiation and propagation angles of subsurface cracks in gcr15 bearing steel under rolling contact, *Wear* 133 (2) (1989) 205–218.
- [61] T. Yoshioka, T. Fujiwara, Paper ii (i) measurement of propagation initiation and propagation time of rolling contact fatigue cracks by observation of acoustic emission and vibration, in: *Tribology Series*, Vol. 12, Elsevier, 1987, pp. 29–33.
- [62] G. Lundberg, A. Palmgren, Dynamic capacity of rolling bearings, *Polytechnica Scandinavica. Mechanical Engineering Series* 1 (3) (1947) 1–52.
- [63] E. Ioannides, T. A. Harris, A new fatigue life model for rolling bearings, *Journal of Tribology* 107 (3) (1985) 367–377.
- [64] K. Hashimoto, T. Fujimatsu, N. Tsunekage, K. Hiraoka, K. Kida, E. C. Santos, Effect of inclusion/matrix interface cavities on internal-fracture-type rolling contact fatigue life, *Materials & Design* 32 (10) (2011) 4980–4985.
- [65] L. Zhang, B. G. Thomas, Inclusions in continuous casting of steel, in: *XXIV National Steelmaking Symposium*, Morelia, Mich, Mexico, Vol. 26, Citeseer, 2003, p. 28.
- [66] R. Allen, Standard test methods for determining the inclusion content of steel (e45), *Annual Book of ASTM Standards, Metal-Mechanical Testing; Elevated and Low Temperature Tests; Metallography*.
- [67] deutsche norm, Microscopic examination of special steels using standard diagrams to assess the content of non-metallic inclusions, Tech. rep., International Organization for Standardization (1985).
- [68] D. Hetzner, Developing astm e 2283: Standard practice for extreme value analysis of nonmetallic inclusions in steel and other microstructural features, in: *Bearing Steel Technology-Advances and State of the Art in Bearing Steel Quality Assurance: 7th Volume*, ASTM International, 2007.
- [69] Olympus, Non-metallic inclusion analysis in steel, <https://www.olympus-ims.com/en/applications/nmi-analysis/>, [Last Accessed 13th Jul 2018] (2016).
- [70] D. Dowson, M. Priest, G. Dalmaz, A. Lubrecht, *Tribological Research and Design for Engineering Systems: Proceedings of the 29th Leeds-Lyon Symposium*, Elsevier, 2003.
- [71] G. Stachowiak, A. W. Batchelor, *Engineering tribology*, Butterworth-Heinemann, 2013.
- [72] K. Johnson, *Contact Mechanics*, University of Cambridge Press, 1985.
- [73] B. J. Hamrock, D. Dowson, *Ball bearing lubrication: the elastohydrodynamics of elliptical contacts*.
- [74] M. Qiu, L. Chen, Y. Li, J. Yan, *Bearing Tribology: Principles and Applications*, Springer, 2016.
- [75] E. V. Zaretsky, *Rolling bearing life prediction, theory, and application*, 2013.
- [76] N. Tsushima, K. Maeda, H. Nakashima, H. Kashimura, New steels now in use for automotive rolling element bearings, in: *Bearing Steels: Into the 21st Century*, ASTM International, 1998.
- [77] E. Ioannides, An analytical formulation for the life of rolling bearings, *Acta Polytechnica Scandinavica*,

Mechanical Engineering Series.

- [78] J. Tao, T. Hughes, H. Evans, R. Snidle, N. Hopkinson, M. Talks, J. Starbuck, Elastohydrodynamic lubrication analysis of gear tooth surfaces from micropitting tests, *Journal of tribology* 125 (2) (2003) 267–274.
- [79] G. E. Totten, *Handbook of hydraulic fluid technology*, CRC Press, 2011.
- [80] W. Greenert, The toroid contact roller test as applied to the study of bearing materials, *Journal of Basic Engineering* 84 (1) (1962) 181–189.
- [81] H. Fessler, E. Ollerton, Contact stresses in toroids under radial loads, *British journal of applied physics* 8 (10) (1957) 387.
- [82] G. Moyer, J. Morrow, Relation of fatigue to pitting failure of surfaces in rolling contact, *Appl. Mech. Rep* 565.
- [83] N. Popinceanu, E. Diaconescu, S. Cretu, Critical stresses in rolling contact fatigue, *Wear* 71 (3) (1981) 265–282.
- [84] K. Fujita, A. Yoshida, Surface durability of case-hardened nickel-chromium steel rollers under pure rolling and sliding-rolling contacts, *Wear* 52 (1) (1979) 37–48.
- [85] J. Martin, A. Eberhardt, Identification of potential failure nuclei in rolling contact fatigue, *Journal of Basic Engineering* 89 (4) (1967) 932–942.
- [86] J. Martin, S. Borgese, A. Eberhardt, Microstructural alterations of rollingbearing steel undergoing cyclic stressing, *Journal of Basic Engineering* 88 (3) (1966) 555–565.
- [87] J. Bush, W. Grube, G. Robinson, Microstructural and residual stress changes in hardened steel due to rolling contact, *Trans. ASM* 54 (1961) 390–412.
- [88] G. Vasilca, V. Raszillier, A study of dark etching area (d.e.a.) type structure modification of material and hertzian contact area induced by ball bearing type motion, *Wear* 19 (1972) 1–15.
- [89] H. Schlicht, Material properties adapted to the actual stressing in a rolling bearing, *Ball and roller bearing engineering* 1 (1981) 24–29.
- [90] H. Swahn, P. Becker, O. Vingsbo, Martensite decay during rolling contact fatigue in ball bearings, *Metallurgical transactions A* 7 (8) (1976) 1099–1110.
- [91] M. Cerullo, Application of dang van criterion to rolling contact fatigue in wind turbine roller bearings under elastohydrodynamic lubrication conditions, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 228 (12) (2014) 2079–2089.
- [92] G. E. Morales-Espejel, V. Brizmer, Micropitting modelling in rolling–sliding contacts: application to rolling bearings, *Tribology Transactions* 54 (4) (2011) 625–643.
- [93] V. Bhargava, G. Hahn, C. Rubin, Rolling contact deformation, etching effects, and failure of high-strength bearing steel, *Metallurgical Transactions A* 21 (7) (1990) 1921–1931.
- [94] A. Voskamp, E. Mittemeijer, Crystallographic preferred orientation induced by cyclic rolling contact loading, *Metallurgical and Materials Transactions A* 27 (1996) 3445–3465.
- [95] S. Turteltaub, A. Suiker, Transformation-induced plasticity in ferrous alloys, *Journal of the Mechanics and Physics of Solids* 53 (8) (2005) 1747–1788.
- [96] N. Arakere, G. Subhash, Work hardening response of m50-nil case hardened bearing steel during shakedown in rolling contact fatigue, *Materials Science and Technology* 28 (1) (2012) 34–38.
- [97] A. Voskamp, Material response to rolling contact loading, *J Tribol* 107 (1985) 359–364.
- [98] R. Vegter, J. Slycke, The role of hydrogen on rolling contact fatigue response of rolling element bearings, *Journal of ASTM international* 7 (2) (2009) 1–12.
- [99] A. Bower, K. Johnson, Plastic flow and shakedown of the rail surface in repeated wheelrail contact, in: *Mechanics and Fatigue in Wheel/Rail Contact*, Elsevier, 1991, pp. 1–18.
- [100] G. Hahn, V. Bhargava, Q. Chen, The cyclic stress-strain properties, hysteresis loop shape, and kinematic hardening of two high-strength bearing steels, *Metallurgical Transactions A* 21 (2) (1990) 653.
- [101] J. Williams, The influence of repeated loading, residual stresses and shakedown on the behaviour of tribological contacts, *Tribology International* 38 (9) (2005) 786–797.
- [102] S. Kulkarni, G. Hahn, C. Rubin, V. Bhargava, Elasto-plastic finite element analysis of three-dimensional pure rolling contact above the shakedown limit, *Journal of Applied Mechanics* 58 (2)

- (1991) 347–353.
- [103] G. Popescu, G. E. Morales-Espejel, B. Wemekamp, A. Gabelli, An engineering model for three-dimensional elastic-plastic rolling contact analyses, *Tribology transactions* 49 (3) (2006) 387–399.
 - [104] K. D. Van, M. Maitournam, Steady-state flow in classical elastoplasticity: applications to repeated rolling and sliding contact, *Journal of the Mechanics and Physics of Solids* 41 (11) (1993) 1691–1710.
 - [105] M. M.-H. Yu, B. Moran, L. M. Keer, A direct analysis of three-dimensional elastic-plastic rolling contact, *Journal of tribology* 117 (2) (1995) 234–243.
 - [106] C.-C. Yu, B. Moran, L. Keer, A simplified direct method for cyclic strain calculation: repeated rolling/sliding contact on a case-hardened half plane, *Journal of tribology* 118 (2) (1996) 329–334.
 - [107] C. Sakae, L. Keer, Application of direct method for a nonlinear-kinematic-hardening material under rolling/sliding line contact: constant ratchetting rate, *Journal of the Mechanics and Physics of Solids* 45 (9) (1997) 1577–1594.
 - [108] M. Yu, B. Moran, L. Keer, A direct analysis of two-dimensional elastic-plastic rolling contact, *Journal of tribology* 115 (2) (1993) 227–236.
 - [109] E. Melan, Der spannungszustand eines mises-henckys chen kontinuums bei veraderlicher belastung, *Sitzber Akad. Wiss. Wien Iia* 147 (1938) 73–87.
 - [110] W. Koiter, A new general theorem on shakedown of elastic-plastic structures, *Proc. Koninkl. Ned. Akad. Wet. B* 59 (1956) 24–34.
 - [111] D. Pham, D. Weichert, Shakedown analysis for elastic-plastic bodies with limited kinematic hardening, in: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 457, The Royal Society, 2001, pp. 1097–1110.
 - [112] Y. Jiang, B. Xu, H. Sehitoglu, Three-dimensional elastic-plastic stress analysis of rolling contact, *Journal of tribology* 124 (4) (2002) 699–708.
 - [113] A. M. Kumar, G. T. Hahn, C. A. Rubin, A study of subsurface crack initiation produced by rolling contact fatigue, *Metallurgical Transactions A* 24 (2) (1993) 351.
 - [114] S. Turteltaub, A. Suiker, Transformation-induced plasticity in ferrous alloys, *Journal of the Mechanics and Physics of Solids* 53 (8) (2005) 1747–1788.
 - [115] G. E. Morales-Espejel, V. Brizmer, Micropitting modelling in rolling–sliding contacts: application to rolling bearings, *Tribology Transactions* 54 (4) (2011) 625–643.
 - [116] P. Rabaso, T. Gauthier, M. Diaby, F. Ville, Rolling contact fatigue: experimental study of the influence of sliding, load, and material properties on the resistance to micropitting of steel discs, *Tribology Transactions* 56 (2) (2013) 203–214.
 - [117] F. Sadeghi, P. C. Sui, Internal stresses in elastohydrodynamic lubrication of rolling/sliding contacts, *Journal of tribology* 111 (1) (1989) 180–187.
 - [118] J. Lord, R. Larsson, Effects of slide-roll ratio and lubricant properties on elastohydrodynamic lubrication film thickness and traction, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 215 (3) (2001) 301–308.
 - [119] D. Nuruzzaman, M. Sheikh, Ehl oil film thickness under rolling-sliding contact, *Journal of Mechanical Engineering* 38 (2007) 58–60.
 - [120] O. I. de Normalización, *Rolling Bearings: Dynamic Load Ratings and Rating Life*, ISO, 2007.
 - [121] ISO, *Rolling bearings dynamic load ratings and rating life*, Tech. rep., Draft International Standard ISO/DIS 281 (1989).
 - [122] W. Cheng, H. Cheng, T. Mura, L. M. Keer, Micromechanics modeling of crack initiation under contact fatigue, *Journal of tribology* 116 (1) (1994) 2–8.
 - [123] W. W. Cheng, H. S. Cheng, Semi-analytical modeling of crack initiation dominant contact fatigue life for roller bearings, *Journal of tribology* 119 (2) (1997) 233–240.
 - [124] A. Vincent, G. Lormand, P. Lamagnere, L. Gosset, D. Girodin, G. Dudragne, R. Fougères, From white etching areas formed around inclusions to crack nucleation in bearing steels under rolling contact fatigue, in: *Bearing steels: into the 21st century*, ASTM International, 1998.
 - [125] Y. Jiang, H. Sehitoglu, A model for rolling contact failure, *Wear* 224 (1) (1999) 38–49.
 - [126] L. M. Keer, M. Bryant, A pitting model for rolling contact fatigue, *Journal of lubrication technology*

- 105 (2) (1983) 198–205.
- [127] G. Lormand, P. Meynaud, A. Vincent, G. Baudry, D. Girodin, G. Dudragne, From cleanliness to rolling fatigue life of bearings a new approach, in: *Bearing Steels: Into the 21st Century*, ASTM International, 1998.
- [128] R. Zhou, H. Cheng, T. Mura, Micropitting in rolling and sliding contact under mixed lubrication, *Journal of Tribology* 111 (4) (1989) 605–613.
- [129] R. S. Zhou, Surface topography and fatigue life of rolling contact bearing, *Tribology Transactions* 36 (3) (1993) 329–340.
- [130] J. W. Ringsberg, Life prediction of rolling contact fatigue crack initiation, *International Journal of fatigue* 23 (7) (2001) 575–586.
- [131] F.-J. Ebert, Fundamentals of design and technology of rolling element bearings, *Chinese Journal of Aeronautics* 23 (1) (2010) 123–136.
- [132] N. R. Paulson, J. A. Bomidi, F. Sadeghi, R. D. Evans, Effects of crystal elasticity on rolling contact fatigue, *International Journal of Fatigue* 61 (2014) 67–75.
- [133] A. Vijay, N. Paulson, F. Sadeghi, A 3d finite element modelling of crystalline anisotropy in rolling contact fatigue, *International Journal of Fatigue* 106 (2018) 92–102.
- [134] N. Raje, F. Sadeghi, R. G. Rateick, M. R. Hoeprich, A numerical model for life scatter in rolling element bearings, *Journal of Tribology* 130 (1) (2008) 011011.
- [135] B. Jalalahmadi, F. Sadeghi, A voronoi finite element study of fatigue life scatter in rolling contacts, *Journal of tribology* 131 (2) (2009) 022203.
- [136] B. Zhang, H. Boffy, C. H. Venner, Multigrid solution of 2d and 3d stress fields in contact mechanics of anisotropic inhomogeneous materials, *Tribology International*.
- [137] A. Olver, The mechanism of rolling contact fatigue: an update, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 219 (5) (2005) 313–330.
- [138] B. Jalalahmadi, F. Sadeghi, V. Bakolas, Material inclusion factors for lundberg-palmgren-based ref life equations, *Tribology Transactions* 54 (3) (2011) 457–469.
- [139] K.-H. Kim, S.-D. Park, J.-H. Kim, C.-M. Bae, Role of spheroidized carbides on the fatigue life of bearing steel, *Metals and Materials International* 18 (6) (2012) 917–921.
- [140] J. Guan, L. Wang, Y. Li, C. Zhang, L. Gu, Influence of rough surface on damage evolution and fatigue life of m50-bearing steel containing a spherical inclusion, *International Journal of Damage Mechanics* (2019) 1056789519842367.
- [141] I. Krupka, P. Sperka, M. Hartl, Effect of surface roughness on lubricant film breakdown and transition from ehl to mixed lubrication, *Tribology International* 100 (2016) 116–125.
- [142] B. Allison, A. Pandkar, Critical factors for determining a first estimate of fatigue limit of bearing steels under rolling contact fatigue, *International Journal of Fatigue* 117 (2018) 396–406.
- [143] H. Fu, W. Song, E. I. Galindo-Nava, P. E. Rivera-Díaz-del Castillo, Strain-induced martensite decay in bearing steels under rolling contact fatigue: modelling and atomic-scale characterisation, *Acta Materialia* 139 (2017) 163–173.
- [144] A. Bhattacharyya, G. Subhash, N. Arakere, Evolution of subsurface plastic zone due to rolling contact fatigue of m-50 nil case hardened bearing steel, *International Journal of Fatigue* 59 (2014) 102–113.
- [145] K. Sugino, K. Miyamoto, M. Nagumo, K. Aoki, Structural alterations of bearing steels under rolling contact fatigue, *Transactions of the Iron and Steel Institute of Japan* 10 (2) (1970) 98.
- [146] D. Morris, F. Sadeghi, Y.-C. Chen, C. Wang, B. Wang, A novel approach for modeling retained austenite transformations during rolling contact fatigue, *Fatigue & Fracture of Engineering Materials & Structures* 41 (4) (2018) 831–843.
- [147] R. Österlund, O. Vingsbo, Phase changes in fatigued ball bearings, *Metallurgical Transactions A* 11 (5) (1980) 701–707.
- [148] H.-K. Lorösch, Influence of load on the magnitude of the life exponent for rolling bearings, in: *Rolling contact fatigue testing of bearing steels*, ASTM International, 1982.
- [149] H. Swahn, P. Becker, O. Vingsbo, Electron-microscope studies of carbide decay during contact fatigue in ball bearings, *Metal Science* 10 (1) (1976) 35–39.

- [150] T. Lund, Structural alterations in fatigue-tested ball-bearing steel, *Jernkontorets Ann* 153 (7) (1969) 337–343.
- [151] H. Fu, P. E. Rivera-Díaz-del Castillo, A unified theory for microstructural alterations in bearing steels under rolling contact fatigue, *Acta Materialia* 155 (2018) 43–55.
- [152] A. Grabulov, U. Ziese, H. Zandbergen, Tem/sem investigation of microstructural changes within the white etching area under rolling contact fatigue and 3-d crack reconstruction by focused ion beam, *Scripta Materialia* 57 (7) (2007) 635–638.
- [153] J. Beswick, Measurement of c levels in structurally transformed sae 52100 ball-bearing steel by microprobe analysis, *Praktische Metallographie* 12 (4) (1975) 200–206.
- [154] A. Muroga, H. Saka, Analysis of rolling contact fatigued microstructure using focused ion beam sputtering and transmission electron microscopy observation, *Scripta metallurgica et materialia* 33 (1).
- [155] W. Solano Alvarez, Microstructural degradation of bearing steels, Ph.D. thesis, University of Cambridge (2015).
- [156] G. D.Scott, B.Loy, Metallurgical aspects of rolling contact fatigue, *Proc. Inst. Mech. Eng. Conf. Proc* 181 (1966) 94–103.
- [157] J.-H. Kang, B. Hosseinkhani, R. H. Vegter, P. E. Rivera-Díaz-del Castillo, Modelling dislocation assisted tempering during rolling contact fatigue in bearing steels, *International Journal of Fatigue* 75 (2015) 115–125.
- [158] G. F. V. Voort, *ASM Handbook Volume 9: Metallography and Microstructures*, ASM International, Burlington, 2004.
- [159] H. S. Mitamura, N. H. Takaki, Microstructural development in bearing steel during rolling contact fatigue, *Materials Science Forum* (2007) 4225–4260.
- [160] H. Fu, Microstructural alterations in bearing steels under rolling contact fatigue, Ph.D. thesis, University of Cambridge (2017).
- [161] W. F. Smith, J. Hashemi, *Foundations of materials science and engineering*, McGraw-Hill, 2011.
- [162] N. H. Forster, L. Rosado, W. P. Ogden, H. K. Trivedi, Rolling contact fatigue life and spall propagation characteristics of aisi m50, m50 nil, and aisi 52100, part iii: metallurgical examination, *Tribology Transactions* 53 (1) (2009) 52–59.
- [163] H. Fu, E. Galindo-Nava, P. Rivera-Díaz-del Castillo, Modelling and characterisation of stress-induced carbide precipitation in bearing steels under rolling contact fatigue, *Acta Materialia* 128 (2017) 176–187.
- [164] R. O. Ritchie, Influence of microstructure on near-threshold fatigue-crack propagation in ultra-high strength steel, *Metal Science* 11 (8-9) (1977) 368–381.
- [165] J. Buchwald, R. Heckel, An analysis of microstructural changes in 52100 steel bearings during cyclic stressing(microstructural changes in 52100 steel bearing inner rings during cyclic stressing, obtaining thickening rate data on white-etching regions and lenticular carbides), *ASM Transactions Quarterly* 61 (1968) 750–756.
- [166] M. Evans, White structure flaking (wsf) in wind turbine gearbox bearings: effects of butterflies and white etching cracks (wecs), *Materials Science and Technology* 28 (1) (2012) 3–22.
- [167] K. Hiraoka, T. Fujimatsu, K. Hashimoto, S. Fukumoto, A. Yamamoto, Crystallographic analyses on cracks initiated by rolling contact fatigue in high carbon chromium bearing steel, in: *Materials Science Forum*, Vol. 561, Trans Tech Publ, 2007, pp. 2151–2154.
- [168] V. Šmeļova, A. Schwedt, L. Wang, W. Holweger, J. Mayer, Microstructural changes in white etching cracks (wecs) and their relationship with those in dark etching region (der) and white etching bands (webs) due to rolling contact fatigue (rcf), *International Journal of Fatigue* 100 (2017) 148–158.
- [169] M. Maalekian, E. Kozeschnik, A thermodynamic model for carbon trapping in lattice defects, *Calphad* 32 (4) (2008) 650–654.
- [170] J. Braza, P. Pearson, C. Hannigan, The performance of 52100, m-50, and m-50 nil steels in radial bearings, *Tech. rep.*, SAE Technical Paper (1993).
- [171] W. Solano-Alvarez, H. Bhadeshia, White-etching matter in bearing steel. part ii: distinguishing cause and effect in bearing steel failure, *Metallurgical and Materials Transactions A* 45 (11) (2014) 4916–

- 4931.
- [172] H. Harada, T. Mikami, M. Shibata, D. Sokai, A. Yamamoto, H. Tsubakino, Microstructural changes and crack initiation with white etching area formation under rolling/sliding contact in bearing steel, *ISIJ international* 45 (12) (2005) 1897–1902.
 - [173] A. Grabulov, R. Petrov, H. Zandbergen, Ebsd investigation of the crack initiation and tem/fib analyses of the microstructural changes around the cracks formed under rolling contact fatigue (rcf), *International Journal of Fatigue* 32 (3) (2010) 576–583.
 - [174] Y.-S. Su, S.-X. Li, S.-Y. Lu, L.-B. Wan, Phase transformation in white etching area in rolling contact fatigue, in: *MATEC Web of Conferences*, Vol. 165, EDP Sciences, 2018, p. 11004.
 - [175] P. Becker, Metallography of structural-changes in ball-bearing steels caused by rolling-contact fatigue, in: *Ultramicroscopy*, Vol. 2, 1976, pp. 113–113.
 - [176] M.-H. Evans, An updated review: white etching cracks (wecs) and axial cracks in wind turbine gearbox bearings, *Materials Science and Technology* 32 (11) (2016) 1133–1169.
 - [177] F. Manieri, K. Stadler, G. E. Morales-Espejel, A. Kadiric, The origins of white etching cracks and their significance to rolling bearing failures, *International Journal of Fatigue* 120 (2019) 107–133.
 - [178] S. Ganti, B. Turner, M. Kirsch, D. Anthony, B. McCoy, H. Trivedi, V. Sundar, Three-dimensional (3d) analysis of white etching bands (webs) in aisi m50 bearing steel using automated serial sectioning, *Materials Characterization* 138 (2018) 11–18.
 - [179] K. Johnson, Formation of shear bands in ball-bearing races, University of Cambridge Department of Engineering, 1988.
 - [180] S. Borgese, A study of the growth mechanism of lenticular carbides in cyclically stressed 52100 steel, *Journal of Lubrication Technology* 92 (1) (1970) 54–58.
 - [181] T. Ochi, Y. Kusano, Change in microstructure and properties in the rolling contact fatigue of bearing steel, *Tech. rep.* (1999).
 - [182] A. H. Cottrell, B. Bilby, Dislocation theory of yielding and strain ageing of iron, *Proceedings of the Physical Society. Section A* 62 (1) (1949) 49.
 - [183] J. Lyman, A shear-stress normal-strain cycle relation to fatigue in bending, torsion, and rolling contact, *Journal of Lubrication Technology* 92 (4) (1970) 567–571.
 - [184] A. Voskamp, R. Österlund, P. Becker, O. Vingsbo, Gradual changes in residual stress and microstructure during contact fatigue in ball bearings, *Metals Technology* 7 (1) (1980) 14–21.
 - [185] E. Lindahl, R. Österlund, 212 transmission electron microscopy applied to phase transformations in ball bearings, *Ultramicroscopy* 9 (4) (1982) 355–364.
 - [186] R. Tricot, J. Monnot, M. Lluansi, How microstructural alterations affect fatigue properties of 52100 steel, *Met Eng Q* 12 (2) (1972) 39–47.