Vehviläinen, Milla; Rahkola, Pekka; Keränen, Janne; Halme, Jari; Sopanen, Jussi; Liukkonen, Olli; Holopainen, Antti; Tammi, Kari; Belahcen, Anouar

Adapting geometry-based polygonal contacts for simulating faulty rolling bearing dynamics

Published in:
Mechanism and Machine Theory

DOI:
10.1016/j.mechmachtheory.2023.105552

Published: 01/02/2024

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
Adapting geometry-based polygonal contacts for simulating faulty rolling bearing dynamics

Milla Vehviläinen a,*, Pekka Rahkola a, Janne Keränen a, Jari Halme a, Jussi Sopanen b, Olli Liukkonen b,c, Antti Holopainen c, Kari Tammi d, Anouar Belahcen e

a VTT Technical Research Centre of Finland Ltd., P.O. Box 1000, FI-02044 VTT, Finland
b Lappeenranta-Lahti University of Technology, FI-53850 Lappeenranta, Finland
c ABB Oy, FI-00380 Helsinki, Finland
d Aalto University, Department of Mechanical Engineering, FI-02150 Espoo, Finland
e Aalto University, Department of Electrical Engineering and Automation, FI-00076 Espoo, Finland

A R T I C L E   I N F O

Keywords:
Multibody simulation
Polygonal contact method
Ball bearing
Data generation
Fault analysis

A B S T R A C T

Rolling bearings are a leading cause of equipment breakdowns in electrical machines, underscoring the significance of predictive maintenance strategies. However, the given methods require high-quality big data, which is challenging to acquire, especially for faulty cases. Simulation models offer an alternative by generating large data sets to complement experimental data. However, bearings involve complex contact-related phenomena, such as slipping and clearance. Therefore, generating realistic data comparable to the real-world necessitates accuracy. Our study presents a multibody simulation system of a motor bearing, incorporating a geometry-based polygonal contact method (PCM), which accurately captures nonlinear bearing dynamics and allows for the simulation of various contact geometries. We introduce a systematic approach to adjust the PCM contact properties for rolling bearings, referencing the well-established Hertzian theory. Both healthy and faulty bearings with a local outer ring fault were simulated. The simulated output was a relative shaft displacement, experimentally validated using a capacitive sensor. Our model successfully demonstrates the potential to employ geometry-based contacts for generating realistic data on faulty bearings with the aim of predictive maintenance.

1. Introduction

Rolling bearings play a critical role in all rotating machinery [1], contributing to 40% of electrical machine breakdowns [2,3] and up to 90% in small machinery systems [4]. As they pose a relatively high risk of failure [5], proper condition monitoring (CM) is necessary to avoid further consequences. In addition to traditional CM, where only the current state of a machine is evaluated, a common approach has been to apply predictive maintenance (PdM) for diagnosing early-stage faults before they occur and accumulate severe damage [6].

PdM exploits CM data with predictive data analytics, such as machine learning (ML), to distinguish certain fault-related characteristics from the normal behaviour of the machine in an automated and proactive way [7,8], thus reducing the risk of...
equipment downtime and increasing efficiency. However, PdM with ML relies on big data of high quality. While advanced ML techniques can provide powerful predictive capabilities, the success of PdM depends more on the quality and amount of available data. More data usually result in more accurate predictions of the current state of a machine [9,10]. Especially acquiring such data on faulty scenarios appears as a practical challenge. Obtaining valid data on faulty cases is non-trivial because bearings, being critical components, are usually replaced before they break completely, making it challenging to collect sufficient data. Besides, most manufacturers have not systematically collected data from their applications and, if they have, they are often reluctant to share it publicly [10,11]. Hence, not much experimental data exist from a statistical perspective.

As an alternative, simulation models can generate large data sets with low variable costs and potentially complement bulky and expensive experimental setups [12]. Once developed, comprehensive models can outperform experimental data collection regarding process speed and cost-efficiency, especially in long-term use cases. Moreover, computational models enable the simulation of multiple configurations with varying parameters providing greater precision and control over more variables than an experimental system [5]. This, in turn, allows exploring the causal role of different components and their features in ways that would otherwise be unrealisable. Generally, experimental data is still needed to validate and augment the simulation models, but the current trend has been to find out if it can be at least partly replaced by simulation data [13].

Various analytical simulation models have been developed to study the ball bearing dynamics with local faults [14–16] as well as with combined faults [17–19] and distributed faults, such as surface waviness [20,21]. Analytical models often adopt a range of presumptions to simplify the model, which leads to a limited number of degrees of freedom (DOFs). For example, rolling elements of bearings are usually treated as a combined set of elements rather than considered individual entities. In reality, however, bearing dynamics involve complex nonlinear phenomena, and to simulate those accurately, more DOFs are needed [22,23]. Therefore, numerical methods are suggested for simulating all bearing elements including their physical interactions with other elements [24].

Especially contact description plays a critical role in bearing dynamics [25]. It is affected not only by the contact geometry but also by multi-clearances, friction and lubrication while rolling and slipping occur independently or simultaneously. The contact description directly affects the accuracy of the output data. For example, Wen et al. [24] presented a 68 DOF rigid body model of a ball bearing, while Wagner et al. [26] introduced a ball bearing with 18 rigid DOFs and an elastic outer ring contributing an additional 250 DOFs, using numerical multibody simulation (MBS) in both studies.

MBS, as a more complex and accurate alternative to analytical models, is a computational technique used to study the dynamics of mechanical systems consisting of interconnected bodies [27]. It is particularly suitable for bearing models, where simulation output is sensitive to the accurate representations of bearing components with tight clearances and strict tolerances. MBS offers a range of contact models that can be adapted to different bearing types and operating conditions, allowing for more accurate predictions of bearing behaviour and performance. The most widely applied Hertzian contact has been used to model the contacts between bearing rings and rolling elements [13,28–30]. Nonetheless, Hertzian theory works best for simple shapes, with low-impact dynamics [31]. However, a common denominator of the dynamic bearing models in the past has been the simplified fault geometry defined by a mathematical function or a geometric notch with a rectangular cross-section, as in [5,12,24]. Additionally, a fault size in past approaches is often relatively large compared to the rolling element diameter, which diverges from real faults that appear gradually and are of arbitrary shape in random locations.

An alternative methodology for the Hertzian theory is the geometry-based polygonal contact model (PCM), a 3D multipoint contact for detecting contacts on rigid and flexible bodies [32,33], which is based on the actual surface geometries of the contacting bodies. It takes into account the local material properties and surface topography to represent the contact dynamics with greater fidelity, thus offering a more accurate and realistic representation of the whole bearing compared to the traditional methods.

PCM has been applied in many contact-critical MBS applications, for example, in the biomedical field [34], aerospace engineering [35–37], aeronautical applications [38] and transportation systems [39]. In the previous applications, the relative velocity between contacting bodies has been low compared to bearings with typical rotational speeds. The model needs to update contact calculations at each time step, so decreasing the step size increases computing time exponentially [40]. Conversely, too large step size may accidentally ignore some contact points, resulting in incorrect contact penetrations [41]. Previously, the computational effort was not sufficient for such simulation, and simpler approaches had to be settled for instead [42].

In this study, we exploit the PCM approach to model bearing contacts between rolling elements and raceways for a more accurate and realistic dynamics description. The primary contribution is to investigate if this particular contact method can be applied to bearing applications. To our best knowledge, there is no previous research on using PCM for such applications where the velocity between contacting surfaces is relatively high. Our aim is to carry out a simulation model that is capable of generating synthetic data for the PdM of ball bearings. Although the PCM is meant for even complex and arbitrarily shaped faults, here we model a simple fault from which we also have the measured validation data.

Using the PCM, polygonal geometry inherently has a significant impact on the contact properties, especially with small penetrations. To make the PCM suitable for bearing simulations, we introduce a systematic approach to adapt the contact properties to meet application-specific needs. For this, the well-established Hertzian contact theory was used as a reference point to determine the PCM parameters.

The bearing under study is a standard single-groove ball bearing, located at the non-drive end (NDE) of an electrical motor. The fault is an artificially made local single-point fault on the outer ring. The simulation model is validated with an experimental test setup where the radial relative shaft displacement data was measured by a capacitive displacement sensor. We simulate both the healthy and faulty bearing at different velocities and obtain a radial shaft position to correspond to the capacitive data. Both the measured data and the simulated results were analysed in the time and frequency domain. In the latter, spectral analysis
for experimental and simulation data included employing Hanning’s window to prevent spectral leakage and using Fast Fourier Transformation (FFT) to separate the signals into distinct frequencies.

The remainder of the article is organised as follows. Section 2 introduces the bearing dynamics for fault analysis, and Section 3 explains the state-of-the-art behind the contact modelling. In Section 4, the experimental test bench for validation is presented, followed by Section 5, where the simulation model is described. Section 6 specifies the data and its processing. Results are drawn in Section 7, and finally, Section 8 concludes.

2. Bearing fault analysis

A typical ball bearing consists of an inner ring (IR), an outer ring (OR), rolling elements (REs) and a cage that maintains equidistance between the REs. The size of the bearing and the number of its REs vary according to the application. In industrial electromechanical applications, the OR is typically fixed in place, and the IR rotates along the shaft [43]. Here we outline the impact of real-world phenomena, such as axial load, clearance and slipping, on the fault analysis of a ball bearing. Incorporating these phenomena into a bearing model necessitates taking translations and rotations to an axial direction into account in a three-dimensional coordinate system.

Determining bearing faults in different parts, namely those located in REs, cage or rolling raceways, is based on the theory of characteristic fault frequencies that depend on bearing size, loading conditions and the rotational frequency of a rotating machine. When an RE rolls over a fault located on a rolling raceway, it induces periodical impulses that appear at the characteristic fault frequencies [44]. In the frequency domain, the periodic, rotation-related impulse responses are seen as a peak amplitude at the characteristic frequency. Each main bearing element has its own characteristic fault frequency. In this study, the focus is on the OR faults that generate the disturbance at the OR failing frequency, also known as the ball pass frequency for the OR (BPFO) [2]:

\[
\text{BPFO} = \frac{N_{RE}}{2} \cdot f_r \cdot \left(1 - \frac{D_{RE}}{D_p} \cdot \cos \theta \right),
\]

where \( f_r \) denotes the rotational frequency of the IR that is expected to rotate along the shaft, \( D_p \) is the centre diameter of the cage (pitch diameter), \( D_{RE} \) is the diameter of one RE, \( N_{RE} \) denotes the number of REs in the bearing and \( \theta \) is the contact angle (Fig. 1) that is a result of an inclination of the raceways relative to the bearing axis, caused by an axial load. The larger the contact angle, the more the contact points are shifted away from the rolling raceway centres, and thereby, the respective radii to the contact points are adapting so that the radius of the contact point at the outer raceway, \( R_{oc} \), shortens and the corresponding radius to the inner raceway contact point, \( R_{ic} \), increases (Fig. 1). The degree of the contact angle is naturally affected by the internal clearance and axial loading.

BPFOs assume ideal continuous contacts ignoring real-world bearing phenomena, namely slip and radial clearance. On the one hand, the clearance affects so that the axial preload and inclined contact area between REs and raceways increase the contact angle \( \theta \). The angular velocity of the cage increases in consequence. Moreover, radial clearance induces fluctuations in contact forces, potentially resulting in a scenario where only certain REs are constantly in contact with both IR and OR. The REs without proper contacts are driven by the cage, and the others are driving it [45]. The cage ensures that the REs maintain an average angular velocity while exhibiting some degree of slip. On the other hand, because of the slip, the cage angular velocity is reduced. Estimating or validating the precise amount of slip can be challenging. Slipping in bearings typically increases with velocity, as the centrifugal
forces acting on the REs increase [46]. With normal lubrication, slipping is assumed to dominate the cage velocity more than the clearance. Both clearance and slipping inherently affect the characteristic frequencies, causing a random deviation compared to those calculated through the theoretical Eq. (1). Studies have shown that the typical deviation ranges from 1%–2% [47] without additional loading.

3. Dynamic contact description in MBS

MBS studies the dynamic motion of mechanical systems caused by external forces. It compounds numerical methods for solving nonlinear systems in the time domain. The systems are composed of rigid bodies that are connected to each other via kinematic joints or force elements. The defining characteristic of the bodies is that they undergo large translational and rotational displacements. MBS can be used for studying the kinematic and dynamic motion characteristics of a wide variety of systems in different fields of engineering. More related MBS theory is introduced, for instance, in [27,48].

Contact between colliding bodies is a complex nonlinear phenomenon involving normal and tangential force components. In MBS, the numerical contact representation is often approximated by neglecting contact stress and deformation between rigid bodies for better computational performance. Common contact methods are based on contact forces defined as continuous functions of the relative penetration between colliding surfaces in a normal direction [40]. Such contact forces control the volume of penetration via spring–damper elements. The spring component determines the elasticity at the contact point, while the damper component accounts for energy dissipation.

The most widely accepted contact model in MBS for bearing contacts is the Hertzian contact stress theory [23], where the cross-sectional area of contact between two spheres is assumed to be an ellipsoid due to contact deformation. The size of the ellipsoid, depending on the radii of the contacting spheres and the elastic modulus of materials, is proportional to the magnitude of the contact stress [49]. The majority of previous bearing contact models use Hertzian theory because of its simplicity and comparable validity [50]. Thus, Hertzian equations provide a feasible contact definition, which is often used as a reference to contact modelling. The following Hertzian Eqs. (2)–(6) are particularly formulated for the contact between a sphere and a bearing raceway [51,52].

Hertzian variables $S$ and $\gamma$ are defined based on geometrical dimensions and material properties of the contacting shapes:

$$S = \frac{1}{2} \left( \frac{1}{R_{RE}} - \frac{1}{r_c} + \frac{1}{r_a} \right),$$

$$\gamma = \frac{(1 - \nu) \cdot 2}{E},$$

where $R_{RE}$ and $r_c$ are the radii of the RE and the contacting rolling raceway, and $R_c$ is the radius to the contact point. Young’s modulus and Poisson’s ratio are denoted by $E$ and $\nu$, respectively. The latter equation assumes that the given materials are identical in both contacting bodies. As the Hertzian contact is presented as an elliptical area, the semimajor and semiminor axes $a$ and $b$ are calculated as follows:

$$a, b = 1.145 \cdot n_{a,b} \cdot (F \cdot S \cdot \gamma)^{1/3},$$

where $F$ is the contact force. Assisting variables $n_a$ and $n_b$ are further formulated in [52]. The same applies to $n_c$ and $n_d$, which are used for solving the maximum contact pressure $p_h$ and the penetration depth $u_h$:

$$p_h = 0.365 \cdot n_c \cdot \left( \frac{F}{S^2 \cdot \gamma^2} \right)^{1/3},$$

$$u_h = 0.655 \cdot n_d \cdot \left( \frac{F^2 \cdot \gamma^2}{S} \right)^{1/3}.$$

However, the Hertzian theory is mostly applicable to simple and uniform geometries modelled by elliptical, line and point contacts with low-impact dynamics [31]. It uses geometrically defined points on the body surface and their normal direction to locate the contacts. This means that the shapes of the contacting bodies need to be modelled as simple volumes, such as cylinders or spheres, which is why Hertzian is generally applicable for healthy bearings or those with fairly simple fault geometries. In bearing applications, however, real raceway defect geometries are rather arbitrary and non-symmetric. Besides, precise manufacturing tolerances and finer geometry play a critical role when it comes to the proper functioning of an MBS system, which is difficult to realise with Hertzian’s continuous contact.

A more advanced contact model called the polygonal contact method (PCM), originally introduced in [32], is an approach where the body surfaces are described by polygonal meshes composed of triangular discretisation, a common internal representation of spatial geometries in computer graphics. PCM provides a contact algorithm specifically intended for dynamic MBS cases to efficiently handle polygonal surfaces with arbitrary complexity. The ultimate practical advantages of PCM are the possibility to modify the model geometry directly through a CAD design, and the method does not necessitate the assumption of continuous contacts. The former is especially substantial with faulty components. The latter facilitates the simulation of more realistic bearing dynamics even when sudden contact losses occur, which is also a real-world phenomenon. With bearings, this allows us to effortlessly take instances, such as radial and axial clearance as well as slipping, into account.

The basic condition for geometrical contact is that at least one pair of intersecting polygons exists. The typical but computationally intensive operation in existing contact methods is to test every triangle on a primary surface against each triangle on a
secondary surface for an intersection at a particular timestep and position [32]. Most of the computing time is consumed in the contact detection phase due to its iterative nature [31]. In contrast, the PCM intersection detection algorithm uses a bounding volume (BV) strategy to avoid as many intersection tests as possible. BV is a preprocessing phase where an entire surface is divided into smaller subareas until the final intersecting areas are found. Then the computationally heavier analysis is performed only for these parts. In case there is no contact in a certain subarea, its triangles are no longer included in the further collision tests. The BV algorithm is also applied at each particular timestep and position. Lastly, and only in case of contact, the corresponding force and torque components are calculated.

In PCM, the contact forces in the normal direction are determined by the elastic layer stiffness, which is used to calculate the force vector for each polygon within the contact patch. The elastic layer stiffness model is a rigid body model which is assumed to be covered by a thin layer of elastic springs — one for each polygon. Eqs. (7)–(10) introduce the contact description according to [32] and Eq. (11) shows the integration of the stiffness scale function. Material properties Young's modulus $E$ and Poisson's ratio $\nu$ are constant for each polygon and define linear-elastic characteristics for the thin layers. With homogeneous compressible materials ($\nu < 0.45$), the elastic modulus $K$ can be derived as follows:

$$K = \frac{1 - \nu}{(1 + \nu)(1 - 2\nu)} \cdot E. \tag{7}$$

PCM neglects tangential shear stress in a layer of thickness, which results in a direct relation of the normal penetration and pressure:

$$p = \frac{K}{b} \cdot u, \tag{8}$$

where $u$ is the normal penetration, $p$ is the pressure and $b$ is the elastic layer thickness. Both $u$ and $p$ are assumed to be constant for each triangular contact element. The normal force $F_k$ in an individual contact element $k$ then results in:

$$F_k = c_k \cdot A_k \cdot u_k, \tag{9}$$

where $A_k$ is the area of the contacting polygon and $c_k$ is the corresponding contact stiffness. In addition, contact stiffness $c_k$ for a single polygon $k$ depends on the polygon size and the elastic layer:

$$c_k = \frac{K}{b} \cdot A_k. \tag{10}$$

The total stiffness $c_{tot}$ of a certain contact patch is determined as a combination of primary and secondary surface stiffness [53]:

$$c_{tot}(u) = \frac{c_1 \cdot c_2}{c_1 + c_2} \cdot f(u), \tag{11}$$

where $c_1$ is the normal stiffness of the contacting polygon on the primary surface and $c_2$ is the corresponding stiffness on the secondary polygon. Practically, this means that in order to ensure consistent contact stiffness across all polygons on both surfaces, their polygonal discretisation should result in equally sized polygons. The total stiffness $c_{tot}$ can be further scaled by using an input function $f(u)$ that is dependent on the normal penetration.

The fundamental difference between the Hertzian and PCM contact models lies in their distinct ways of calculating the overlapping volume between contacting bodies. In the Hertzian, the overlapping volume is approximated by mathematical equations (Eqs. (4)–(6) with multipliers) whereas the PCM derives this directly from the geometries. Specifically, the Hertzian contact area increases linearly with respect to penetration, while the PCM contact area depends on the contacting geometries, and with small penetrations, the polygonal discretisation particularly affects the accuracy. This affects inherently so that in addition to area-penetration relation, contact pressure descriptions of these contact methods are incomparable. However, with bearing applications, contact stiffness plays an important role. In the Hertzian theory, contact stiffness is given by the ratio of the applied load to the elastic deformation of the surfaces. This is different from the PCM, where each polygon has its own local stiffness, and the overall contact stiffness is obtained by integrating the local stiffness values over the entire contact area. Using the scale function $f(u)$ in Eq. (11) the PCM contact stiffness, represented as the force-penetration relationship, can be adapted for the application in question despite the polygon size, up to some point. To achieve a proper scaling for our bearing application, we use the Hertzian theory (Eqs. (4)–(6)) as the foundation to obtain the reference for the contact stiffness in the ball bearing. Simultaneously, other PCM properties, such as the layer depth $b$ and a proper polygon size for contacting components, can be selected, recognising that the polygon size exerts the most significant impact on scalability in this context.

In addition to contact forces in a normal direction, tangential friction forces need to be considered for a complete contact description. Most contact approaches assume frictionless surfaces [40], so a common way to model tangential forces is to apply a separate friction description. Similarly, PCM can also be combined with a selected friction description. Coulomb's friction is a widely used method for modelling contacts in rolling bearings [13,30]. It assumes that the frictional force between two bodies is proportional to the normal force acting between them, and the frictional force is limited by a maximum value called the friction coefficient. We establish a regularised single friction coefficient value that remains constant, irrespective of the shaft rotation speed, thereby eliminating stiction-friction. This friction description has also been referred to as Bernard friction, a friction force model that mitigates the discontinuity inherent in Coulomb's law by employing a piecewise linear relationship between force and velocity [54], as in Eq. (12) for the tangential friction force:

$$F_t = \begin{cases} \mu \cdot F_n \cdot \min(k_e \cdot v_t, 1) & \text{if } v_t \geq 0, \\ \mu \cdot F_n \cdot \max(k_e \cdot v_t, -1) & \text{if } v_t < 0, \end{cases} \tag{12}$$
Mechanism and Machine Theory 192 (2024) 105552

M. Vehviläinen et al.

where \( \mu \) is the friction coefficient and \( F_n \) is the normal contact force. The relative tangential velocity of the contacting surfaces is represented as \( v_t \), and \( k_\epsilon \) represents the slope of the linear transition at low relative tangential velocities. Overall, Coulomb’s dry friction model provides a simple yet accurate representation of the frictional behaviour of the contacts, and it is numerically efficient and straightforward to apply in dynamic simulation systems [55].

In general, PCM can be applied in any open-source platform designed for MBS. Additionally, Simpack software already provides a separate polygonal contact force element called FE199 [53], particularly based on PCM. Despite the efficient contact detection algorithm, PCM demands significant computational effort in typical bearing applications. With multiple computer processor cores, the calculation can be divided into multiple threads that can be executed in parallel, reducing the overall computing time. Contact analysis especially is a task that benefits from parallelisation [56].

4. Experimental test bench

The experimental test bench consists of two electrical machines connected to each other by a coupling that eliminates external radial loads (Fig. 2). The motor is an 11 kW three-phase ABB machine with a nominal speed of 1500 rpm. The load machine acts as a generator but does not apply any torque load, and in this case, it is considered additional inertia. The total rotor system mass of the motor is about 37 kg and the corresponding polar mass moment of inertia is 0.103 kgm\(^2\). The rotor balance class is G2.5 according to ISO 21940-1 [57], which means that with the particular machine, the mass centre is allowed to locate radially about 15 \( \mu \)m aside from the rotation centre. This balancing class is rather directive and intended for the original manufacturer to balance a single rotor. The required balancing may be difficult to achieve within the present lab circumstances as the repeating re-adjusting of the bearings and the connection to the load machine inevitably cause additional unbalance, so this value is evaluated with caution. Further, there are uncertainties associated with the actual balancing condition of the entire rotor system, including the coupling.

There was no external loading to the system except for the gravitational force. The NDE bearing was preloaded axially by a wave spring with 229 N/mm spring constant resulting in 445 N. The axial translation of the shaft was fixed at the drive end (DE) and free at the NDE to consider possible thermal expansion. However, the temperature was kept at about 32 \( ^\circ \)C to imitate the situation without the thermal effect. The temperature inside the bearing shields was measured from the NDE cooling air by a separate temperature sensor.

The rotor shaft was supported by two single-groove ball bearings, between which the distance was about 430 mm. We concentrated particularly on the bearing at the non-drive end (NDE), where the measured data contain less noise and disturbance from the load machine. The bearing was a standard SKF 6209 ball bearing with ten rolling elements and a bore size of 45 mm (Fig. 3). Not all geometrical bearing dimensions are shared by the manufacturer, which is why some complementing values were experimentally measured [58]. RE diameter was measured using a digital length measuring machine, whereas groove radii and ring radii were measured with a CNC coordinate measuring machine. The measured values were carried out with a resolution of 0.05 \( \mu \)m. The specification parameters of the bearing are listed in Table 1 with rounded values and visualised in Fig. 4.

The faults in the bearings were represented by a single 2 mm diameter hole drilled through the OR (Fig. 3). This kind of drilling easily causes a tearout, which is a small irregular piece of material torn out from the surrounding area. The axial position of the hole was 1 mm from the centreline towards the DE to locate the fault to the rolling track of the balls even under the axial preload (Figs. 3 and 4). This is relevant mostly in the simulation of the bearing because, in the experimental setup, the tearout from drilling the hole would amplify the contact with the REs, even though the REs are not exactly rolling over the fault.

To measure the relative displacement of the rotor shaft at the NDE, we employed a capacity sensor, shown in Fig. 5, originally designed to accurately control active magnetic bearing applications. The displacement range of the sensor was from \(-1\) to 1 mm with a resolution of 0.16 \( \mu \)m and a sampling rate of 20 kHz. The sensor consists of four capacitive plates positioned around the rotor shaft. The distance from the plates to the shaft was measured radially as a voltage signal of a resonance circuit. Voltage signals were
Fig. 3. The standard SKF 6209 bearing with a 2 mm diameter fault on the outer ring was located 1 mm beside the axial centreline. The fault location was adjusted between 0°, 180° and 270°.

Table 1
Dimensions of SKF6209 ball bearing. Note that the values are rounded.

<table>
<thead>
<tr>
<th>Denotation</th>
<th>Dimension</th>
<th>Value [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_Ball</td>
<td>Ball diameter</td>
<td>11.90</td>
</tr>
<tr>
<td>R_OR1</td>
<td>Outer ring outer radius</td>
<td>42.50</td>
</tr>
<tr>
<td>R_OR2</td>
<td>Outer ring inner radius</td>
<td>37</td>
</tr>
<tr>
<td>R_IR1, R_s</td>
<td>Inner ring inner radius, shaft radius</td>
<td>22.50</td>
</tr>
<tr>
<td>R_IR2</td>
<td>Inner ring outer radius</td>
<td>56</td>
</tr>
<tr>
<td>R_OG</td>
<td>Outer raceway radius</td>
<td>38.09</td>
</tr>
<tr>
<td>R_IN</td>
<td>Inner raceway radius</td>
<td>26.17</td>
</tr>
<tr>
<td>R_p</td>
<td>Pitch diameter</td>
<td>32.13</td>
</tr>
<tr>
<td>r_o</td>
<td>Outer groove radius</td>
<td>6.50</td>
</tr>
<tr>
<td>r_i</td>
<td>Inner groove radius</td>
<td>6.14</td>
</tr>
<tr>
<td>c_d</td>
<td>Radial clearance</td>
<td>0.03</td>
</tr>
<tr>
<td>x_shift</td>
<td>Axial fault location</td>
<td>1.00</td>
</tr>
</tbody>
</table>

converted into digital values, namely micrometres. Because of the number of plates, the capacitive surface is larger and, thereby, the averaging effect increases the accuracy. This also reduced the effect of the surface roughness of the shaft, unlike many commercially available proximity sensors. A more detailed description of the capacitive sensor is reported in [59].

Although capacitive sensors offer improved access to bearing frequencies, shaft displacement is sensitive to the disturbance caused by motor-specific characteristics, manufacturing tolerances and manual installation. Surface irregularities and out-of-roundness of a shaft, collectively referred to as mechanical motor runout [60], are inherent features of experimental setups.

To compensate for runout, it needs to be extracted from the measured data during post-processing. Note that different compensating methods may yield marginally dissimilar outcomes, which is why we tested the effect of runout compensation only on the healthy data with 1500 rpm. We applied the so-called slow roll method, where the runout is measured separately at a slow velocity disregarding external forces, such as centrifugal forces. Mechanical runout remains constant despite the varying velocity. An encoder was used to obtain the displacement with respect to the angular shaft position at the runout velocity of 150 rpm as well as at 1500 rpm.

The study conducted four experimental runs to measure different health conditions, including the bearing in its healthy state and with an OR fault positioned at three distinct angular positions (Fig. 3). Therefore, the healthy bearing was run once, after which the faulty bearing was run three times. The location of the fault was changed simply by rotating the faulty bearing 90° between the runs. The capacitive sensors were temporally removed and reinstalled every time the angular position or the whole bearing was changed. This was done with the special assembly jig to ensure a constant air gap between the sensor frame and the rotor shaft. Subsequently, the velocity was gradually accelerated and occasionally levelled out to a constant velocity to measure a three-second section at three different velocities. The velocities were 750, 1125 and 1500 rpm. The motor rotation direction was clockwise from the NDE. Table 2 represents the measurement sets.
Fig. 4. Cross-section of the ball bearing under study with the dimensions available by the manufacturer as well as the additional measurements of the geometry.

Fig. 5. The capacitive sensor that was used to measure the relative radial displacement at the NDE.

Table 2
Implementation of the experimental measurements. From each run 1–4, capacitive displacement data at three distinct velocities were extracted.

<table>
<thead>
<tr>
<th>Run</th>
<th>Condition</th>
<th>Speed 1</th>
<th>Speed 2</th>
<th>Speed 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>Healthy</td>
<td>750 rpm</td>
<td>1125 rpm</td>
<td>1500 rpm</td>
</tr>
<tr>
<td>Run 2</td>
<td>OR fault at 0°</td>
<td>750 rpm</td>
<td>1125 rpm</td>
<td>1500 rpm</td>
</tr>
<tr>
<td>Run 3</td>
<td>OR fault at 180°</td>
<td>750 rpm</td>
<td>1125 rpm</td>
<td>1500 rpm</td>
</tr>
<tr>
<td>Run 4</td>
<td>OR fault at 270°</td>
<td>750 rpm</td>
<td>1125 rpm</td>
<td>1500 rpm</td>
</tr>
</tbody>
</table>

* The data at the runout velocity of 150 rpm was collected simultaneously with the encoder measurement.
5. Simulation model

We replicated the experimental case with the MBS approach, demonstrating how the contact properties, such as contact stiffness, can be scaled to be suitable for bearing application using the PCM, available in Simpack software. The complete simulation system represents the motor rotor assembly supported by two ball bearings, and it has 57 DOFs in total. The RE–IR and RE–OR contacts were modelled using the PCM. At this point, we did not model the contacts between the rolling elements and cage pockets. This was because specific dimensions for the cage geometry were unavailable, and we found that the benefits gained from including cage contacts when simulating a fault on the outer ring were not significant enough to justify the additional computational load. For the same reason, the fault was modelled as a clean hole without the tearout geometry. Besides, the tearout does not represent natural types of defects. The given limitations are listed as follows:

- The system is limited to the motor, and the load machine mechanics are excluded.
- The DE bearing is modelled as a bushing element.
- The simplified lubrication representation, characterised by Coulomb’s friction, is simulated by using a lower friction coefficient to mimic the lubrication effects.
- The system consists of rigid bodies.
- Contacts between the cage pockets and REs are neglected. Consequently, REs are simulated with 5 DOFs by combinations of spherical and planar joints.

5.1. Geometries

The rotor assembly was modelled as the rotor and the shaft combined into a uniform geometry. The DE bearing was also modelled as one body. The NDE bearing under the focus consisted of the OR, IR, ten REs and the cage. All parts in the system were considered rigid solids as we are not interested in vibrations or phenomena that require flexible structures in this study. Fig. 6 shows the healthy bearing model. The coordinate system of the model was set so that the bearing rotates about the axial \( Z \)-axis, and axes \( X \) and \( Y \) form the radial direction.

The rotor assembly and the DE bearing were modelled as native Simpack bodies whose geometry is the most efficient for the Simpack solver. These parts were determined by simplified geometries and material properties to obtain masses and inertia, and they were not in touch with other components. Additionally, a separate point mass was added to the rotor to represent the rotor balance class. This was realised with a 7 g mass located radially 80 mm away from the rotor centre which shifted the mass centre by about 15 μm, aligning with the balance class G2.5. We acknowledge, however, that this balance class, while readily available, is likely to underestimate the extent of imbalance due to the aforementioned phenomena associated with the experimental measurements.

In contrast, the NDE bearing parts consisted of polygonal surface models which were CAD geometries imported to the MBS software as wavefront objects (obj-files), a format for encoding 3D geometry through polygonal meshes. Most commercial CAD modelling software provides the corresponding export format by default, but their automatic surface discretisation often results...
in large deviations in polygon size and shape. As a consequence, we employed a separate finite element meshing tool by Abaqus software for a finer discretisation of the models. Fig. 7 illustrates the polygonal representation of an RE generated with an automated CAD software export and remeshed with the meshing tool. In this example of a sphere, the automated meshing created polar regions that resulted in smaller and sharper polygons near the poles. Using the separate meshing tool, it was possible to define a constant polygon size throughout the sphere. The OR and IR geometries are not as prone to this issue as the RE geometry, but for consistent quality, their raceway surfaces were separately remeshed as well (Fig. 8). Moreover, the faulty OR was modelled simply by adding the 2 mm hole to the CAD geometry. Similarly to the experimental case, the fault was located axially 1 mm aside from the radial centreline (Fig. 3).

5.2. Connections and force elements

One end of the rotor shaft was connected to a simplified DE bearing to reduce the numerical load. This was implemented by a bushing element with radial and axial linear stiffness and viscous damping. In the axial direction, stiffness and damping values were set to zero so that the NDE of the shaft carried the axial load. In the radial direction, we configured the bushing properties so that they allowed the radial displacement of the shaft at the DE to be equivalent to the maximum radial clearance in the actual bearing. As a result, the values of \( 2 \times 10^7 \) N/m and \( 2 \times 10^4 \) Ns/m were used for the radial stiffness and damping, respectively.

The other end of the shaft was fixed to the IR of the NDE bearing, and the cage was attached from its centre point to the IR by a revolute joint. The cage maintains an equal angular distance between the REs. Each RE was attached to the cage by a spherical joint in tandem with a planar YZ, which allowed rotations about all three axes as well as axial and radial translations. As a result, each RE has 5 DOFs with respect to the cage, as demonstrated in Fig. 9.

For the simulation purposes, the OR was fixed to the ground, although in the test set-up, the OR was axially free. Furthermore, the REs interacted with both OR and IR through geometrical contact forces (described in Section 5.3), which inherently restricted the planar movements of the REs in the radial direction by the groove geometries. Fig. 10 visualises the system dynamics.

In addition to the contact forces and the bushing element, the axial preload applied to the shaft at the inner ring is modelled as an external spring-damper force. The preload of 445 N corresponds to the axial force, and the spring constant of 229 N/mm represents the axial stiffness. Typical ball bearings can be loaded radially without axial load, which at low velocities causes periodical sudden losses in RE–raceway contacts, which in real life are damped by the lubrication. The axial preload typically reduces slip and ensures more continuous contact for all REs. In our case, the axial preload allows for simulating the contacts with simplified lubrication. The radial load is caused purely by the gravitational force and the inertia forces.
5.3. Description of contacts

The load torque acting on the shaft is velocity-dependent viscous torque. The driving torque applied to the shaft is defined by a PI controller in a feedback loop to control the shaft rotation speed. This is to obtain a smooth acceleration up to the desired constant velocity. The acceleration phase is not useful for the current study, but because MBS is a time-dependent method, we need to model it to achieve a steady state. Finally, a sensor element records the relative radial displacement of the shaft at the NDE. The output data corresponds to the experimental data measured by a capacitive sensor.

5.3. Description of contacts

The NDE bearing contains twenty contacts in total — ten between REs and OR and another ten between REs and IR. The contact force element applied in our model is labelled as FE199 by Simpack, and its implementation is based on the PCM [32]. Here we present a procedure to adapt the PCM approach to be suitable to our bearing case, referencing the Hertzian theory. The procedure is generally valid for modelling geometrical contacts in different bearings and in other applications as well. Defining contact parameters included three primary iterative processes. First, the polygon size for the contacting surfaces was selected, and second, the contact stiffness was adjusted to align with the well-established Hertzian contact theory. Lastly, the friction coefficient was iterated. In addition, the remainder contact parameters were defined.
The contacting components, REs, IR and OR, are assumed to have the same material properties and elastic modulus defined by Young’s modulus and Poisson’s ratio in Eq. (7). Using the PCM, polygonal discretisation is also expected to be identical on the contacting surfaces.

Polygonal discretisation affects the resolution of a model and normal contact stiffness, which in the elastic layer is linearly dependent on the polygon area (Eq. (10)). When it comes to the resolution, the more polygons, the more accurate results can be obtained at the cost of computational efficiency. The polygon size fundamentally limits the size of faults that can be detected. Besides, the size of the polygons can be thought to correspond to the real surface roughness of the contacting areas. In theory, an approximate polygon size can be calculated to match a certain surface roughness. In reality, at a steady-state condition, the surface roughness, including a possible tearout, is strongly damped by the lubrication, so it is not necessary to replicate such details with the simulation model for this study. Note that the replication with a realistic surface roughness would still be possible if the focus were on surface details with a magnitude of micrometres. The normal contact stiffness, however, imposes the primary constraint on the polygon size. Within this constraint, we adopted the Hertzian approach as a reference point to determine an optimal polygon size that is accurate enough to imitate a real bearing and still computationally reasonable to simulate.

The contact stiffness was resolved as a function of penetration depth using the scale function \( f(u) \) in Eq. (11). Since the material properties and geometrical polygon area are constant parameters, the elastic layer depth \( b \) and the stiffness scale function \( f(u) \) were adjusted so that contact stiffness aligned with the Hertzian contact stiffness. This was then repeated for multiple polygon sizes.

Unlike the Hertzian, the PCM assumes a linear relationship between contact pressure and penetration. Because these contact models are rather different, it is not possible to match contact pressure for all possible penetration values. Therefore, we determined a representative contact pressure value, defined by an average loading. The average normal contact force in the simulated bearing case was about 250 N for one RE. Using this force, we obtained the pressure of 1350 MPa and penetration of 4.5 \( \mu \)m through the Hertzian Eqs. (5) and (6), respectively. The derived pressure and penetration values were used in Eq. (8) to determine the elastic layer depth \( b \) for the PCM, resulting in 0.4 mm.

From Eq. (6), it can be seen that the general shape of the Hertzian stiffness, represented as contact force as a function of penetration, is formulated as a constant number multiplied by the penetration \( u_h \) to the power of 1.5. The stiffness functions for the RE–OR and RE–IR contacts were \( 24 \cdot u_h^{1.5} \) and \( 34.3 \cdot u_h^{1.5} \), respectively. In the PCM, the exponent of the polygonal area-related contact stiffness function exceeds 1.5 without additional scaling. To align these with the Hertzian, we used the stiffness scale function \( f(u) \) to accomplish the same function shape with the exponent of 1.5.

After fitting both Hertzian and PCM force-penetration data sets to the exponential function forms, the final stiffness scale function \( f(u) \) was derived as the ratio of the Hertzian function to the unscaled PCM function. After this procedure, the stiffness \( f(u) \) was adjusted so that the stiffness graph of the PCM followed the Hertzian stiffness at least at the derived normal force and penetration. Figs. 11a and 11b compare the Hertzian contact stiffness to PCM results derived separately for RE–OR and RE–IR contacts with three different polygon sizes of 0.1 mm, 0.3 mm and 0.5 mm. As the scaled functions were relatively close to each other, their deviations from the Hertzian is visualised in Figs. 11c and 11d. Table 3 lists the scale functions.

With the largest polygon size, the contact stiffness was no longer properly scalable which was noticeable particularly with the RE–IR contact. We observed that 0.5 mm polygons caused numerical problems, such as nonlinearities in the contact pressure. This was likely induced by the too sparse discretisation, making the contact occasionally too tight due to less clearance compared to RE–OR contacts but which can also be seen as a disturbance in Fig. 11b. Smaller polygon sizes of 0.1 mm and 0.3 mm were both feasible, but the latter was twice as fast to simulate. To respect the computational speed, we selected a polygon size of 0.3 mm for our components. Fig. 12 shows a close-up view of one RE–OR contact.

The damping properties for contacts were defined through compression and expansion area-related damping constants, based on [61,62]. The penetration range over which the transition from compression to expansion occurs was set by a transformation depth parameter. Damping parameters are challenging to define precisely as they are strongly associated with lubrication. However, the damping does not affect the main characteristics of the dynamic behaviour, such as natural frequencies and system mode shapes, which are the most relevant features for this study.

We selected the regularised Coulomb’s friction to apply the tangential forces. The friction model is regularised piecewise linearly with respect to the relative friction velocity and it takes a friction coefficient and the friction regularisation velocity as inputs. The velocity value sets the point, below which the friction force changes to a damping force. However, the friction coefficient value cannot be determined analytically and needs to be estimated. This was implemented by running bearing simulations in an iterative manner with a minimal preload that is just enough to maintain the contacts. The friction coefficient was adjusted until the cage slip, due to contact geometry and elastic deformation from load, was at around 1%, which represents the lower end for slip suggested in [47]. The resulting slip percentage calculation was based on the output angular velocity of the cage and the shaft. No other parameters than the friction coefficient were varied during the iteration. We then used the derived friction coefficient value in

<table>
<thead>
<tr>
<th>Polygon size</th>
<th>RE–OR</th>
<th>RE–IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 mm</td>
<td>( 1.23 \cdot 10^6 \cdot u^{-0.68} )</td>
<td>( 1.18 \cdot 10^6 \cdot u^{-0.68} )</td>
</tr>
<tr>
<td>0.3 mm</td>
<td>( 2.26 \cdot 10^6 \cdot u^{-0.85} )</td>
<td>( 4.08 \cdot 10^6 \cdot u^{-0.90} )</td>
</tr>
<tr>
<td>0.5 mm</td>
<td>( 28.92 \cdot 10^6 \cdot u^{-2.45} )</td>
<td>( 127.04 \cdot 10^6 \cdot u^{-1.99} )</td>
</tr>
</tbody>
</table>
Fig. 11. Subfigures (a) and (b) show the adjusted contact stiffnesses for RE–OR and RE–IR contacts with different polygon sizes compared to contact stiffness by the Hertzian theory. The corresponding deviations from the Hertzian are depicted in Subfigures (c) and (d).

In addition to the aforementioned parameters, the maximum penetration variable $u_{\text{max}}$ sets the numerical limit for the contact description [53]. This parameter is mostly to ease the computational performance rather than being related to real-world physics. The parameters used in our bearing contacts are listed in Table 4. Note that the values have been specifically tuned for this particular case, and they do not represent any general ball bearing with different dimensions or different loading conditions.

Based on the exact geometry of the REs and the raceways, the inherent maximum contact angle of this particular bearing obtained from the simulation model was eight degrees. The value is needed to calculate the cage slip percentage and the theoretical characteristic fault frequencies. The contact angle can be assumed to be at its maximum because the axial load is high enough to cause the maximum axial shift to the IR.

Fig. 12. Polygonal contact between an RE and OR.

our actual simulations with the full axial preload. This also represents a simplified lubrication as the detailed lubrication was omitted in this study.

In addition to the aforementioned parameters, the maximum penetration variable $u_{\text{max}}$ sets the numerical limit for the contact description [53]. This parameter is mostly to ease the computational performance rather than being related to real-world physics. The parameters used in our bearing contacts are listed in Table 4. Note that the values have been specifically tuned for this particular case, and they do not represent any general ball bearing with different dimensions or different loading conditions.

Based on the exact geometry of the REs and the raceways, the inherent maximum contact angle of this particular bearing obtained from the simulation model was eight degrees. The value is needed to calculate the cage slip percentage and the theoretical characteristic fault frequencies. The contact angle can be assumed to be at its maximum because the axial load is high enough to cause the maximum axial shift to the IR.
### Table 4
Contact parameters used for the PCM.

<table>
<thead>
<tr>
<th>Denotation</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{1,2}$</td>
<td>Poisson’s ratio</td>
<td>0.3</td>
</tr>
<tr>
<td>$E_{1,2}$</td>
<td>Young’s modulus</td>
<td>208 GPa</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>Elastic layer depth</td>
<td>0.4 mm</td>
</tr>
<tr>
<td>$c_{c}$</td>
<td>Compression-related damping</td>
<td>400 MNs/m²</td>
</tr>
<tr>
<td>$c_{e}$</td>
<td>Expansion-related damping</td>
<td>60 MNs/m²</td>
</tr>
<tr>
<td>$t_{d}$</td>
<td>Transformation depth for damping</td>
<td>1 μm</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Friction coefficient</td>
<td>0.0015</td>
</tr>
<tr>
<td>$v_{f}$</td>
<td>Regularised velocity for friction</td>
<td>0.002 m/s</td>
</tr>
<tr>
<td>$u_{\max}$</td>
<td>Maximum penetration</td>
<td>0.5 mm</td>
</tr>
</tbody>
</table>

#### 5.4. Time integration

The varying parameters of the simulation model are simply the velocity request for the PI controller in all given simulation runs and, in addition, the initial position angle of the OR in the faulty cases. We run three simulations with the healthy OR and nine with the faulty OR, replicating the experimental case (Table 2). The duration of the simulated period is eight seconds, starting from zero speed. From these, three-second sections at steady-state velocities are extracted.

Time integration is used for solving the model in the time domain to obtain the complete nonlinear dynamic behaviour of the full system. We use a backward differentiation formula integrator, an implicit multistep integration scheme based on a solver type whose benefits are robust and fast performance for low orders 1–2 as well as variable step size and orders 1–5. This is called SODASRT 2 in Simpack. Even though the variable step size records samples according to how rapidly the model stages are changing, the output data is sampled with the fixed step size user-defined by the sampling rate, which in our case is set to the same 20 kHz used for the capacitive sensors.

#### 6. Data and processing

All in all, three healthy and nine faulty scenarios, including three different OR fault positions, were obtained from measurements and simulations with the varying input parameters being the angular shaft velocity and the angular position of the OR fault. The three-second displacement data from both capacitive measurements and simulated shaft position is presented in micrometres. With the 20 kHz sampling frequency, each data consists of 60,000 samples.

The time series data, namely the waveform signal, records the movement of the rotating shaft in the time domain, where the amplitude is related to the loading and limited by the clearance between the REs and ring grooves. We carry out the X- and Y-direction displacement waveforms separately. By presenting them together as a polar graph, we can visualise the radial shaft movement, often referred to as a rotor path. As the mechanical runout has a significant influence on the experimental measurements, especially on the time-dependent waveform and rotor path, it was compensated for one particular healthy measurement case at 1500 rpm to have a closer comparison.

For both measured and simulated data processing, we applied an identical Hanning’s window and FFT to analyse their frequency content, a popular way for bearing fault analysis. Hanning windowing was used to reduce the impact of spectral leakage and to improve the accuracy of our frequency analysis. We applied three windows over the data, with a 50% overlap, resulting in a 0.67 Hz frequency resolution. The use of overlapping windows ensures that each time point is included in multiple windows, which improves the robustness of the analysis and helps to reduce noise in the data. FFT was utilised to convert the waveform signals to the frequency domain and deconstruct signals into separate frequencies.

By applying FFT, we can gain insights into the frequency spectrum and identify any patterns or trends, particularly in constant-velocity steady-state data. The frequency domain data is not as prone to external noise as the waveform signal, which is why it is more convenient in fault analysis. The OR fault is expected to appear as frequency components at the characteristic fault frequency, as in Eq. (1), and its harmonics.

#### 7. Results

We divide the validation process of our MBS model into the healthy case and the faulty case. Initially, the healthy case was validated in the time domain by analysing both measured and simulated rotor paths and waveform signals. Our healthy bearing analysis focused mainly on the velocity, from which we also have runout-compensated data. Subsequently, frequency spectrums for healthy cases were carried out.

For the faulty case, our interest mostly lies in the frequency spectrums by the FFT analysis and the BPFOs at the given velocities. Additionally, to demonstrate the general capability of our simulation model to handle bearing slip, we show how it can be identified through the fault frequencies. Finally, we present the computational results.
Fig. 13. Rotor paths from the healthy cases at 1500 rpm. Subfigures (a) and (b) are the graphs before and after the runout compensation. In turn, (c) and (d) are the simulated counterparts with normal and tenfold unbalance, respectively.

Fig. 14. Healthy displacement waveforms at 1500 rpm from the measured runout-compensated and simulated models with and without the tenfold rotor unbalance mass.

7.1. Validating the healthy bearing

We analysed the healthy time domain data more closely, particularly at 1500 rpm. Fig. 13 presents subgraphs a–c, which illustrate the three rotor paths derived from the measured data both before and after compensating for the runout, as well as the simulated path. In the original measured data, the shaft displacement is about ±8 μm, whereas after the compensation, it was reduced by half to ±4 μm. Despite the compensation, the measured displacements still remained higher than the corresponding simulated values, which fall below ±2 μm.

Our interpretation is that because it is challenging to completely remove runout through post-processing, the remaining effects continue to increase the displacements. An additional cause could be the NDE bearing that is installed axially free, meaning that its position likely adapts depending on the other component tolerances (a small angle of the outer ring with respect to the X or Y axes). These reasons may also be attributed to the non-round shape of the measured rotor paths. Furthermore, it should be noted that the ideal rigid body model does not currently account for all real-world inclinations, misalignments or eccentricities, which could stem either from the motor or from the load machine. Although the given phenomena are replicable, we do not have enough information to do so. Without the periodical irregularities in the rotation, the axial preload presses evenly all RE contacts aligning and centring in the middle, which results in the tight and uniform shape of the simulated rotor path. The measured runout compensated and simulated displacement waveforms of the healthy bearing case are depicted with blue and orange colours in Fig. 14, respectively. The difference in rotor path sizes is also shown here in the waveform amplitudes.

Based on the measured data, it seems evident that the electrical machine is subjected to some additional loads causing the larger displacement amplitude. It is also reasonable to suspect that there has been more unbalance in the measurements than allowed by the balancing class. The effect occurs at the rotational frequency, which implies an unbalance-related phenomenon naturally existing in experimental set-ups. To demonstrate this, we conducted a separate test where the rotor unbalance was increased tenfold compared to the G2.5 balance class. As a result, the new simulated shaft displacement waveform, marked with dark grey in Fig. 14, resembles the measured waveform both in shape and amplitude. The corresponding rotor path was obtained (Fig. 13d), which also confirms that the bearing clearance can be included to the dynamics simulation using the polygonal contact. This indicates that with correct
loading definitions, the simulation model is capable of imitating the dynamics of the experimental setup. However, in our case, it is not certain which details cause the eccentricity phenomenon similar to the tenfold rotor unbalance. Some speculations may relate to the runout compensation method or the load machine causing some external loads. For this reason, all of the following simulation cases were run with the unbalance that was defined according to the rotor balancing standard.

We conducted separate frequency analyses for both X- and Y-displacements. Nonetheless, to simplify the analysis, only the most relevant for each case is reported here. Therefore, Fig. 15 shows the healthy amplitude spectrums of the shaft displacements for all given velocities in the Y-direction, which is naturally more subjected to the gravitational force. The rotational frequency components $f_r$ and their harmonics are marked with vertical dotted lines. The BPFOs are also marked with grey to indicate that there are no distinctive amplitudes at those frequencies. Table 5 lists the rotational frequencies and the theoretical fault frequencies for the different velocities.

At lower frequencies, the measured displacement spectrums exhibit an elevated intensity level compared to the simulated spectrums. This phenomenon can be partially attributed to the factors previously discussed, such as lubrication discrepancies and eccentricity resulting from runout or unbalance, among others. These factors can be expected to exert an influence, particularly near the rotational frequencies. Nevertheless, at higher frequencies, where the fault frequencies occur, the intensity levels of both the measured and simulated spectra seem to align more closely.

### 7.2. Validating the faulty bearing

Similarly to the healthy case, we applied the FFT to the displacement data either in the X- or Y-direction, depending on which is more sensitive for each type of bearing fault. Specifically, when the fault is on the side of the bearing (at 270°), the lateral X-displacement provides a more informative measure of the fault than the vertical Y-displacement. On the other hand, when the fault is located at the top or bottom of the bearing (at 0° or 180°), the vertical Y-displacement is a more useful measure of the fault.
The frequency analysis results for faulty bearings at 750, 1125 and 1500 rpm are presented at the three distinct velocities in Figs. 16, 17 and 18, respectively. The rotational frequency components and the characteristic OR fault frequencies BPFOs and their harmonics are again marked with dotted lines. At each velocity, we compare the simulated shaft displacement spectrum to the measured spectrum with varying OR fault positions. To note, the measured spectrums were derived from the non-compensated raw displacement data.

Examining both the measured and simulated spectra of shaft displacements, it seems evident that the OR fault at 180° results in the most distinct amplitudes at the BPFOs and their harmonics across all three rotational velocities. This can be attributed to the gravitational force which inherently sustains a stronger contact between the REs and the fault when the fault is located at the bottom. However, when the fault is located at 0° or 270°, there is a noticeable decrease in the amplitudes at the BPFOs, particularly in the measured spectrum. This is explained by the fact that at these velocities, the centrifugal forces are still insufficient to cause significant fluctuations in the shaft position, which are necessary for maintaining proper contact between the faulty raceway and the REs. Additionally, it is likely that faulty signals are somewhat attenuated by lubrication or external noise sources, which are not incorporated into our dynamic bearing simulations in detail. As a result, regardless of the OR fault location, the fault frequencies appeared more distinctly in the simulated spectrums.

In addition, the measured fault frequencies seem to appear at moderately lower frequencies compared to the theoretical values, which insinuates slipping. The amount of slipping also increases with higher velocities, typical in dynamic behaviour. With the simulated spectrums, the BPFOs occur rather accurately at the expected values and with clear amplitudes, although the slip is less visible than in the measured spectrums.

To demonstrate the slip more closely in the simulated results, Fig. 19 shows the same simulated frequency spectrums as what was presented in Figs. 16–18. But now the results are grouped by velocities, and axis breaks are used to zoom in on the theoretical BPFOs and their two harmonics, again marked with black dotted lines. The healthy displacement spectrums in both the X- and Y-direction were added for comparison. The results imply that also in the simulated spectrum, the slip appears larger as the velocity increases. In this figure, we can also see the same as seen with the measured spectrums, which is that, at most frequencies, the fault at the bottom caused slightly higher characteristic fault frequency amplitudes than the faults located at the top or the side.

It is again noteworthy that the bearing model simulates contacts with simplified lubrication, where the slip is naturally smaller than the real lubricated contact. Nonetheless, the amount of slip was not validated in this study because of the simplified lubrication description and the lack of sufficient validation data to develop our model for more accurate lubrication. Table 6 lists the theoretical cage velocities and the simulated counterparts with the corresponding slip percentage. The values were derived from the healthy
Table 6
Simulated cage speeds compared to the theoretical cage speeds, and the corresponding slip percentages.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>306.07</td>
<td>305.52</td>
<td>0.18</td>
</tr>
<tr>
<td>1125</td>
<td>459.11</td>
<td>458.21</td>
<td>0.19</td>
</tr>
<tr>
<td>1500</td>
<td>612.14</td>
<td>610.85</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 7
Resulting computing times for the simulated cases.

<table>
<thead>
<tr>
<th>Velocity [rpm]</th>
<th>Condition [healthy/faulty]</th>
<th>Duration of the simulated period [s]</th>
<th>Mean step size [μs]</th>
<th>Clock time [h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>healthy</td>
<td>8</td>
<td>2.27</td>
<td>9.54</td>
</tr>
<tr>
<td>750</td>
<td>faulty</td>
<td>8</td>
<td>2.27</td>
<td>9.21</td>
</tr>
<tr>
<td>1125</td>
<td>healthy</td>
<td>8</td>
<td>1.91</td>
<td>11.61</td>
</tr>
<tr>
<td>1125</td>
<td>faulty</td>
<td>8</td>
<td>1.94</td>
<td>10.87</td>
</tr>
<tr>
<td>1500</td>
<td>healthy</td>
<td>8</td>
<td>1.70</td>
<td>12.94</td>
</tr>
<tr>
<td>1500</td>
<td>faulty</td>
<td>8</td>
<td>1.71</td>
<td>12.39</td>
</tr>
</tbody>
</table>

cases at the three distinct velocities. Although the slip percentages are comparatively small, the trend is that it increases with the shaft velocity, which is plausible.

The simulations were run using a desktop computer equipped with an Intel Xeon W-2255 processor. Table 7 lists the final computing times, run with four cores. Note that the simulating times did not markedly differ between the faulty cases with different OR fault positions, so average values of the three given faulty cases per velocity are drawn here.

In general, simulation models are powerful tools for exploring and predicting complex systems, but they require plenty of input parameters to function properly. Some of these parameters may not have clear objective values if they cannot be directly measured or referred to as unambiguous values, which means that they must be estimated with a trial and error technique to achieve the best possible results or be excluded for simplification. Variables, such as the friction coefficient, the rotor unbalance or misalignments
8. Conclusion

We presented an approach that employs the geometry-based polygonal contact method to simulate ball bearing dynamics. We presented a procedure to adjust the contact stiffness to ensure its alignment with the well-established Hertzian contact theory. To the best of our knowledge, this study is the first to present a surface geometry-based polygonal contact in a multibody simulation model of a bearing application. The model was used to analyse an outer ring fault with varying positions and at varying velocities by generating radial shaft displacement data. The simulation results were evaluated against experimental displacement data measured by a capacitive position sensor.

The experimental case was successfully replicated within a reasonable computation time of 10–12 h, depending on the shaft rotation speed. Our simulations proved that the polygonal contact method can be adapted specifically for bearings, both healthy and faulty, as long as the selected polygon size is small enough to carry out a certain accuracy. The polygonal contact method produced plausible results, successfully imitating real-world phenomena such as clearance and slipping, whose detailed modelling was made possible through the utilisation of the polygonal contact method.

In the time domain, the displacement amplitudes were noticeably smaller than in the measured data due to unknown unbalance. Modifying the unbalance in the simulation model, however, improved the results. This also increased the rotor path pattern which confirms the clearance. In the frequency domain, the displacement spectrums showed the same fault frequencies and their harmonics as the experimental data, particularly when the fault was located at the bottom. The measured frequency level was overall slightly higher near the rotational frequencies, possibly due to minor, yet inevitable, differences between the experimental and simulated system, such as background noise and motor runout that were disregarded in the simulations. Additionally, the existence of slipping...
Fig. 19. The simulated shaft displacements of the faulty cases in the frequency domain. A zoomed view of the BPFOs and their harmonics.

phenomena in the simulation results was demonstrated by a slower cage velocity than the theoretical velocity, although the amount of slip was underestimated and not validated due to the simplified lubrication description and lack of validation data.

The main challenge in creating accurate simulation models lies in the deep understanding of the complex nonlinear dynamics of simulated systems. Despite the uncertainties and unknown parameters in the measured system, our model demonstrated the capability to generate realistic data of a ball bearing, which can be elaborated for further research and analysis. This study highlights the potential of our simulation model as a valuable tool in studying ball bearing dynamics and faulty cases. Overall, the use of the polygonal contact method offers several practical advantages in bearing simulations, including the ability to accurately represent complex and arbitrary geometries and, therefore, more realistic bearing dynamics. These benefits can contribute to a better understanding of the behaviour of the system under different conditions.

Future research can further refine the model to address the time domain interpretations and enhance its accuracy in simulating real-world bearing dynamics — that is, adding even more degrees of freedom to the system by including contacts between rolling elements and the cage as well as including a more detailed lubrication description. Moreover, flexible parts, such as the OR, can be modelled to obtain vibrations for further analysis. In future work, more realistic data can also be generated from different real fault cases with arbitrary geometries and under varying loading conditions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.
Acknowledgements

The authors acknowledge the financial support from the Research Council of Finland through two projects: ‘Intelligent Techniques in Condition Monitoring of Electromechanical Energy Conversion Systems’ (ESTV) and ‘The Center of Excellence in High-Speed Electromechanical Energy Conversion Systems’ (HiECSs) with the decision numbers 331199 and 346441, respectively. Additionally, this research received funding from the EU ECSEL Joint Undertaking under ‘Arrowhead Tools’ project with a decision number 826452.

References


