A novel method for diagnosing rolling bearing surface damage by electric impedance analysis

G. Martin, F. Becker, T. Schirra, E. Kirchner
Technische Universität Darmstadt, Institute for Product Development and Machine Elements
Otto-Berndt-Straße 2, 64287 Darmstadt, Germany

Abstract
This paper presents a novel condition monitoring method for rolling bearings, based on measuring the electric bearing impedance. The method can diagnose the presence of damage by frequency-domain analysis, and its extension along the raceway by time-domain analysis. The latter enables the assessment of the severity and the progression of bearing damage. A fatigue test shows that the occurrence of pittings in the bearing raceways causes characteristic peaks in the impedance signal, and that the duration of the peaks increases during damage progression. A second test series with artificial damage shows that the duration of the peaks depends on the bearing load and the length of the damage along the raceway and confirms the explanation hypothesis.

1 Introduction
Predictive maintenance is a centerpiece of the digitization of mechanical engineering. According to RANDALL, it offers the possibility of far-reaching cost savings compared with the traditional maintenance strategies of corrective or time-based preventive maintenance [1].

Rolling bearings are machine elements that play an important role here, since they are used in almost every machine and are directly in the flow of forces. For example, approx. 37% of all damage in machine tool spindles is rolling bearing damage [2]. Condition monitoring of rolling bearings has been the subject of research for decades and has produced a large number of measurement procedures and evaluation methods [1]. Vibration monitoring is still the dominant method in practice [3].

In this paper, an alternative measurement method to detect bearing damage is investigated, in particular pitting damage. This method takes advantage of the inherent electrical impedance of rolling bearings. The electrical impedance of rolling and journal bearings has found some applications as a sensor technology in the past years [4–8]. The fact that no additional sensor technology has to be integrated in the machine element facilitates the integration of the sensory function without requiring additional installation space, so that the user does not have to make any changes to the design [9–11].

1.1 Electric impedance of rolling bearings
The electrical properties of elastohydrodynamic (EHL) contacts depend on the lubrication state [12,13]. GEMEINDER distinguishes three states depending on the lubrication film thickness [13],
which are shown in Figure 1. In the case of metallic contact (boundary friction and mixed friction), the contact is described as an ohmic resistance $R_p$. If the lubricating film separates the contact partners completely, two states can occur: If the electric field strength exceeds the breakdown field strength of the lubricant, a so-called breakdown occurs, often called *EDM current* due to its similarity with the manufacturing process electric discharge machining. In this state, the contact is modeled as a capacitance $C_p$, with the resistance of the discharge channel $R_{p,EDM}$ connected in parallel. If the electric field strength is below the breakdown field strength, the contact is also modeled as a resistor $R_{p,EHL}$ and a capacitor $C_p$ connected in parallel, but the resistance of the lubrication film is much larger than the resistance during breakdown. In the case of a thin lubricating film, the two states can occur in rapid alternation, depending on the local proximity of surface asperities; this is modeled in Figure 1 as a switch.

![Electric model of the EHL contact as a function of the lubrication film thickness](image)

In the case of a sufficiently thick lubrication film, the contact can be modeled as a plate capacitor, whose plate thickness is the lubrication film thickness and whose plate area is the Hertzian area, cf. Figure 2. Film thickness and Hertzian area, and thus also the capacitance, depend on the load. In this study, the complex impedance is measured, which reflects the entire electric behavior of the bearing, including resistive and capacitive terms.

![Electric model of the EHL contact in a ball bearing.](image)

### 1.2 Effect of surface properties on the electric impedance of rolling bearings

This paper deals with the detection of surface damage by measuring the bearing impedance, i.e. it investigates the relationship between surface properties and the bearing impedance.
Therefore, this paragraph summarizes the current state of research in the relationship between surface properties, such as surface roughness and damage, and the electric properties of rolling bearings and EHL contacts in general.

TEN NAPEL and BOSMA measure the contact capacitance and show that the capacitance is greater with rough surfaces than with a smooth surface, and attribute this to a geometric effect, according to which the capacitance of a plate capacitor with a rough surface is greater than with a smooth plate surface [14]. SCHMIDT arrives at similar results in extensive experimental investigations on a two-disc test rig [15], but, in contrast to TEN NAPEL and BOSMA, he attributes this to an actual reduction in lubrication film thickness.

Resistance measurement has been used extensively to determine the occurrence of metallic contacts. Based on the approach that the resistance of a lubricating film is orders of magnitude higher than the resistance at metallic contact of surface asperities, FUREY developed a method to detect and measure the incidence of metallic contacts [16]. FUREY used the mean value of the resistance as a measure of the time fraction of metallic contacts. TALLIANN et al. further refined this method by counting the number of metallic contacts [17]. HEEMSKERK et al. use the same approach to calculate the percentage time of metallic contact, which is referred to as the PCT value [18].

Important for this work, which investigates the effect of surface properties on the electric impedance, are some results on the behavior of the resistance and capacitance during run-in and before failure. Since surface asperities are smoothed during run-in, run-in corresponds to a change in surface properties. LORD and LARSSON measured the change in resistance and PCT value during run-in for different surface treatments. In all cases, the PCT value decreases, and the mean value of the resistance increases. It can be deduced that surface roughness influences the resistance of EHL contacts. TUOMAS and ISAKSSON performed a fatigue test, while measuring capacitance and detecting contacts by the resistance method [19]. The absolute value of the capacitance does not show any changes before failure, but the incidence of metallic contacts, which had initially dropped to zero after running-in, increases before failure, cf. Figure 3. The authors do not provide any further information on the type of bearing damage that occurred, and furthermore only one fatigue test is presented. A generalizable rule can therefore not be derived from this test. These results point to a relationship between the electric properties of the bearing and incipient bearing damage, but they also show the need for further research to gain a better understanding of the relationship between bearing damage and bearing impedance.
In summary, there is limited and isolated knowledge about the actual relationship between surface properties and impedance. Therefore, there is a need for a systematic study on the suitability of rolling bearing impedance for the investigation of surface damage in rolling bearings. The goal of this study is to investigate how the impedance of rolling bearings is influenced by surface damage. To this end, two series of test were conducted. In a fatigue test described in sections 2.2 and 4.1, five angular contact ball bearings were run from new condition to failure, and during damage progression, in order to detect anomalies in the impedance signal in damaged condition. In a second test series described in sections 2.3 and 4.2, artificial damage was introduced into undamaged bearings by laser structuring, in order to investigate the relationship between surface deviations and the impedance signal under controlled conditions.

2 Materials and Methods
This section describes the test setup and the test procedures of the two test series.

2.1 Test rig and impedance measurement
The tests described in this paper are conducted on the ATHENE bearing test rig at the institute for product development and machine elements pmd. Figure 4 shows a cross-section of one test station of this test rig. All bearing seats consist of two parts, separated by an electrically isolating layer. The inner part of the bearing seats can be contacted by an electric probe. Not shown in Figure 4 are the oil system with oil pump, oil reservoir, oil temperature control and oil temperature sensor, the bearing temperature sensors, the hydraulic cylinders for radial and axial load, the drive motor and the accelerometers mounted axially and radially. The drive motor is connected to the test rig shaft by a jaw coupling. The compensating elements of the coupling are made of an electrically insulating elastomer; therefore, the test rig shaft is electrically insulated from the drive shaft.
The bearings whose electrical properties are measured are called test bearings, and all other bearings are called support bearings.

Figure 4. Test station of the bearing test rig

The test setup is shown schematically in Figure 5. A total of four rolling bearings are mounted on the shaft. The hydraulic cylinders load the bearings with forces $F_a$ and $F_r$. The radial force is introduced via the common bearing seat of the inner bearings and is supported by the outer bearings. The radial load of each bearing, denoted $F_r$, is thus 50% of the introduced force $F_{r,z}$. The axial load $F_a$ is introduced via the bearing on the cylinder side and supported by the other outer bearing. Figure 5 also shows the test rig components that provide a defined electrical path through the test bearing. The shaft can be contacted by a slip ring and the bearing seats can be contacted by contact pins.

The electrical setup used is a modified version of the test setup described in [20]. For the tests with artificial damage, three bearings with ceramic balls and one bearing with steel balls were used, in order to eliminate any parasitic effects of the support bearings, cf. Figure 5. A sinusoidal voltage of 2.5 MHz was applied to the shaft and the current through the bearing was measured. The setup for the fatigue tests uses two test bearings with steel balls, and two support bearings with ceramic balls, as shown in Figure 6.

Figure 5. Schematic representation of the test stand in the setup for tests with artificial damage.
2.2 Design and procedure of the fatigue tests

The fatigue tests were conducted in two phases, schematically shown in Figure 7. Five angular contact ball bearings (subsequently named bearing #1-5) of the type FAG 7205B-XL-TVP were used. In the first phase, these each of these bearings was run from new condition to failure. In the second phase, the tests were continued for another 30 minutes in order to let the damage progress. The lubricant was FVA reference oil III A.

For the first phase, two test bearings are initially installed in new condition and the test is carried out until an increased vibration level indicates a bearing damage (initial damage). During this stage, an impedance measurement is performed every minute. Based on the impedance data, the failed bearing can already be identified, so that only the damaged bearing is removed and replaced by a new bearing. The test is now continued until the next bearing failure. During this first phase, the radial load is $F_r = 3,000$ N, the axial load is $F_a = 28,000$ N (38,000 N in the case of bearing #1), the speed is $n = 4,000$ min$^{-1}$ and the oil temperature is $T = 30^\circ$C.

After the initial damage, the failed bearings are disassembled, degreased and the geometry of the damage is measured. For this purpose, an Alicona InfiniteFocus reflected-light microscope is used, which can determine 3D topographies of surfaces by focus variation. For the second phase of the experiment, damage progression, the failed bearings are installed again in the test rig and the impedance is measured for another 30 min. During this time, the impedance measurement is performed approximately every 20 seconds. After that, the bearings are disassembled again, and the damage geometry is measured again. Figure 8 shows an example of the resulting pittings after the initial damage has occurred, and after damage progression. During this phase, the speed is reduced to $n = 2,000$ min$^{-1}$ in order to obtain a slower damage progression, while the other operating conditions remain unchanged.
2.3 Design and procedure of the tests with artificial damage

For this test series, artificial surface structures are introduced into ball bearings to produce damage that is comparable to real pittings in terms of geometry. The structures are produced by laser ablation with a LaseBox system from the manufacturer Östling Markiersysteme. An example cross-section of the resulting geometry is shown in Figure 9. Here, a line-shaped geometry was modeled so that this figure represents the smallest geometry that can be generated. It is noticeable that the lowest point does not occur in the center of the structure, but off-center at two points.

The bearings used were of the type SKF 6205 C3 TN9 and the lubricant was FVA reference oil IIIA. Each bearing was first conditioned by a run-in procedure to flatten surface asperities. After disassembly and degreasing, the laser structures were introduced and the bearing was reassembled.

Both the geometry of the artificial surface structures and the operating conditions were varied, in order to enable a complete study of the relationship between the surface geometry, the
operating conditions and the impedance. A two-level full factorial test design with six factors was used: Three factors represent the operating conditions radial bearing load $F_r$, oil temperature $T$ and rotation speed $n$, whose values are shown in Table 1.

<table>
<thead>
<tr>
<th>Factor level</th>
<th>Radial load $F_r$ in N</th>
<th>Oil temperature $T$ in °C</th>
<th>Rotational speed $n$ in min$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta$</td>
<td>750</td>
<td>40</td>
<td>1,250</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>3,500</td>
<td>70</td>
<td>4,000</td>
</tr>
</tbody>
</table>

Three factors describe the surface structures. These structures are intended to be representative of real-world rolling bearing damage, especially pitting. The pittings that occurred in the fatigue tests have an irregular, shell-shaped geometry. To keep the number of geometry parameters low, the structural geometry is reduced to the three variation parameters length $l_p$ (in rolling direction), width $w_p$ (in axial direction) and depth $t_p$ (perpendicular to the raceway surface). From these three parameters results the structure, which is schematically illustrated in Figure 10. The choice of the factor levels is based on the characteristic dimensions of the EHL contact. Therefore, the high levels ($\Theta$) of the length $l_p$ and width $w_p$ are set to values slightly lower than the respective dimensions of the Hertzian area at the high level ($\Theta$) of the radial load $F_r$.

Figure 10. Definition of the parameters describing the artificial surface structures

The depth of the structure $t_p$ is determined based on the lubricant film thickness. Numerical investigations by PAUSCH showed that in EHL contacts with pure rolling, a negative influence on the lubricant film already occurs at structure depths in the range of the lubricant film thickness [21]. Although the depth of real pittings is significantly greater than the lubricant film thickness, surface damage with a significantly lower depth does occur in other damage patterns, e.g. particle indentations, so that the investigation of such low structure depths is relevant to real bearing damage. Therefore, the following condition for specifying the structure depth was set: The lower factor level $t_p\Theta$ should be lower than the film thickness at a specific lubrication film thickness of $\Lambda = 3$, at which point a complete separation of the contact partners without contact of surface asperities can still be assumed. From this requirement results the lower level of the structure depth $t_p\Theta = 1$ μm.

A summary of the factor levels of the structural geometry is given in Table 2.
Table 2. Factor levels of the parameters describing the artificial surface structures.

<table>
<thead>
<tr>
<th>Factor level</th>
<th>Length $l_p$ in μm</th>
<th>Width $w_p$ in mm</th>
<th>Depth $t_p$ in μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>⊖</td>
<td>80</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>⊕</td>
<td>400</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

3 Theory and calculation

In both test series, characteristic peaks occurred in the impedance signal, which could be linked to the damage, see section 4. In this section, the method for evaluating the duration of these peaks is described.

Figure 11 shows an exemplary peak from the tests with artificial damage. The beginning of a peak $t_1$ is clearly marked by a sharp increase in the imaginary part of the impedance. The end of the peak, on the other hand, cannot be clearly determined. The imaginary part first reaches a minimum, then rises again to a secondary maximum and then decreases again. The end of this second falling edge cannot be clearly defined and is therefore not suitable for a quantitative analysis. For the real part, the signal is similar with reversed signs. To obtain a clearly mathematically defined criterion, the minimum between the main maximum and the secondary maximum of the imaginary part is therefore defined as the end of the peak $t_2$. Mathematically, $t_1$ and $t_2$ are defined as the times at which the slope of the imaginary part of the impedance disappears, i.e. the condition

$$\frac{d \text{Im}(Z)}{dt} = 0$$

(3.1)

is satisfied. The duration of the peak is thus the difference between the beginning and the end of the peak,

$$\Delta t = t_2 - t_1.$$  

(3.2)

Figure 11 shows the determination of $\Delta t$ on an exemplary measurement.

In a second analysis step, the distance which the rolling element travels relative to the damaged race during the time $\Delta t$, i.e. the rolling distance $\Delta l$, is calculated. This analysis is the basis for the evaluation of the size of the damage in section 4.

Figure 12 represents the essential geometric relationships for the calculation of the rolling distance for the inner race. Within a period $t$ the cage, and with it also the rolling element set, moves by an angle
\[ \varphi_K = 2\pi f_K t, \]  
\[ \varphi_{IR} = 2\pi f_{IR} t. \]  

and the inner race moves by an angle

\[ \varphi_{IR} = 2\pi f_{IR} t. \]  

Since the rotation speeds of the rolling element set \( f_K \) and the inner race \( f_{IR} \) are different, there is an angular difference

\[ \Delta \varphi = \varphi_{IR} - \varphi_K \]  

between the inner race and the rolling element set.

By multiplying by the raceway radius of the inner ring \( d_{IR}/2 \), the distance traveled by the rolling element relative to the inner race in the time period \( t \) is obtained,

\[ \Delta l = \Delta \varphi \cdot d_{IR} = (\varphi_{IR} - \varphi_K) \cdot d_{IR}. \]  

Substituting (3.3) und (3.4) into (3.6) provides an equation for the rolling distance traveled during the duration of a peak \( \Delta t \),

\[ \Delta l = \pi (f_{IR} - f_K) d_{IR} \Delta t. \]  

The rolling distance on the outer ring can be determined analogously. For the case of a stationary outer ring the rolling distance can be calculated by

\[ \Delta l = \pi f_K d_{OR} \Delta t, \]  

d\( _{OR} \) being the raceway diameter of the outer ring.

Figure 12. Quantities used for the calculation of the rolling distance for the inner race.

4 Results and discussion

This section describes and discusses the results of the test series, whose setup and test procedures were described in section 2.

4.1 Results of fatigue tests

Figure 13 shows impedance measurements at different stages of the life of bearing #1. Immediately after the start of the test, a large number of fast spikes occur in the impedance signal. Such spikes are known to be caused by contact of surface asperities, as described in section 1.2. Therefore, the spikes in this signal can also be attributed to such asperity contact.
regardless whether the spike is caused by a direct metallic contact or by an electrical breakdown event due to an excessively high electric field in the insulating lubricant near the asperity. As described in section 2.2, the first test phase ends with the occurrence of a first pitting, which is called initial damage. Immediately before the initial damage, the signal is much steadier, because surface asperities were flattened during run-in. However, some spikes are visible in the real and imaginary parts of the signal, their temporal spacing corresponds to the reciprocal of the ball-pass frequency on the outer race, BPFO. Since the pitting occurred on the outer race, it is possible that these spikes occurred because a crack or other surface irregularity was already present at the location of the eventual damage, and its over-rolling caused a breakdown of the lubrication film and consequently a metallic contact. Immediately after the initial damage, fast spikes again dominate the signal. Whether this is due to the damage or to the intermediate disassembly is unclear. In addition, distinctive peaks are visible in the signal, their temporal spacing corresponds to the BPFO frequency. This peak is even more pronounced in the measurement after 30 minutes damage progression, and the duration of the peak also appears to be longer than immediately after the initial damage. The peak is much more pronounced in the imaginary part than in the real part. These peaks have a significantly longer duration than the spikes associated with metallic contact. Therefore, they will be referred to as slow peaks in the following. A similar behavior can be observed in principle for the other test runs. The negative values of the real part of the impedance are caused by an offset in the measuring setup. For the analysis in this paper, only the gradient is relevant, which is not affected by the offset.

Figure 13. Impedance measurements at different stages of the bearing life for bearing #1

In the next analysis step, the slow peaks occurring during damage progression are investigated in more detail. Figure 14 shows representative slow peaks at different times of the damage progression for bearing #1. The beginning $t_1$ and the end $t_2$ of these peaks are marked in the plots, clearly showing an increase in the duration of the peaks $\Delta t$. 
Using the calculations described in section 3, the rolling distance of the ball during the time $\Delta t$ is calculated, cf. Figure 15. The rolling distance increases monotonously from the beginning to the end of the measurements. In the case of bearings #1 and #2, the increase is not continuous, but occurs in steps interrupted by plateau phases. It is possible that an increase in pitting length takes place at the time of these jumps.

Figure 15 also shows the actual length of the pittings, derived microscope measurement. There is a remarkable concordance between the actual length and the duration of the peaks. This supports the hypothesis that the impedance peaks are indeed caused by the over-rolling of the pittings. In addition, it leads to the hypothesis that the duration of the peaks corresponds to the duration of the over-rolling event.

G. Martin et al.: Preprint submitted to Tribology International by Elsevier
4.2 Results of tests with artificial damage

To investigate the hypothesis that the duration of the peaks corresponds to the duration of the over-rolling event, tests with artificial surface structures of predefined dimensions were conducted, as explained in section 2.3.

Figure 16 shows the effects of the six factors on the duration of the peaks, expressed as rolling distance $\Delta l$. First, it should be noted that the confidence intervals of the effects are small compared to the effect strength, i.e., the effects observed in the experiment can be estimated with low uncertainty. All six factors have significant effects. The further analyses concentrate on the two factors with largest effects, i.e. the structure length $l_p$ and the radial load $F_r$.

![Figure 16. Effects of the factors on the duration of the peaks, expressed as rolling distance $\Delta l$.](image1)

The strong effect of the structure length $l_p$ is plausible and reflects the finding from section 4.1. A physical explanation for the effect of the radial load $F_r$ will be given below, illustrated by Figure 17. The perturbation of the EHL contact by the surface structure begins when the leading edge of the Hertzian area gets into contact with the edge of the structure at time $t_1$. Correspondingly, the perturbation ends when the trailing edge of the Hertzian area leaves the structure. Between these instances, the rolling element moves relative to the inner ring by a distance of

$$\Delta l = 2b + l_p. \quad (4.1)$$

Since the semi-minor axis $b$ depends on the load, the distance $\Delta l$ also depends on the load.

![Figure 17. Physical model for the influence of the radial load on the duration of an impedance peak](image2)

According to this model, the effect of the structure length on the rolling distance determined should correspond exactly to the difference of the two factor levels of the structure length $\Delta l$. 

G. Martin et al.: Preprint submitted to Tribology International by Elsevier
The effect of the radial load should correspond to the difference between the dimension of the Hertzian area 2b at the factor level \( F_{r} \) und at the level \( F_{r} \). The difference of the actual structure lengths is \( \Delta l_{p} = 355 \mu m \) and the difference of the Hertzian dimensions is \( \Delta(2b) = 114 \mu m \). These values are within the confidence interval of the effects, i.e., there is no statistically significant difference between these values and the experimental values.

The analysis so far is based on the effects, i.e. the change of the rolling distance caused by a change of the factors. For verification purposes, the prediction of the physical model according to eq. (4.1) is compared to the experimental results; the influences of all other factors are eliminated by averaging. For all four combinations of load and structure length, there is a good agreement between the physical model and the experimental results.

![Graph](image)

**Figure 18.** Comparison of the experimentally observed rolling distance and the physical model

In the fatigue tests described in section 4.1, the temporal spacing of the impedance peaks corresponds to the ball-pass frequency on the inner race. This gives rise to the hypothesis that the location of the damage can be identified by a spectral analysis of the impedance. Therefore, surface structures (depth \( t_{p} \), width \( w_{p} \) and length \( l_{p} \)) were also introduced on the outer race and on a rolling element. The characteristic impedance peaks could also be observed on the outer race and rolling element. Figure 19 shows a comparison of the impedance spectra for the three damage locations. The ball-pass frequencies on the outer (BPFO) and inner race (BPFI) as well as the ball spin frequency (BSF), calculated according to Randall [1], are clearly elevated. In the case of the inner race, strong sidebands are visible, which are due to the modulation of the rolling element load caused by the rotation of the inner race. This indicates that the temporal spacing of the impedance peaks depends on the damage location and that the damage location can be identified by spectral analysis.
5 Conclusions

The goal of this study was to investigate how the impedance of rolling bearings is influenced by surface damage. From the tests described in this paper, several conclusions about this relationship can be drawn. Characteristic peaks occur in the impedance signal when real pitting damage or artificial surface structures are present on the raceway surface. The temporal spacing of these peaks corresponds to the ball-pass frequencies on the damaged element, enabling the localization of the damage. The geometric length of the damage is directly related to the duration of the peaks. By calculating the rolling distance, the distance traveled by the rolling element relative to the race during the duration of a peak, it is possible to infer the length of the damage and thus monitor damage progression. Lastly, the rolling element load influences the duration of the peaks. In applications with varying loads, this needs to be considered when evaluating the damage length. A simple analytical model for the relationship, which explains the peak duration as a function of the rotation speed, the length of the damage and the bearing load was developed and validated.

These conclusions show that electric impedance analysis is capable of diagnosing not only the presence of surface damage in rolling bearings, but also its extension along the raceway.
6 References


