Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics – state of the art

J Halme* and P Andersson
VTT Technical Research Centre of Finland, Espoo, Finland

The manuscript was received on 15 April 2009 and was accepted after revision for publication on 27 October 2009.

DOI: 10.1243/13506501JET656

Abstract: Rolling bearing operation is affected by friction, wear and lubrication mechanisms, fluid dynamics and lubricant rheology, material properties, and contact mechanics. Changes in rolling surfaces occur due to plastic deformation, rolling contact wear, and rolling contact fatigue. Wear particles can be formed and mixed into the lubricant. Increased levels of vibrations due to surface degradation can be monitored by sensors. Rolling contact wear and rolling contact fatigue during rolling bearing operation can be diagnosed by combining measured and interpreted condition monitoring data with theory, and conclusions drawn thereof can support a continuous prognosis for the remaining bearing life. In the present work, connections between bearing diagnostics and tribological mechanisms are outlined.

Keywords: rolling bearings, wear, rolling contact fatigue, condition monitoring, diagnostics

1 INTRODUCTION

Rolling bearings have a great influence on the industrialized world. The bearing life theory has been developed more than a century ago, starting from the testing of rolling bearings by Stribeck in 1896, followed by work by Goodman in 1912 [1, 2]. A fatigue limit concept was introduced by Palmgren in 1924 [3]. In 1939, Weibull presented his statistical theory of the strength of materials as a mathematical tool for analyses on the bearing life scatter [4]. In the years 1947–1952, Lundberg and Palmgren [5, 6] published their life theory of rolling bearings, which was adopted as the formula for the basic rating life of rolling bearings, \( L_{10} \), in millions of revolutions, or

\[
L_{10} = \left( \frac{C}{P} \right)^p
\]  

(1)

where \( C \) is the basic dynamic load rating for the bearing, \( P \) is the equivalent bearing load, \( p \) is an exponent with the value \( p = 3 \) for ball bearings, and \( p = 10/3 \) for roller bearings [1, 7, 8]. The \( L_{10} \) value expresses the number of millions revolutions that 90 per cent of a bearing population is expected to survive under conventional operating conditions before the first incidence of fatigue pit formation at any of the rolling surfaces [9].

The lifetime of a rolling bearing is nowadays usually estimated not only by calculations of the basic rating life, but rather by calculations of the modified rating life \( L_{nm} \) [9] using the formula

\[
L_{nm} = a_1 \times a_{ISO} \times L_{10}
\]  

(2)

where \( a_1 \) is the modification factor for reliability other than 90 per cent and \( a_{ISO} \) represents parameters such as load, speed, oil viscosity, material fatigue limit, and content of contamination particles in the oil. From a statistical point of view, failures can be initiated in 10 per cent of a bearing population (at \( a_1 = 1 \)) before reaching the \( L_{nm} \) limit. The exact prediction of the first initiation of a failure in a specific bearing is practically not possible. For fatigue damages that occur before the calculated number of revolutions is achieved, bearing diagnostics is a powerful tool for avoiding undesired consequences of bearing failures. Furthermore, if a bearing becomes subjected to wear instead of rolling contact fatigue, the lifetime cannot be mathematically predicted on a statistical basis.

*Corresponding author: VTT Technical Research Centre of Finland, P. O. Box 1000, MK6, FIN-02044 VTT (Espoo), Finland. email: jari.halme@vtt.fi
2 ROLLING CONTACT MECHANICS

2.1 Loading of elastic solid bodies

When two non-conforming solids of geometrically ideal shapes are brought into contact, the initial contact is a point or a line, which can be analysed with the Hertzian contact theory \[173\]. The contact pressure from the normal load causes principal stresses and shear stresses beneath the contact surface (Fig. 1).

Three features of the contact mechanics can be seen in Fig. 1. First, under normal static loading, the angle of principal or maximum shear stress is \(\pm 45^\circ\) in the middle of the normal pressure (however, in a rolling and sliding contact, tangential forces affect the angle of the principal shear stress). Second, under the Hertzian contact pressure, the shear stress reaches its maximum below the surface at a depth of 0.48\(a\) and the maximum shear stress value is 0.31 \(\times p_0\) with steel.

Third, in the surface \(z = 0\) at the centre of contact \((r = 0)\), the radial stress \(\sigma_r\) is 0.8 \(\times p_0\) \[11, 12\].

2.2 Non-Hertzian normal contact of elastic bodies under rolling/sliding contact

When a rolling surface is compressed in one direction, its material tends to expand in the transverse directions according to Hooke's law and Poisson's ratio, and this creates tangential forces and local slip between contact surfaces at the trailing edge of the rolling surface \[10, 13\].

In deep groove ball bearings, the raceway geometry is a semi-conforming groove, and the contact area is a curved ellipse with three slip zones (see Fig. 2); a central zone with a micro-slip in the direction of rotation of the ball and two outer zones with a micro-slip in the opposite direction \[10\]. Similarly, micro-slip zones
Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics

2.4 Dynamical impacts and vibrations in structures surrounding rolling contacts

Rolling contacts comprise dynamically varying stress fields from surface roughness and contaminant particles. A wave motion initiated from any point of dynamic stress propagates as spherical fronts of pressure and shear waves \[10, 15\]. The interaction of pressure and shear waves with the free surface gives rise to the Rayleigh wave, which is a surface wave. In rolling contact, fatigue crack formation and contaminant particle crushing introduce high-frequency vibrations. Figure 4 shows the principle for the propagation of high-frequency vibration waves.

During the progress of the wave fronts, their energy remains almost constant, with the exception of minor losses due to hysteresis. In rolling contacts, the Rayleigh surface waves dominate and expand circularly on the material surface. Partly the waves propagate through the interface to the bearing housing and partly the energy reflects back to the bearing \[16\].

In terms of the response properties, a theoretical impulse of infinitely small duration contains equal energy at all frequencies and a linear system is completely specified by its impulse response \[17, 18\]. In a mechanical system, like a rolling bearing, an incoming impact excites all nominal vibration frequencies of the system below the frequency of the impact, and as a consequence of this secondary excitation of higher frequencies within the system occurs. Most of the energy of the vibration is bound to the nominal frequencies of the structure and the displacement amplitude is at its highest at the lower frequencies. The sustaining frequencies are typically measured with appropriate vibration acceleration sensors and the propagating wave fronts with acoustic emission (AE) sensors. The analyses of these are discussed in chapter 6.

3 LUBRICATION OF ROLLING BEARINGS

3.1 Lubrication regimes

Entrapment of a liquid into a converging gap at a sliding or rolling contact creates a hydrodynamic pressure and a lubricant film to separate the contact surfaces and reduce the contact pressure. The ratio between the oil film thickness and the combined surface roughness in the contact, or the lambda (\(\lambda\)) value, determines the lubrication mechanism and its effect \[19–22\]. An overview of lubricating mechanisms and the Stribeck curve are shown in Fig. 5 \[2, 22, 23\]. Transitions from one lubrication mechanism to another may occur for reasons of operation, running-in or wear \[20, 24, 25\].

3.2 Boundary lubrication

If the hydrodynamic action is too weak for separating the contact surfaces from each other (\(\lambda < 1\), and...
the surfaces are only wetted by oil, the load is carried mainly by solid-to-solid contacts and boundary films on the surface asperities [14, 23, 26, 27]. Boundary lubrication relies on lubricant and surface chemistry and the ambient atmosphere, on the formation rate of the boundary films and on their mechanical properties [14]. As a result of the solid-to-solid contact conditions, some wear is to be expected by tribochemical or mechanical wear [27]. In rolling bearings, boundary lubrication can occur locally in micro-slip zones and at the end planes of axially loaded rollers. Boundary lubricated contacts lack any damping effect, and boundary lubricated contacts can act as a vibration sources.

3.3 Mixed lubrication

Mixed lubrication is characterized by boundary lubricated conditions as part of Elastohydrodynamic (EHD) or hydrodynamic lubrication [28]. In mixed lubrication, the solid-to-solid proportion of the contact can be lubricated by a solid-like or viscous-like boundary film [22].

3.4 EHD lubrication

EHD lubrication (EHL) refers to lubrication conditions based on Reynolds’ equation for hydrodynamic oil pressure build-up [24], and a combination of elastic deformation at contact surfaces of low geometric conformity, and an increase in oil viscosity under high pressure in the contacts. The slip-to-roll ratio in a rolling contact influences the oil film formation [29, 30]. Conditions of EHL is the basic assumption behind bearing life calculations [9, 20, 23, 24, 31].

In the EHL regime, the oil film thickness is slightly higher than the combined surface roughness of the rolling elements and the raceways, or $\lambda > 1$. In practice, details in the texture of the contact surfaces and the kinematics of the contact may cause large variations in the $\lambda$-value requirement for appropriate EHD lubrication [20, 29]. The minimum oil film thickness $h$ under viscous–elastic EHD conditions can be calculated using the formula [24]

$$h = 3.63 \times U^{0.68} \times G^{0.49} \times W^{-0.073} \times (1 - e^{-0.68k})$$ \hspace{1cm} (3)

where $U$ is the dimensionless speed parameter (viscosity $\times$ velocity/elasticity modulus $\times$ radius), $G$ is the dimensionless materials parameter (elasticity modulus $\times$ pressure $-$ viscosity coefficient), $W$ is the dimensionless load parameter (load/elasticity modulus $\times$ radius), and $k$ is the ellipticity parameter for the contact junction area. From formula (1), it is obvious that the oil film thickness under conditions of EHD lubrication is only slightly dependent on the load and the elasticity modulus, while the speed and oil viscosity have a strong effect [24, 32].

Micro-elastohydrodynamic lubrication means an elastic smoothening of the original surface roughness and an improvement in the EHD lubrication conditions in comparison with the nominal $\lambda$-value calculated for the unstressed surfaces [20]. The rolling elements eject some of the oil from the raceways and some of the oil is lost because of the centrifugal action. Starvation occurs in rolling bearings if the time between subsequent events of contact between rolling elements and raceways is too short for ensuring a sufficient replenishment of lubricant to support a fully flooded EHD film formation [20, 23, 33, 34]. In unfavourable cases, debris particles can agglomerate at the inlet zone of an EHD contact and cause oil inlet blockage and starvation [35]. As the severity of starvation increases, the pressure
spike at the outlet zone of the EHD contact (Fig. 6) becomes smaller, the film becomes thinner and flatter over most of the contact surface [23], the area of load-carrying oil pressure becomes smaller, and the load-carrying capacity is reduced [24]. In grease-lubricated rolling bearings, fully flooded conditions can occur in freshly lubricated bearings, while after a short time the bearing is normally running under starved conditions [19, 36].

Operation under EHL or μ-EHL conditions causes numerous dynamic stress field elements, which act as weak sources of vibrations. The thin oil film provides only weak damping properties in rolling bearings. Starvation reduces the oil film thickness, which increases the stress and vibration level.

3.5 Hydrodynamic lubrication

In hydrodynamic lubrication, sliding surfaces are completely separated by a fluid film, which is thick in comparison with the surface roughness and thicker than in EHD lubricated conditions. In rolling bearings, hydrodynamic lubrication occurs in contacts between rolling elements and their cages, and between roller ends and bearing race flanges in roller bearings with axially loaded rollers. The thick
oil film in hydrodynamically lubricated tribosystems provides a significant degree of damping of dynamic loads.

### 3.6 Influence of contaminated lubricants

Lubricants contaminated by hard, sharp particles cause abrasive contact conditions similar to (or more severe than) those of boundary lubricated contacts. Each particle that is larger than the oil film thickness can form a contact spot, which transmits part of the load and becomes surrounded by a local stress field. Lubricants contaminated by solid particles are known to cause abrasive wear both in hydrodynamically lubricated thick-film contacts [37] and EHL contacts [38–40]. Lubricant cleanliness data have been included as a parameter in rolling bearing life calculations [9].

### 4 WEAR MECHANISMS AND SURFACE QUALITY IN ROLLING BEARINGS

Several wear mechanisms [41] can be active in rolling bearings. Additionally, the surface texture can be changed by plastic deformation.

#### 4.1 Fatigue wear

Fatigue wear can occur in surfaces that are dynamically loaded. Under elastic contacts, the fatigue process usually requires high number of cycles, while under plastic contact, a low-cycle fatigue mechanism can be expected [42].

#### 4.2 Rolling contact fatigue

Dynamic contact conditions, together with possible tangential stresses and hoop stresses, and intensified by local asperity or contaminant particle contacts, cause local surface and subsurface shear stress fields, plastic deformation and dislocations, and eventually the nucleation of rolling contact fatigue cracks at defects or inclusions in the micro structure. The initial fatigue cracks are usually inclined to the surface [23, 43]. Intersection of fatigue cracks leads to pit formation and wear. The rolling contact fatigue process is usually self-propagating through surface roughening, dynamic loads, and a decrease in the load-carrying area. The formation of dislocations and cracks, and the subsequent relative motion between fractured surfaces, gives rise to vibrations in terms of AE.

The life of rolling bearings, in terms of rolling contact fatigue, can be estimated on statistical basis, from the bearing design, load, speed, and lubrication...
conditions [1, 7, 9, 23, 24]. Present calculation procedures do not consider uncontrolled factors such as electric currents, water in the lubricating oil and hydrogen embrittlement of the rolling surfaces, tribochemical wear or corrosion of the bearing, strong external vibrations during operation or standstill, or any influence of intermittent operation [9, 44–46]. In roller bearings, the roller profile and its pressure distribution can affect the bearing life [9, 47]. A stress-life method, which takes into consideration the entire stress situation in a rolling contact, has been developed in recent years [1].

Under less severe rolling contact conditions, micro-pitting can occur as the dominating surface deterioration effect [43]. Below the threshold for fatigue crack formation, surface changes in rolling contacts is limited to roll polishing by a mild plastic deformation or micro-abrasion.

4.3 Adhesive wear

Under poor lubrication, rolling bearings may suffer from adhesive wear at roller ends and in micro-slip zones. In these sliding contacts, strong adhesive junctions between surface asperities may be formed due to frictional heating, and adhesive wear may occur [14, 42].

4.4 Abrasive wear

Solid, hard particles can cause local stress peaks and shorten the life of the bearing [35, 38, 39, 48, 49]. Even particles smaller than the mean film thickness may cause abrasive wear. Furthermore, abrasive wear may take place in contacts between rough surfaces, for instance as micro-abrasion (or sliding micro-fatigue) between rolling elements and their cages. The additional solid-to-solid contact spots in the EHD or mixed lubricated contact form local stress risers [1], which under dynamic loading act as vibration sources.

4.5 Trubochemical wear

In tribochemical or corrosive wear, chemical or electrochemical reactions in the tribological contacts result in the formation of soft or brittle surface layers, which are easily removed by subsequent contact events.

5 STAGES OF WEAR AND VIBRATIONS DURING ROLLING BEARINGS SERVICE LIFE

The processes of rolling contact wear and rolling contact fatigue in rolling bearings during different stages of the operational life of the bearings are described in the following, and a schematic illustration is presented in Fig. 7.

5.1 Running-in wear of rolling bearings

Under favourable running-in conditions, a ‘roll polishing’ effect occurs in rolling bearings, mainly through plastic deformation of surface asperities. The roll polishing effect gradually disappears as the surface roughness decreases and a transition from mixed lubrication into EHD lubrication conditions occurs. The running-in process can be re-activated by changed operational conditions [24, 25, 29, 43, 50]. Each incident of plastic deformation of a surface asperity during running-in forms a dynamic stress field and a vibration source.

Micro-abrasive and sliding fatigue wear in the micro-slip regions (Fig. 2) of rolling contacts are responsible for part of the surface alterations, for rolling contact wear, and for wear particle formation in the running-in of rolling bearings [51]. The plastic deformation caused by abrasion and the micro cracking caused by fatigue are weak sources of vibrations.

Abrasive wear and surface indentations are caused by particles in the oil during the running-in stage of the bearing [52]. Abrasive wear by contaminant particles during the running-in stage of a bearing usually roughens the rolling surfaces and produces more wear particles, and therefore weakens the preconditions for subsequent operation of the bearing [25]. Crushing of particles, abrasion by particles, and contact dynamics due to surface roughening contribute as sources of vibration in a rolling contact.

Under running-in at stress levels leading to local rolling contact fatigue at areas of asperity interaction, surface roughening by micropitting can eventually occur. The small depth of the micropits gives a low volume loss, although an abundance of wear debris of small size. Micropitting increases the vibration level of the bearing, through contact dynamics, due to the surface roughening and the increasing contents of wear particles in the contacts.

Depending on the rolling conditions during running-in and the initial smoothness of the surfaces, the surfaces smoothness can be either enhanced or decreased. The wide band root mean square (RMS) values of the vibration acceleration and the noise generated directly from the surface asperities decrease if the surface asperities are smoothened during running-in. Whether this is detectable or not depends on the background noise and other vibration sources of the bearing. In addition to the changes in the RMS values, the frequency content of vibrations and noise originating from the surface roughness can change during running-in.

5.2 Wear and fatigue in rolling bearings under steady-state conditions

Most of the operational lifetime of a rolling bearing is supposed to occur under steady-state conditions
after the running-in stage. The run-in surfaces are the smoothest and the vibrations are at their lowest level. The operating conditions during the steady-state phase will finally determine whether the bearing will be subjected to rolling contact wear only or if the contact mechanism will develop into a process of rolling contact fatigue (Fig. 7).

As in the running-in stage, micro-slip on part of the rolling contact is responsible for part of the abrasive wear and sliding fatigue wear as well as part of the wear debris formation in steady-state operation [51].

Abrasive wear and surface indentations can roughen the rolling surfaces and shorten the bearing life [1, 35, 38-40, 48]. Crushing of particles, abrasion by particles, and increased contact dynamics due to surface roughening contribute as sources of vibration under steady-state operation, usually with increasing intensity.

When a bearing operates above fatigue threshold conditions, the bearing will be subjected to rolling contact fatigue cracking. The dislocation and crack formation and the subsequent relative motion between fractured surfaces give rise to vibrations in terms of AE from the surfaces subjected to the fatigue process.

When the calculated rating life for a bearing has been reached, the probability for a first occurrence of pitting is 10 per cent (factor $a_1 = 1$), by definition, although the likelihood for fatigue pitting in practice is reduced by built-in safety margins in the calculation procedure. Upon the onset of pitting, the steady-state of the bearing operation is terminated and the bearing moves into a stage of accelerated fatigue wear.

### 5.3 Accelerated rolling contact fatigue and rolling bearing failures

Choi and Liu [53] have divided the last stage of the rolling contact fatigue process of a bearing into two periods. The first period shows no significant change in the vibration amplitude, as the crack initiation and propagation occur below the surface. The second period shows a significant increase in the vibration amplitude, due to the formation and progress of a spall on the surface.

![Fig. 7 Upper image: rolling bearing wear volume accumulation from running-in towards major rolling contact wear or towards rolling contact fatigue spall formation. Lower image: the corresponding evolution from running-in to rolling contact fatigue failure at rolling bearing raceway surface and subsurface (inserted photograph: rolling contact fatigue failure in roller bearing)](image-url)
Once initiated and developed into a first local pitting failure, the running conditions of the bearing evolve, through the formation of fatigue wear particles, surface roughening, and dynamic loads, into a stage of progressive pitting and failure development. The acceleration of the fatigue and pitting process is accompanied by an increase in the vibration level, rate of wear particle formation, and wear particle size [54].

The surface roughening increases the tangential forces on the rolling elements in the bearing. The increase in the vibration level, surface roughness, and tangential forces in the bearing can lead to secondary damages such as raceway fractures, typically originating from fatigue pits, and roller cage breakages. In the most severe scenario, cage breakages block the motion of the rolling elements.

6 ROLLING BEARING MONITORING AND DIAGNOSTICS

6.1 Vibration analysis

Analysis of vibrations from rolling bearings is typically based on the results of vibration acceleration and AE measurements, which are analysed in the time domain or in the frequency domain, or in both. Prior to sampling, the vibration signals are anti-alias filtered for preventing frequencies higher than half the sampling frequency so as to appear at the lower frequencies.

Analyses in the frequency domain are based on spectrum analyses, which present the vibration data as a function of discrete frequency components. The spectrum is calculated by using fast Fourier transformation. In the simplest form of vibration analyses, the RMS values are monitored. In practice, further processing, such as signal averaging, filtering, envelope calculation, calculation of amplitudes of selected frequency components, cepstrum calculations, synchronized rotation vector and profile monitoring, etc., is needed. After these actions, the characteristic features (i.e. the vibration parameters are calculated). The number of different parameters, methods, practices, and their combinations in vibration analysis is unlimited. In the following, the use of the most common methods and parameters are shortly reviewed in relation to certain factors interfering with the operation of rolling bearings.

6.1.1 Defects in rolling surfaces

When a rolling element passes a local fatigue crack, pit or spalled area, or some other irregularities on the race surface, an impulse occurs and this excites vibrations with frequencies governed by the resonances of the structure. A defect in any of the active rolling surfaces periodically produces periods of vibrations due to the impulses from the over-rolling of the surface defects. The periods of vibrations are repeated at certain intervals, the frequencies of which depend on the bearing geometry features, the rotational velocity, and the location of the defect or defects [55–57]. Figure 8 shows the theoretical generation of cyclically repeated periods of vibrations due to bearing defects, and their envelope curves under constant load.

Envelop curves are in this context visualizations of the dynamics within vibrations originating from point defects on rolling surfaces. The envelope curves are determined from vibration acceleration data in the time domain, which are filtered using a band pass filter in the expected natural bearing frequency range, which typically lies between 500 and 3000 Hz. The purpose of the filtering is to remove the background noise around the natural frequencies, which are modulated due to periodic impacts from point defects at the rolling contacts. After performing a Hilbert transformation of the filtered vibration acceleration data, an envelope signal is obtained. Finally, the spectrum of the envelope signal is analysed in order to determine the characteristic, statistically relevant frequency of the cyclic variations of the envelope curve.

In practical rolling bearing applications, the frequencies of the cyclic impacts vary in a random manner, due to slip, varying speeds, and varying load angle [56, 58, 59]. Consequently, the inconsistency in the cyclic impacts may complicate the detection of the nominal bearing defect frequencies from the raw vibration spectrum. However, the envelope spectrum or its square can be successfully used for indicating the defect frequencies even with random slip [59, 60].

The methods for the determination of vibration parameters representing the rolling surface defects in bearings have obtained particular attention in recent years.

Kiral and Karagülle [61, 62] used artificial dynamic loading models for healthy and defected bearings, for analysing vibration responses of the rolling bearing structure by a finite-element package. Band-pass filtering of the vibration acceleration increased the ratio of the RMS value, the crest factor, and the Kurtosis value.

Heng and Nor [63] have demonstrated that statistical parameters are sensitive to the rotational velocity of the bearing. The authors suggested that plots showing the crest factor against the Kurtosis value could be used for classifying defects at the rolling surfaces of rolling elements, outer races, and inner races.

According to the results of experiments with faulty outer races of rolling element bearings, by Abdullah and co-workers [64], the statistical RMS parameter and the maximum values of the vibration acceleration and the AE measurements increase with the size of the defect, but not linearly with the bearing speed and load. By a similar study, Li and co-workers [65] showed that a suitable frequency band for the vibration acceleration RMS was 3–5 kHz.
Fig. 8 Rolling bearing defect vibration and envelope curves for a radial rolling bearing under a fixed radial load, a constant rotational velocity of the inner race, and a static outer race. The abbreviations BPFO (ball pass frequency outer race) and BPFI (ball pass frequency inner race) refer to the peak frequencies of the envelope curve caused by inner and outer race defects, respectively. The abbreviations BSF (ball spin frequency) and FTF (fundamental train frequency) refer to the peak frequency of the envelope curve caused by a rolling element defect or a cage defect, respectively. RPM stands for the rotational velocity.

In a study by Orhan et al. [57], a spectral analysis of the vibration acceleration was successfully used for detecting ball bearing looseness and defects in ball bearings and cylindrical roller bearings. The authors observed that when a rolling surface defect reached an advanced stage, amplitudes of high frequency components disappeared due to self-hammering of sharp surface flaws originating from the defects.

Khemili and Chouchane [66] used an adaptive noise cancellation technique to increase the signal-to-noise ratio of a vibration acceleration signal from rolling element bearings with artificially caused defects at the outer and inner races. Li and Li [67] showed that both the short-time energy strength and zero-crossing rates of the AE signals increases upon an evolution from healthy rolling bearings to ones with damages on the outer races and rollers. Using a rolling bearing test rig, Al-Dossary et al. [68] have shown that the transient AE burst duration correlated with the size of a defect at the outer raceway.

Choy et al. [69] studied damage identification on artificially caused damages on the inner race and a ball of a ball bearing in a constantly loaded bearing test rig. The change from a healthy bearing into a faulty one was clearly detected with both the traditional vibration frequency analysis methods and a Poincare map. While no significant change was recognized in the frequency pattern with an increasing degree of bearing damage, the Poincare map covering more than 2000 shaft revolutions provided an accurate quantification of the damage level and a reliable indication of the location of the damage in the bearing.

By performing test rig studies on rolling element bearings, McFadden and Toozhy [70] showed that synchronous averaging of vibration acceleration could be used for determining the location of the inner race fault distribution as well as the variations within the set of rolling elements.

Ericsson et al. [71] analysed vibration acceleration signals from 103 measurements carried out at laboratories and industrial companies, on bearings of which the true condition was known. As a result of the analysis, the wavelet, the envelope, and the synchronous time domain averaging, each having an error rate of about 10 per cent, were considered the best methods for vibration data analysis. For extremely slowly rotating rolling bearings with an outer race defect, the synchronous time domain averaging was found to be the best method, while the other methods studied were found unreliable.

At low speeds, the band energy of the envelope spectrum has shown to be a powerful tool for outer and inner race defects but not for rolling element defects [61, 62]. Jamaludin and Mba [72, 73] analysed...
the AE signals from tests with extremely slowly rotating rolling bearings with and without artificially caused defects on the outer race, on the inner race, and on a roller, and concluded that the faulty bearing surfaces were identified by signatures of modelled autoregressive coefficients of the AE.

Shiroishi et al. [74] compared healthy bearings with ones having artificially produced raceway defects, by using the envelope method and an adaptive line enhancer method for reducing the wide-band noise of the envelope signal. It was shown that the wide-band noise reduction made the defect amplitude more visible. Vibration acceleration measurements were found to be more sensitive to the fault types studied than were the AE measurements, especially with an inner race fault, in which case there are more interfaces effectively attenuating the high-frequency vibrations. Williams et al. [75] obtained similar results when running new rolling bearings to failure.

In a review study by Tandon and Choudhury [56] about rolling bearing fault detection, it is suggested that out of the time domain statistical parameters, the Kurtosis value is the most effective one for defect identification. At the initial stage of bearing failure, high-frequency techniques such as the vibration acceleration envelope analysis, the AE measurements, and the shock pulse method (SPM) have been successfully applied.

By applying vibration analysis methods and a test rig, Tandon and Kumar [76] have studied bearings with outer races having artificially caused defects. At high rotational speed, the SPM gave a better sensitivity in comparison with the RMS value of the vibration velocity.

Tandon et al. [77] found that, with different load and defect sizes, the peak amplitude of the AE showed the largest increase when the bearing was changed from a healthy one to a faulty one. The SPM gave the second-best response, while the RMS of the vibration velocity showed the lowest response in this comparison. Zhen et al. [78] studied the condition monitoring of healthy and faulty rolling bearings, and concluded that the SPM method is effective in estimating the rolling bearing condition.

Rubini and Meneghetti [79] compared the vibration acceleration spectrum, the envelope and the wavelet analysis methods on a double-row ball bearing with artificial raceway defects, and found that an inner race defect could be detected with the spectrum analysis method, and that an outer race defect could be detected by the same method if the loading was increased. Both the envelope analysis method and the wavelet analysis method were successfully used for the determination of all the fault types. While the edges of the cracks were smoothed during repeated over-rollings by rolling elements, the amplitude of not only the envelope analysis but also the wavelet analysis results decreased.

Applying statistical, spectrum, and wavelet analysis methods, Prabhakar et al. [80] have investigated ball bearings with artificial inner race and outer race defects. The RMS and Kurtosis results increased when the bearing was changed from a healthy one to one with defects. The vibration acceleration spectrum showed an outer race defect, but an inner race defect remained invisible due to some unknown change in the measurement practice. By wavelet analysis, it was possible to detect all the fault types and the wavelet analysis method was effective in detecting both single and multiple defects in a bearing.

As indicated by the above examples, the number of applications and applied methods, results, and resulting analyses is large. So far, no method of vibration analysis has been identified as universally superior to all the others, for which reason each new application will require an evaluation of pre-selected candidate methods prior to the selection of the most appropriate one.

6.1.2 Starvation

At a high product of rolling velocity and viscosity, the oil film thickness shows a tendency to level off and even decrease with increasing velocity [33]. The oil film thinning due to the starvation generates quick pressure pulses, which excite nominal frequencies of the bearing structure. Using a bearing test installation, Mba [81] has demonstrated that high-frequency vibration, or the AE energy, correlates to the calculated film thickness in a rolling element bearing. According to Miettinen et al. [82], a high base oil viscosity in a lubricating grease leads to increased starvation, as indicated by increased AE pulse count rates.

6.1.3 Contaminated lubricants

The condition monitoring of rolling bearings with lubricants containing quartz (SiO₂) dust has been studied by Maru et al. [83]. Tests with clean oil revealed that the RMS value at the high-frequency (600–10 000 Hz) band was less sensitive to the load and speed than at the other frequencies. The RMS values increased with the contamination particle concentration up to a stabilization limit. With increasing size of the contamination particles in the tests, the vibration level first increased until an upper particle size limit was reached, after which the vibration level decreased, probably due to a restricted entry of larger particles into a contact with certain oil film thickness, or due to particle settling by gravity.

Furthermore, Maru et al. [84] compared high-frequency RMS values of the vibrations from rolling bearings lubricated with contaminated oil. With clean oil, the high-frequency RMS values were higher for the roller bearings that were studied than for the ball bearings studied, due to a lower lambda value for the roller
bearings. However, the increase ratio for the high-frequency RMS values, when changing from clean oil to contaminated oil, were higher for the ball bearings than for the roller bearings.

Akagaki et al. [85] analysed vibration acceleration in a bearing test rig with contaminated oils, and observed that the shape of the probability density function changed, depending on the increased particle size of the white fused alumina contaminant. When increasing the contaminant particle size, the wear rate of the bearing increased to a certain limit.

6.1.4 Grease lubrication

In comparison with oil lubrication, the use of lubricating greases may increase or reduce the noise in measurement signals for rolling bearings. The noise level of grease-lubricated deep-groove ball bearings rises with the bearing diameter [86]. Work by Howard [87] showed that, in comparison with a light mineral oil, all lubricating greases considered as clean gave higher RMS values of the velocity of the bearing-structure-borne vibration at a high-frequency band, from 1.8 to 10 kHz, while at a lower frequency band (50–300 Hz), the results were quite similar for greases and for oil.

Rolling bearing tests by Wunsch [86] on the noise characteristics of lubricating greases showed that the vibration peak level decreases with an increasing base oil viscosity. Solid lubricants, such as graphite, calcium carbonate, etc., increased as well the vibration peak level of the original bearing. Similarly, bentonite-clay-thickened and polyurea-thickened greases are expected to be very noisy [88].

Work by Komiya [99] on the effect of solid contamination on the noise of grease-lubricated ball bearings showed that the vibration acceleration noise increased with the increase of aluminium oxide particle size with a constant concentration of 1000 ppm. The noise increase tends to occur in the early stage of rotation and tends to decrease when the rotation time is prolonged, which seems to be in good correlation with the corresponding oil-lubricated cases. The particles of the largest size produced a lower vibration value, which is probably due to the smaller number of particles for the same mass percentage. A 10 ppm contaminant particle concentration, regardless of the particle size, had no effect on the RMS values of the vibration acceleration, but increased the number of vibration acceleration pulses exceeding the threshold value. In general, the lower the base oil viscosity, the greater the pulse count of the vibration acceleration.

Miettinen and Andersson [90] investigated the influence of solid contaminant particles as the cause for vibrations in grease-lubricated, statically loaded deep-groove ball bearings. The measured AE levels increased non-linearly with the particle concentration. Small-sized particles generated a higher AE pulse count than did particles with a larger size. The cleaning and re-greasing of a bearing that had been tested with contaminated grease gave a reduction in the AE level, although the AE level remained higher than it was when the bearing was new and lubricated with clean grease.

6.2 Wear particle monitoring

Several methods for the detection and analysis of wear particles in rolling bearings are available. The methods can be divided into off-line laboratory methods and continuous on-line or in-line methods. Hunt [52], and Roylance and Hunt [91], have listed most of the wear particle detection techniques for in-line or on-line analyses. In the following, a selected set of methods for wear particle monitoring are presented and discussed.

6.2.1 Composition of particles

Inductively coupled plasma optical emission spectrometry can be used for elementary analysis and detection of different wear metals and their concentration in wear particles [92]. Atomic emission spectroscopy is suitable for an elementary analysis of a large number of elements. X-ray fluorescence is used for classifying a large number of elements from a sample, which is excited by bombarding with high-energy X-rays. The energy dispersive X-ray spectrometry feature of a scanning electron microscope can be used for identifying wear elements as well as for wear particle sizing and characterization. For particles of more than 1 μm in diameter, an optical light microscope can be used for analysing the visible features of wear particles. In ferrography, ferrous particles are identified on the basis of their sedimentation on a glass sample plate by the combined effect of magnetic attraction and gravity.

6.2.2 Size, shape, and concentration

In the traditional filter blockage technique, a flow of oil with particles passes through a fine mesh of a known pore size, and the change in the flow characteristics indicates how many pores are blocked by particles larger than the pore size [52, 91]. In inductance-based detection, a coil is placed around a pipe with a flow of oil with particles, and metal particles are detected through an inductance change in the coil, which is different for ferrous and non-ferrous particles [52, 91, 93]. Magnetic attraction of wear particles in oil can be used for sensing a particle flow past a magnetic sensor [52, 91, 93].

In ferrography analysis, oil-borne particles are sedimented along a glass slide by the interaction of gravity and magnetic attraction. Based on the location and orientation, form, size, texture, and colour, wear particles and other particles can be identified under a light microscope [52, 91, 93].
For the optical determination of the size distribution of particles in oil, one option is the Fraunhofer light diffraction (also known as forward scatter) technique, by which an array of detectors ahead of a light beam, passed through a mixture of oil and particles, detects the proportion of light diffraction caused by small and large particles [52, 91].

With the optical obscuration method, each particle casts a shadow when passing through a light beam, and a photo detector measures the drop in intensity at a reference surface [52, 91, 94]. According to Roylance and Hunt [91], time-of-transition is another optical technique. As a rotating, focused light beam passes some particles, the time-of-transition is measured and it is proportional to the particle size.

When a flow of oil and particles is directed along a surface coated with a thin conducting layer, the electrical resistance across the thin film is increased as the coating becomes thinner due to erosion by the particles. An ultrasonic beam, which is focused into the oil, is reflected from particles in the oil, and the reflected signal, which depends on the size of the discontinuity, is recorded.

The use of wear particle examination as a diagnostic tool requires an estimate of the likely wear mechanisms and the regimes in which each of the contacts of interest operates [41]. Condition monitoring of an oil-lubricated ball bearing in an accelerated bearing life test was studied by Halme [95], who concluded that during a step from the steady state to the final phase of the life of a ball bearing, the most remarkable acceleration in a wear process was detected through the measured vibration acceleration responses, as well as through the amount of relatively large particles in the oil.

Using a bearing test rig, Harvey et al. [96] studied the wear monitoring of a lubricated taper roller bearing with an artificial fault on the inner race. An accelerated test was run until a total damage occurred, and the wear particle responses were measured with electrostatic sensors and the vibrations were measured with acceleration sensors. When about 10 per cent of the calculated lifetime of the bearing was left, a first steady increase in the measured electrostatic responses was seen. When about 6 per cent of the lifetime was left, both the electrostatic response and the vibration amplitude of the inner race fault frequency increased. When about 99 per cent of the calculated lifetime of the bearing had elapsed, a sharp increase occurred in all the measured parameters.

6.3 Combining calculated rating life and diagnostics information for refined life expectancy estimates

Because of the uncertainty in the determination of the rating life of rolling bearings, a link between the probabilistic life estimations performed by calculations and the recorded and interpreted condition monitoring data, the machine operation statistics, and the maintenance history should be established [97]. For example, Wang and Zhang [98] have reported the results of combining oil monitoring information, the time of operation, and the failure statistics of 30 aircraft engines for predicting the residual life of these engines.

In terms of the reliability theory, the most widely used distribution type calculated for the lifetime of machine components is the Weibull distribution [4, 98, 99]. On the other hand, according to the least sophisticated approach, the operational time is only compared to the recorded lifetime of similar components in similar operations as a mean residual life estimate, which is not very reliable for an individual rolling bearing in practice [100].

A supervised back-propagation method has been successfully tested for the diagnostics of rolling bearing defects by using vibration signatures as input values to a neural network [101–103]. In addition, a successful investigation with an unsupervised adaptive resonance method was reported [102]. Rolling bearing degradation tests were conducted by using vibration envelope features as input vectors for unsupervised self-organizing maps (SOM) [104]. It was shown that by comparing the minimum quantization error of new data to the SOM trained with healthy data, the results were more sensitive for the identification of incipient bearing failure than the statistical vibration acceleration RMS parameter would have been. A supervised back-propagation neural network has been evaluated for the determination of the degradation time and finally the weight application technique for residual life prediction for ball bearings [105]. The results reported showed a better prediction in comparison with the traditional $L_{10}$ calculation technique. However, despite the reported promising results of rolling bearing diagnostics obtained with different kinds of neural networks, it should be noticed that in most cases, the studies have been carried out in a laboratory environment where the data are sufficient in terms of the amount and one-dimensional in terms of the application. In many real industrial applications, the situation is totally different and the generalization of the results and neural network models may not be justified. Furthermore, this limitation in the applicability applies somewhat to rolling bearing diagnostics based on fuzzy logics. As an example, it is reported that, by using vibration features in a test rig, the fuzzy reasoning gave 100 per cent accuracy in distinguishing between healthy rolling bearings and faulty ones [106].

For an initial prediction of the rating life for a rolling bearing, the statistical methods like the ISO 281 calculation procedure are the most reliable ones. During the operation of the bearing, the prognosis for the remaining operational lifetime can be updated and
ensured by employing a combination of the calculated rating life estimate and condition monitoring models that utilize the progress or trend in the recorded and interpreted condition monitoring parameters. The process for updating the lifetime prognosis and the order in which the condition monitoring methods are applied is important; in vibration analyses of rolling bearings, the initiation of sub-surface cracks and minor metal-to-metal contacts usually generate high-frequency vibrations, which can be monitored with AE sensors. When the first sub-surface cracks reach the surfaces, or when surface cracks are initiated, they have sharp edges which cause impacts between the rolling surfaces of the bearings and excite the nominal frequencies of the bearing, or vibrations of typically 500–3000 Hz frequency. The nominal frequencies can be detected by using the vibration acceleration envelope analysis described in section 6.1. Finally, when the sharp edges are flattened and wider cracks and pits are generated, the nominal bearing defect frequencies such as BPFO, BPFI, BSF, or FTF (see section 6.1) become detectable from the vibration velocity spectrum. The high overall vibrations can be measured when the bearing is approaching its final failure.

7 CONCLUSIONS

Friction, wear, and lubrication mechanisms acting at the rolling and sliding contacts are central issues for the operation of rolling contact bearings, and considered in rating life calculations for bearings at specific operational conditions.

During bearing operation, rolling surface roughening and wear particle formation may eventually occur as a consequence of rolling contact wear and rolling contact fatigue, unless the operational conditions are mild enough to limit the surface alterations to polishing.

Rolling contact fatigue, in particular, causes changes in the bearing kinematics by rolling surface roughening and wear particle formation, which in turn increase the bearing vibrations. The vibrations and the wear particle formation can be monitored during the operation of the bearing.

Useful methods for vibration-based condition monitoring of rolling bearings are the vibration acceleration high-frequency techniques, the SPM, the envelope analysis, and the AE analysis for obtaining an early warning. For the detection of a severe bearing defect, the most useful methods are the spectrum analysis of the vibration velocity amplitudes at nominal bearing frequencies, the determination of the RMS value of the vibration velocity, and the analysis of the statistical parameters of the vibration acceleration such as the Kurtosis and RMS values. Measured and interpreted vibration and wear particle data from condition monitoring measurements can be combined with a theoretical rating life estimate, in order to give an updated and more reliable prognosis for the remaining operational lifetime of the bearing.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support from the European Commission, the Sixth Framework programme for Research and Technological Development, under which the present work was carried out as part of Work Package FP6 of the Integrated Project IP017498 DYNAMITE ‘Dynamic Decisions in Maintenance’. Research Professor Kenneth Holmberg of VIT is greatly acknowledged for proposing the present study to be carried out and for encouraging discussions during the progress of the work.

© Authors 2010

REFERENCES

33 Chiu, Y. An analysis and prediction of lubricant film starvation in rolling contact systems. ASLE Trans., 1974, 17, 22–35.
87 Howard, P. Noise testing of rolling element bearings. NLGI Spokesman, 1975, 39, 54–60.
88 Ward, C. Practical aspects of grease noise testing. NLGI Spokesman, 1994, 58, 8–12.