A new computationally efficient model of the non-linear dynamics in harmonic drive reducers

Mechanism and Machine Theory 209 (2025) 105992

Contents lists available at ScienceDirect

Mechanism and Machine Theory

journal homepage: www.elsevier.com/locate/mechmt

Research paper

A new computationally efficient model of the non-linear dynamics in harmonic drive reducers

Roberto Guida *, Antonio Carlo Bertolino, Andrea De Martin, Massimo Sorli

Politecnico di Torino, Deparment of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi, 24, Torino, 10129, Piemonte, Italy

ARTICLE INFO

Keywords:
Harmonic reducer
High-fidelity modeling
Non-linear modeling
Defects modeling

ABSTRACT

Harmonic drives, known for their high precision and compactness, are widely used in applications such as robotics and aerospace. Due to their design, harmonic drives exhibit complex nonlinear dynamics that cannot be described through the traditional equations employed for conventional gear transmissions. Such non-linearities heavily impact both the static and the dynamic performances of the system, requiring in-depth investigation to predict and optimize its expected behavior. This paper presents an accurate and computationally efficient dynamic model able to capture non-linear dynamic effects such as the kinematic errors, hysteresis, internal friction and tooth meshing behavior. Additionally, the proposed model incorporates non-ideal factors such as the internal defects, wear, erroneous assembling and improper lubrication conditions, allowing for a comprehensive simulation of real-world conditions. Results demonstrate the model's ability to replicate the response of the real component across diverse operational scenarios, validated against literature data. The accuracy of the results together with the reduced computational burden make this model a valuable tool for those applications requiring a good fidelity and computational efficiency, such as model-based design applications or prognostic and health management systems.

1. Introduction

A harmonic drive is a specialized gear transmission system renowned for its high precision, compact size, and high torque transmission capabilities, making it an essential component in applications requiring precise motion control. The harmonic drive consists of three primary components: the wave generator, the flexspline, and the circular spline. The wave generator, typically an elliptical cam, deforms the flexible flexspline, which meshes with the rigid circular spline. This deformation allows for smooth, precise motion and high reduction ratios. Its unique operating principle makes the harmonic drive ideal for fields such as robotics, aerospace, medical devices, and industrial automation, where space constraints and accuracy are critical [1].

Harmonic drives have several notable characteristics, including zero backlash, high efficiency, and exceptional positioning accuracy. However, these benefits come with inherent challenges, particularly the complex non-linearities arising from kinematic errors, hysteresis, and tooth meshing dynamics [2]. As a result, accurately modeling the harmonic drive's behavior, both in nominal and faulty conditions, is crucial for enhancing its reliability in applications requiring stringent operational performance.

In recent years, the importance of Prognostics and Health Management (PHM) systems has grown significantly, especially in industries that rely on high-precision equipment. PHM focuses on predicting and preventing faults before they occur, minimizing downtime and ensuring operational efficiency. Data-driven models (DDMs) have emerged as powerful tools in PHM, as they utilize

* Corresponding author.
E-mail address: roberto.guida@polito.it (R. Guida).

https://doi.org/10.1016/j.mechmachtheory.2025.105992
Received 4 November 2024; Received in revised form 22 January 2025; Accepted 7 March 2025
Available online 17 March 2025

operational data to extract health indicators and predict the remaining useful life of components [3]. However, the scarcity of real-world fault data — due to the high reliability and long operational life of harmonic drives — creates a significant challenge for DDMs [4,5]. A robust model capable of simulating both nominal and fault-induced behaviors is necessary to create synthetic data for training machine learning algorithms, enabling predictive maintenance strategies.

Various models of harmonic drives have been proposed in the literature. Early models focused on kinematic errors and simplified empirical relationships, which provided useful insights but lacked the accuracy needed for fault prediction [1,6]. High-fidelity models, such as those based on finite element analysis (FEA), offer detailed simulations of gear mechanics but are computationally expensive, making them unsuitable for real-time PHM applications [7]. More recent models, such as those by Zou et al. [8] and Zhang et al. [9], incorporate refined dynamic responses and account for geometric and internal interactions, but they still fall short in replicating certain fault conditions.

The model developed in [1] focuses on characterizing the mechanical behavior of the harmonic drive from a kinematic perspective. It describes the interaction between the wave generator, flexspline, and circular spline and their influence on gear reduction and precision. The proposed approach uses a lumped parameter model that treats the harmonic drive as a series of interconnected bodies, simplifying the complex internal interactions into more manageable mathematical expressions. The model accounts for basic non-linearities, such as gear meshing stiffness and backlash, but does not deeply address the influence of faults or detailed dynamic responses. This work is fundamental for understanding nominal behavior, but it simplifies the system, especially when compared to more advanced models available today.

In the follow-up work published in 1996, Tuttle and Seering extend the model by incorporating non-linearities into the analysis [6]. This model addresses the more complex behaviors of harmonic drives, including hysteresis, frictional effects, and kinematic errors. The focus here is on developing a non-linear model that better captures the dynamic interaction between the components of the drive during operation. A way to model gear flexibility, especially in the flexspline, is introduced as well as the internal stresses and strains that occur during gear meshing. This model provides deeper insights into the drive's performance under various operating conditions, making it more suitable for analyzing how the system might behave under load variations or minor mechanical deviations. However, while this model improves on its predecessor by introducing non-linearities, it still has limitations in terms of predicting behavior under fault conditions. The focus remains more on normal operational performance rather than fault detection or health management.

The model proposed by Rhéaume in [7] focuses on analyzing the torsional stiffness of harmonic drives using a high-fidelity Finite Element Analysis (FEA) approach. The authors aim to provide an accurate representation of the mechanical behavior of the harmonic drive, specifically in terms of its torsional stiffness, which is a critical factor in its overall performance and precision. Despite its accuracy, the FEA model is computationally intensive, making it less suitable for real-time applications or integration into PHM systems. While it offers detailed insights into mechanical behavior under nominal conditions, its high computational cost limits its practicality for simulating rapid fault conditions or performing extensive real-time fault diagnostics.

In [8] a harmonic drive model that emphasizes the importance of geometry and internal interactions is proposed. The model integrates detailed representations of the wave generator, flexspline, and circular spline, accounting for precise geometric factors and the internal interaction forces between these components. By capturing the contact mechanics and elastic deformation of the flexspline, the model simulates meshing behavior more accurately. It also incorporates non-linear dynamic responses to phenomena such as backlash, contact non-linearities, and elastic properties, offering a more realistic analysis of the harmonic drive's behavior under different operating conditions. Despite these improvements in capturing detailed internal dynamics, the model remains computationally feasible, balancing accuracy and efficiency, making it suitable for design optimization and fault analysis, though not ideal for real-time diagnostics or fault detection.

A refined dynamic model focusing on the dynamic response and vibration analysis of harmonic drives under operational conditions represents the next improvement [9]. This model explores the harmonic drive's dynamic stiffness and compliance, emphasizing the elastic deformation of the flexspline under dynamic loads. A novel aspect of this work is its analysis of the harmonic drive's vibration modes and resonance phenomena, crucial for identifying operational inefficiencies and faults. The model incorporates non-linear dynamics related to contact forces and torsional effects and simulates fault conditions like wear and misalignment, providing insights into how these faults alter the system's dynamic response. While effective for detailed dynamic analysis and fault simulation, its computational complexity limits its real-time application, making it more suitable for offline diagnostics and system optimization.

Advanced mathematical models have been developed to capture the harmonic reducer dynamics, incorporating elements like dual mass systems and compliance mechanisms to simulate real-world behavior [10,11]. Recent studies highlight the use of finite element-inspired methods to represent the dynamic response of individual components, enhancing the fidelity of simulations [12]. Such models have been integrated into simulation frameworks like MATLAB/Simulink, enabling detailed analysis of system parameters and control strategies, such as active disturbance rejection control for minimizing position tracking errors. These advancements offer a robust foundation for designing more precise, reliable, and efficient harmonic drive-based systems across industrial and service robotics [11,12].

The model proposed in this paper introduces a computationally efficient planar model of the harmonic drive, specifically designed to balance accuracy and feasibility for real-time fault detection and predictive maintenance applications. Unlike previous models that either lacked fault simulation capabilities or were computationally intensive, this model integrates key non-linearities such as wave generator misalignment, flexspline deformation, and tooth meshing characteristics, while also simulating fault conditions like tooth wear, flexspline tooth cracks, and rim cracks. These features allow the model to accurately capture both nominal and non-nominal behaviors, providing a detailed understanding of the dynamic interactions within the drive, even under fault conditions.

The model's efficiency makes it ideal for use in PHM systems, where DDMs require large data sets representing fault behaviors to train machine learning algorithms. By generating fault scenarios and replicating real-world operational conditions, this model creates a data lake essential for predictive maintenance strategies. Moreover, the planar approach significantly reduces computational cost compared to FEA models like those in [7], while offering more detailed fault analysis than the lumped-parameter models in [1,8].

While data-driven approaches, such as those utilizing deep learning algorithms (e.g., CNN-LSTM or mixed convolutional neural networks), have gained significant traction in fault detection, the proposed physics-based model offers unique advantages. Data-driven methods rely heavily on large datasets for training, which may not always be available or sufficiently representative of the full spectrum of operating conditions and fault scenarios. Additionally, these methods often act as "black boxes", providing limited insight into the underlying physical phenomena.

In contrast, the proposed model is grounded in the physics of harmonic drive dynamics, allowing for detailed analysis of specific fault mechanisms, such as crack-induced stiffness variations or wear-related meshing degradation. This level of interpretability is crucial for understanding fault propagation and designing targeted maintenance strategies. Furthermore, the model's computational efficiency makes it suitable for generating synthetic datasets, which can complement data-driven methods by addressing gaps in available data. By combining the strengths of physics-based and data-driven approaches, this methodology lays the foundation for hybrid diagnostic frameworks that leverage the predictive accuracy of machine learning with the explanatory power of physical models.

The results, validated against experimental data and literature [2,13,14], demonstrate the model's effectiveness in capturing the dynamic response of harmonic drives under various conditions. This model presents a significant step forward in the development of predictive maintenance systems, enabling real-time fault detection and enhancing the reliability of harmonic drives in critical applications.

2. Mathematical model of the harmonic reducer

The proposed model employs a multibody approach that treats each component of the harmonic reducer as an individual body, considering external loads, internal interactions between the components and friction [15]. This comprehensive framework facilitates a detailed analysis of the dynamics within the reducer allowing to describe not only the macroscopical behavior of the overall system.

The core concept of the model is to represent the interactions in the two meshing zones of the harmonic drive using equivalent wedges within an Eulerian reference system. This reference frame follows the rotation of the wave generator, which, in the chosen configuration, is linked to the reducer's input shaft. The model considers three key components (Fig. 1): the wave generator, the flexspline, and the circular spline. The system is characterized by four degrees of freedom, with the free variables being the wave generator rotation, the flexspline rotation and radial deformation, and the circular spline rotation.

The concept of "equivalent wedges" provides a simplified yet effective representation of the dynamic interactions within the harmonic reducer. In the proposed model, the meshing zones between the flexspline and the circular spline, as well as the deformation induced by the wave generator, are modeled as wedges within an Eulerian reference frame. This reference frame rotates with the wave generator, enabling efficient capture of the periodic engagement of the components.

The chosen four-degree-of-freedom configuration enables the simulation of various operational setups, including cases where either the flexspline or the circular spline is fixed. This flexibility ensures that the model can accurately represent a wide range of real-world harmonic drive configurations. Moreover, the additional degrees of freedom enhance the model's capacity to capture complex dynamic phenomena. This approach also accommodates a broader spectrum of nonlinearities and fault types, including misalignment, wear, and rim crack, conditions that are difficult to model with simpler approaches.

The model begins by focusing on a single interaction zone, which is the merge of the two individual interaction zones between the teeth. This zone encounters a specific segment of the fixed flexspline twice per revolution of the wave generator, capturing the periodic engagement of the components.

This approach accurately represents the load distribution and interaction forces, which are critical for assessing the harmonic reducer's performance under both nominal and faulty conditions.

The chosen configuration of the reducer presents the input shaft linked to the wave generator, the output shaft linked to the circular spline and the flexspline fixed, in this particular configuration the reduction ratio is equal to:

$$\tau = \frac{\dot{\theta}_{\mathrm{WG}}}{\dot{\theta}_{\mathrm{CS}}} = \frac{Z_{\mathrm{CS}}}{Z_{\mathrm{CS}} - Z_{\mathrm{FS}}}; \tag{1}$$

where:

- $\dot{\theta}_{CS}$ is the circular spline rotational speed;
- $\dot{\theta}_{WG}$ is the wave generator rotational speed;
- Z_{CS} is the number of circular spline teeth;
- Z_{FS} is the number of flexspline teeth.

The rotational motion of the elliptical cam (θ_{WG}) is translated into a horizontal displacement (y_{WG}) via an equivalent radius (r_{HR}) that represents the point at which interaction between the wave generator, the flexspline, and the circular spline occurs. As the elliptical cam rotates, the horizontal coordinate tracks the continuous displacement of the wedge representing the cam. The wedge associated with the cam is kinematically linked to the wedge representing the inner ring of the bearing (x_{WG}), which induces

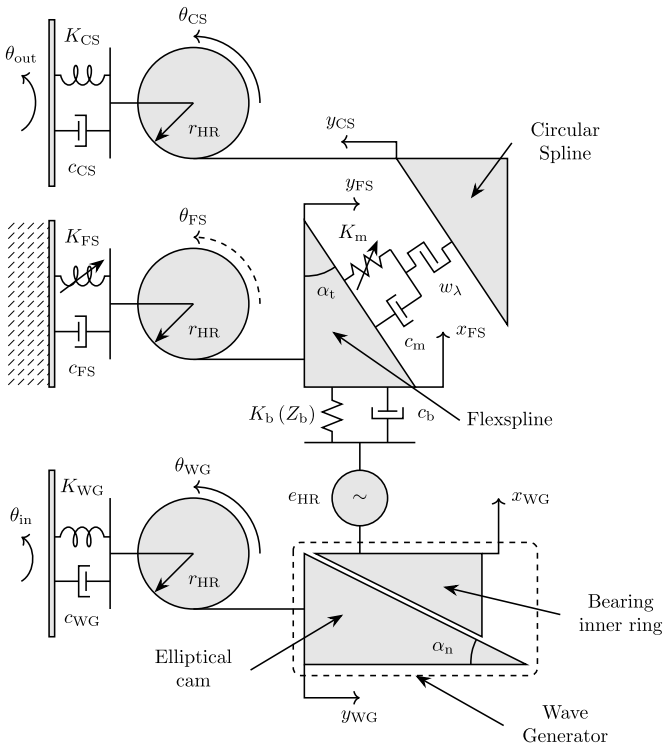


Fig. 1. Harmonic reducer lumped parameters model.

deformation in the outer ring that is rigidly connected to the flexspline and can move only vertically. The transmission ratio in the wave generator-flexspline interaction is described by the equivalent cam angle (α_n).

The teeth of the flexspline can undergo both horizontal movement (y_{FS}), which corresponds to a rotation of the element, and vertical movement (x_{FS}), associated with radial deformation. While the tangential displacement is restricted by the reaction force of the fixed flexspline, the radial displacement is free to continue, leading to the tangential movement of the wedges representing the teeth of the circular spline (y_{CS}). These wedges are free to move tangentially, and this motion can ultimately be converted into the rotational movement of the circular spline itself (θ_{CS}).

From a model point of view the system's inputs are the positions of the input and output shafts. Considering a typical configuration, the input shaft position corresponds to the angular position of the electric motor shaft, while the output shaft position represents the user actual position. The primary outputs of the model are the torques transmitted by both the input and output shafts.

To accurately calculate the dynamic interactions within the system, spring–damper systems of varying fidelity levels are exploited, specifically the model evaluate the following forces and moments:

- Input Torque: The torque applied to the input shaft.
- Restoring Torque of the Flexspline: The torque exerted on the flexspline by the fixed frame.
- Output Torque: The torque transmitted from the reducer to the output shaft.
- Bearing Radial Force: The radial force exchanged between the wave generator and the flexspline through the elliptical bearing.
- Bearing Friction: The frictional forces acting on the bearing.
- Meshing Force: The forces resulting from the interaction between the teeth of the flexspline and the circular spline.
- Meshing Friction: The frictional forces occurring at the gear tooth interfaces.

This multibody approach offers several advantages: by considering each component as a separate body, the model can capture complex interactions and accurately reflect the behavior of the harmonic reducer under various operating conditions. The use of modified spring–damper systems allows for high-fidelity simulations that can account for both linear and nonlinear dynamic effects, providing a more realistic representation of the system's performance. The detailed outputs of the model, including torque and force measurements, can inform predictive maintenance strategies, potentially extending the lifespan of the harmonic reducer and preventing unexpected failures.

2.1. Load arc

The proposed digital replica combines the two interaction zones along the semi-major axes of the wave generator into a single, unified interaction zone. These two areas of engagement define a load arc (β), as illustrated in Fig. 2, within which meshing occurs

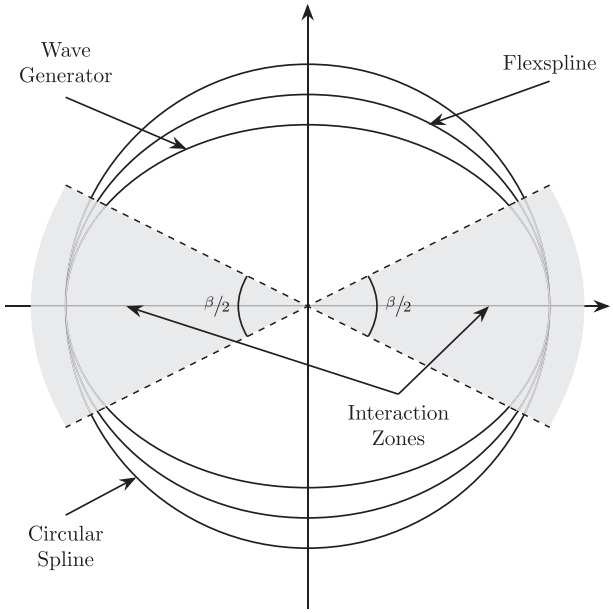


Fig. 2. Interaction zones and load arc.

between the flexspline and the circular spline [16]. This approach simplifies the modeling of the harmonic reducer by treating the interaction zones collectively, providing a more streamlined and efficient analysis of the bearing interaction between wave generator and flexspline and of the meshing dynamics between the flexspline and circular spline.

The unified load arc ensures accurate representation of the load distribution and interaction forces, crucial for understanding the performance and reliability of the harmonic reducer under various operating conditions. In the scientific literature [17–20] the percentage of meshing teeth in a harmonic drive reducer typically ranges between 20% to 40%. This value is a significant factor in achieving the high torque transmission and low backlash properties of harmonic drives, as the meshing teeth are distributed across a larger contact area compared to conventional gears. The actual percentage may vary depending on the specific design of the gear and operational conditions.

It is possible to define a dimensionless tangential coordinate within each interaction zone, which varies from −1 to 1, spanning from one end of the load arc to the other. This coordinate normalizes the position along the interaction zone, providing a consistent framework for analyzing the distribution of forces across the interaction zone.

A first way to determine the load arc is to identify the intersection between a circle with a radius equal to that of the circular spline and an ellipse constructed using the dimensions of the deformed flexspline. Alternatively, literature data can be used to establish a curve that describes the relationship between the applied load and the loading arc of one interaction zone [18], as shown in Eq. (2).

$$\beta = \beta_{\min} + \left(1 - \frac{\frac{2}{3}}{1 + 10 \left| \frac{Q_{\text{red,out}}}{Q_{\text{red,max}}} \right|} \right) \cdot \left(\beta_{\max} - \beta_{\min} \right); \tag{2}$$

where:

- $\beta_{\min}$ and $\beta_{\max}$ are, respectively, the minimum and maximum theoretical loading arcs for the considered application;
- $Q_{\text{red,out}}$ is the external torque on the output axis.
- $Q_{\text{red,max}}$ is the maximum available torque on the output axis obtained from the hysteresis test.

The proposed formula is compared to literature data in Fig. 3. In this graph, it is evident that the number of teeth in the meshing zone increases rapidly with respect to the load, reaching a plateau after a certain level of output torque.

The minimum $(\beta_{\min})$ and maximum $(beta_{\max})$ theoretical loading arcs are determined based on the geometric and operational constraints of the harmonic reducer. These parameters define the range of meshing engagement between the flexspline and the circular spline under varying load conditions.

The loading arc is influenced by the deformation of the flexspline and the external load applied to the system, as described in Eq. (2). Specifically, $\beta_{\min}$ represents the loading arc under no-load conditions, while $\beta_{\max}$ corresponds to the loading arc at the maximum output torque. These values are derived from the literature [18], and the loading arc is modeled using empirical data.

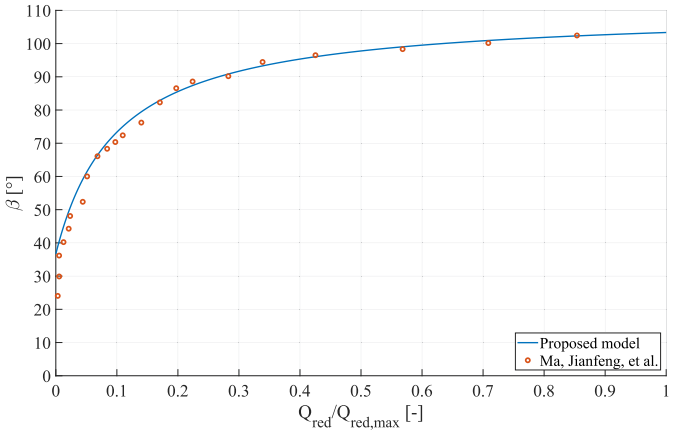


Fig. 3. Load arc respect to normalized output shaft torque [18].

Differently, The theoretical loading arcs could be determined by finding the intersection points between the circular profile of the flexspline and the elliptical profile of the wave generator. These arcs correspond to the meshing zones under conditions of minimum and maximum deformation, where the lengths of the semi-major and semi-minor axes of the ellipse depend on the external load.

The circular profile of the flexspline is described as:

$$x^2 + y^2 = r_{FS}^2; \tag{3}$$

The elliptical profile of the wave generator is given by:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1; \tag{4}$$

where the semi-major axis a and semi-minor axis b depend on the external load.

To calculate the intersection points, substitute $y = \pm\sqrt{r_{FS}^2 - x^2}$ into the elliptical equation, rearrange and simplify to obtain a quadratic equation in x^2:

$$x^2\left(\frac{1}{a^2} - \frac{1}{b^2}\right) + \frac{r_{FS}^2}{b^2} - \frac{1}{b^2} = 0. \tag{5}$$

Solving this equation gives the x-coordinates of the intersection points. The corresponding y-coordinates are found using:

$$y = \pm\sqrt{r_{FS}^2 - x^2}. \tag{6}$$

The angular positions of the intersection points relative to the flexspline center are:

$$\theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{7}$$

The load semi-arc is determined as the angular difference between these points:

$$\beta = 2 \cdot \left|\tan^{-1}\left(\frac{y_2}{x_2}\right) - \tan^{-1}\left(\frac{y_1}{x_1}\right)\right|. \tag{8}$$

Defining the loading arc allows to clearly delineate the two primary zones of interaction among the three components of the system: the interaction between the inner ring and the outer ring of the bearing, and consequently between the wave generator and the flexspline, as well as between the flexspline and the circular spline.

2.2. The wave generator: Elliptical cam and bearing inner ring

The first component to be modeled is the wave generator, the element representing, in the chosen configuration, the input of the motion of the harmonic reducer. For this model, the inner ring of the elliptical deformable bearing is considered fixed to the elliptical cam (Fig. 4). This assumption simplifies the analysis by providing a stable reference point from which the deformation of the bearing and the subsequent motion can be calculated. To accurately capture the dynamics of the wave generator, the following forces and torques are considered:

- Input Torque: this is the torque applied to the input shaft of the reducer. The input torque is a crucial parameter as it drives the entire system.

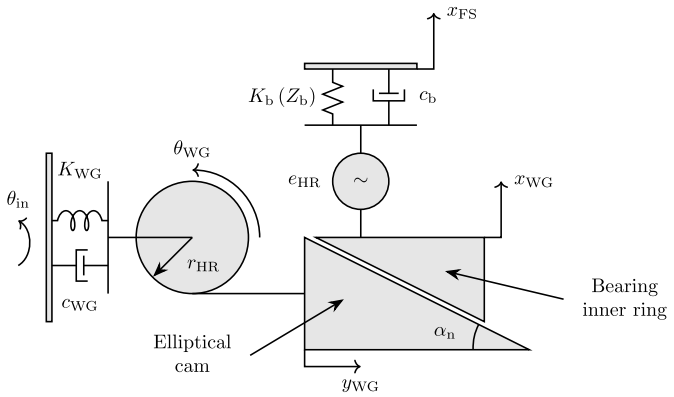


Fig. 4. Wave generator lumped parameters model.

- Force Exchanged with the Flexspline through the bearing interaction: the interaction force between the wave generator and the flexspline. This force results from the de-formation of the flexspline as it conforms to the elliptical shape of the wave generator elliptical bearing. Also the connected friction moment is considered.

The reducer input torque $(Q_{red,in})$, which is the torque applied to the wave generator, is computed using a spring–damper system. This system takes into account both the angular position of the input shaft and the angular position of the wave generator itself [21], as described in Eq. (9):

$$Q_{red,in} = K_{WG} \cdot (\theta_{in} - \theta_{WG}) + c_{WG} \cdot (\dot{\theta}_{in} - \dot{\theta}_{WG})\,; \tag{9}$$

where:

- K_{WG} is the input shaft stiffness;
- c_{WG} is the input shaft damping;
- θ_{in} is the input shaft angular position;
- θ_{WG} is the WG angular position.

The stiffness of the input shaft coupling is calculated considering the stiffness of an hollow shaft, and the damping is directly linked to it by means of the damping factor.

The governing equation of the wave generator rotation dynamic is expressed in Eq. (10):

$$Q_{red,in} - \left(F_b + \frac{F_{fb}}{\tan\alpha_n}\right) \cdot \tan\alpha_n \cdot r_{HR} = J_{WG} \cdot \ddot{\theta}_{WG}; \tag{10}$$

where:

- r_{HR} represents the equivalent radius, which can be considered as the average radial distance from the center of the wave generator to the midpoint between the contact point of the wave generator-flexspline and the contact point of the flexspline-circular spline. This equivalent radius simplifies the representation of the wave generator's deformation effect on the flexspline and provides a consistent geometric reference for modeling the dynamic interactions within the harmonic reducer.
- F_b and F_{fb} are the bearing interaction and friction forces;
- J_{WG} is the WG moment of inertia calculated as the hollow shaft moment of inertia.

The equivalent cam angle, which is linked to the transmission ratio between the wave generator motion and the flexspline single teeth motion, is calculated knowing the reducer ratio and the pressure angle as given by Eq. (11):

$$\tan\alpha_n \cdot \tan\alpha_t = \frac{1}{\tau}; \tag{11}$$

where α_t is the teeth pressure angle.

The radial (x_{WG}) and tangential (y_{WG}) displacements of the wave generator edge can be derived as functions of the WG angular position:

$$x_{WG} = \theta_{WG} \cdot R_{WG} \cdot \tan\alpha_n; \qquad y_{WG} = \theta_{WG} \cdot R_{WG}. \tag{12}$$

where R_{WG} represent the wave generator bearing outer radius.

2.3. Bearing interaction

In this section, the interaction between the wave generator and the flexspline through the deformable bearing will be described. As previously mentioned, an equivalent interaction model is used. The wave generator's motion induces a deformation in the flexspline. Although the flexspline itself does not rotate, its shape deformation follows the rotation of the wave generator, having therefore the same rotational speed. To define the deformation of the flexspline, it is necessary to consider the change in the radial coordinate of the wave generator, and thus in the flexspline. Additionally, there is a friction force that could potentially cause motion in the fixed flexspline.

2.3.1. Bearing radial force

The force transmitted by the bearing consists of two primary components: the elastic restoring force and the damping force (Fig. 4). Together, these forces describe the interaction between the inner and outer ring of the bearing and consequentially between the wave generator and the flexspline.

$$\begin{aligned} F_{\mathrm{b}} &= K_{\mathrm{b}} \cdot \Delta s_{\mathrm{b}} + c_{\mathrm{b}} \cdot \Delta v_{\mathrm{b}}; \\ \Delta s_{\mathrm{b}} &= \theta_{\mathrm{WG}} \cdot R_{\mathrm{HR}} \cdot \tan \alpha_{\mathrm{n}} - x_{\mathrm{FS}} + e_{\mathrm{HR}}; \\ \Delta v_{\mathrm{b}} &= \dot{\theta}_{\mathrm{WG}} \cdot R_{\mathrm{HR}} \cdot \tan \alpha_{\mathrm{n}} - \dot{x}_{\mathrm{FS}} + \dot{e}_{\mathrm{HR}}; \end{aligned} \tag{13}$$

where:

- K_{b} is the bearing stiffness;
- c_{b} is the damping coefficient;
- e_{HR} represents a geometric distortion of the ideal elliptical profile due to the wave generator misalignment.

Hertzian contact theory is employed to determine the bearing single load ball radial stiffness values $\left(Z_{\mathrm{bi}}\right)$, taking into account the elliptical contact area between the ball and the races. As reported in [22] it is not feasible to derive an analytical solution for the parameters defining the contact surface, a numerical solution, initially proposed by Hamrock and Anderson [23] and subsequently adapted by Guay and Frikha, is utilized. Then the total stiffness is calculated considering the total number of load balls $\left(Z_{\mathrm{b}}\right)$.

$$K_{\mathrm{b}} = Z_{\mathrm{b}} \cdot K_{\mathrm{bi}} = Z_{\mathrm{spheres}} \cdot \frac{\beta}{2\pi} \cdot K_{\mathrm{bi}} \tag{14}$$

where $\left(Z_{\mathrm{spheres}}\right)$ is the total number of bearing balls.

The defined model allows for the calculation of the radial displacement of the flexspline, accumulated over the number of half-rotations of the wave generator. However, to determine the actual displacement during rotation, it is necessary to subtract a value proportional to the total number of wave generator rotations.

2.3.2. Misalignment of the wave generator

The interaction through the bearing is significantly influenced by the potential for misalignment between the wave generator and the flexspline, which can arise from mounting errors. Such misalignments may lead to uneven wear, increased friction, and reduced efficiency in the system.

To describe this phenomenon a six-variable model is proposed, in which the parameters of misalignment in terms of displacement along the x and y directions, as well as rotation around these two axes, represent the inputs. By knowing these parameters, it is possible to calculate a roto-translation matrix that links the wave generator position in the presence of misalignment to its nominal angular position [13].

The misalignment is introduced through small angular displacements $\left(\underline{\varepsilon} = \left[\varepsilon_{\mathrm{x}}, \varepsilon_{\mathrm{y}}\right]\right)$ and translational offsets $\left(\underline{e} = \left[e_{\mathrm{x}}, e_{\mathrm{y}}\right]\right)$ applied to the wave generator (Fig. 5). The elliptical motion of the wave generator is discretized into angular steps, and the resulting misaligned configuration is computed.

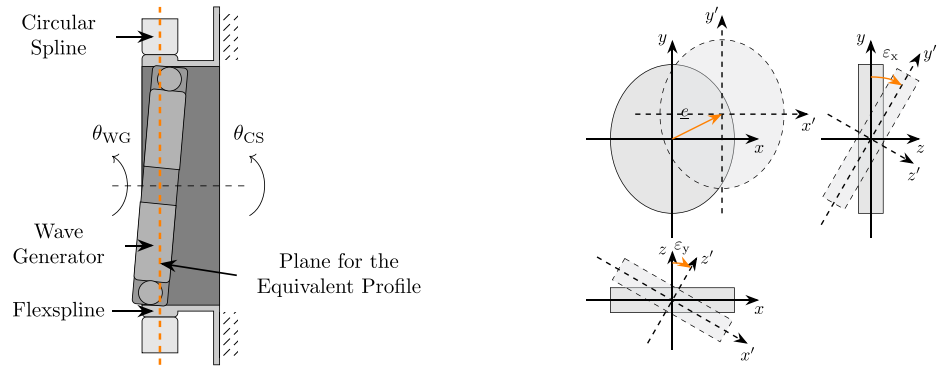
The first matrix rotates the points in the xz-plane around the x-axis by an angle ε_{x}, the second one rotates the points in the yz-plane around the y-axis by an angle ε_{y} and the translational misalignment is applied as a shift in the x and y directions.

$$\underline{\underline{R}}_{\mathrm{x}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varepsilon_{\mathrm{x}}) & -\sin(\varepsilon_{\mathrm{x}}) \\ 0 & \sin(\varepsilon_{\mathrm{x}}) & \cos(\varepsilon_{\mathrm{x}}) \end{bmatrix} \tag{15}$$

$$\underline{\underline{R}}_{\mathrm{y}} = \begin{bmatrix} \cos(\varepsilon_{\mathrm{y}}) & 0 & \sin(\varepsilon_{\mathrm{y}}) \\ 0 & 1 & 0 \\ -\sin(\varepsilon_{\mathrm{y}}) & 0 & \cos(\varepsilon_{\mathrm{y}}) \end{bmatrix} \tag{16}$$

$$\underline{T}_{\mathrm{xy}} = \begin{bmatrix} e_x \\ e_y \\ 0 \end{bmatrix} \tag{17}$$

The final misalignment error is calculated by comparing the nominal and misaligned positions of the wave generator, allowing for an accurate prediction of the kinematic error across multiple angular rotations.



(a) Misaligned wave generator and plane for equivalent profile.

(b) Misalignment model parameters.

Fig. 5. Wave generator misalignment model.

To determine the kinematic error, the 3D profile of the wave generator is obtained and intersected with a plane perpendicular to the input axis. This intersection results in an equivalent planar profile, which is then compared to the nominal profile. By comparing the two profiles, the radial displacement is determined, and the associated kinematic error is calculated as a direct result [24].

The geometric error parameter of Eq. (13) can be therefore linked to the misalignment errors through the following equation:

$$e_{HR} = \left| \left(\underline{\underline{R}}_{y} \cdot \underline{\underline{R}}_{x} \cdot \underline{x} + \underline{T}_{xy} \right) - \underline{x} \right| \tag{18}$$

where $\underline{x} = (x, y)$ represents the nominal coordinates of the wave generator profile in the plane for the equivalent profile.

2.3.3. Bearing friction

If the bearing is considered without friction, there would be no interaction between the outer ring and the inner ring in the tangential direction. However, accounting for this interaction, a torque, and then a force (F_{fb}) is generated on each ring, which tends to slow down the wave generator and move the flexspline circumferentially, motion which is represented in the model by the coordinate y_{FS}. This interaction is dependent not on the relative position of the two rings but on their relative velocity. Specifically, this torque (Q_{fb}) is calculated using the Harris model (M_{Harris}), modified for implementation in the harmonic reducer model [25,26]:

$$Q_{fb} = F_{fb} \cdot r_{HR} = M_{Harris} = M_{lub} + M_{load}; \tag{19}$$

where:

- M_{lub} is lubrication-dependent friction torque;
- M_{load} is the load-dependent friction torque.

The two torques are defined as follows:

$$M_{lub} = f_{b0} \cdot \nu_0 \cdot n; \qquad M_{load} = f_{b1} \cdot F_b \cdot r_{HR}; \tag{20}$$

where:

- f_{b0} is the lubrication friction coefficient, that in first approximation is considered constant and in a condition of perfect lubrication;
- f_{b1} is the load friction coefficient;
- ν_0 is the kinematic viscosity of the lubricant expressed in centiStokes;
- n is the rotational speed expressed in rpm;
- F_b is the bearing radial force defined in Eq. (13).

Adjusting the friction coefficient in the frictional moment associated with lubrication will enable an accurate replication of the increase in internal friction within bearings due to degraded lubricant properties, insufficient lubrication and changing in temperature. This increase leads to a reduction in overall system efficiency. By making this adjustment, it is possible to precisely simulate the effects of suboptimal lubrication on bearing performance.

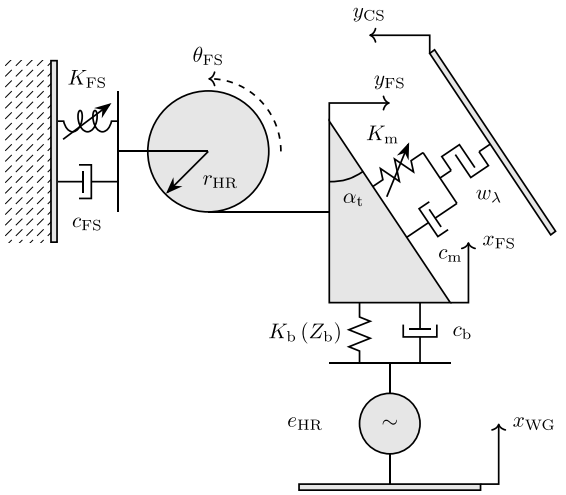


Fig. 6. Flexspline lumped parameters model.

2.4. The flexspline: Bearing outer ring and flexible external gear

The flexspline is a crucial component in the harmonic drive system and must be accurately modeled, being the major contributors to the hysteretic behavior of the reducer [27]. In this model, the outer ring of the elliptical deformable bearing is assumed to be fixed to the flexible gear. This assumption simplifies the analysis by treating the bearing outer ring and the flexspline as an unique compliant structure that deforms under load.

To calculate the dynamics of the flexspline, the forces exchanged with the wave generator, the circular spline, and the fixed frame are considered (Fig. 6). These interactions are critical for understanding the flexspline's deformation and its impact on the overall system performance. The wave generator imposes an elliptical deformation on the flexspline, generating radial and tangential forces. These forces are crucial for the proper meshing of the flexspline with the circular spline. The circular spline interacts with the flexspline through meshing teeth. This interaction generates contact forces that depend on the load transmitted through the gear system. The fixed frame provides a reference point for the flexspline's deformation and opposes to its deformation and rotation. The interaction between the flexspline and the fixed frame involves reaction forces that balance the applied loads and constrain the rotational motion of the flexspline.

By accurately modeling these interactions and the resulting forces, it is possible to predict the flexspline's dynamic response under various operational conditions.

2.4.1. Flexspline torsional stiffness

The flexspline material is modeled as steel, with temperature-dependent properties derived from manufacturer data and literature sources [28,29]. The variation of Young's modulus (E) and Poisson's ratio (ν) with temperature is illustrated in Fig. 7. These properties are represented using Eq. (22), which incorporates experimental data to capture the nonlinear reduction in stiffness as the temperature increases.

The temperature evolution of the harmonic reducer is estimated using a thermal network approach detailed in Section 2.7. Internal heat generation arises from bearing friction $\left(F_{f,b}\right)$ and meshing interactions $\left(F_{f,m}\right)$. Heat dissipation is characterized by the thermal resistance $\left(R_{th,red}\right)$ and thermal capacity $\left(C_{th,red}\right)$, enabling the determination of a bulk temperature for the reducer. This bulk temperature directly influences the flexspline stiffness. By dynamically linking material properties to temperature, the model provides a realistic representation of the reducer's behavior under varying thermal conditions.

The nominal rotational stiffness of the flexspline can be modeled by considering it as a thin-walled hollow shaft, and dividing it by the square of the equivalent radius it is possible to obtain the equivalent linear nominal stiffness $\left(K_{FS}\right)$, and the relative damping $\left(c_{FS}\right)$.

The stiffness of a hollow shaft is directly proportional to the Young modulus and inversely proportional to Poisson ratio. This implies that by varying these parameters, the stiffness of the flexspline cup will change accordingly. Therefore, the nominal stiffness at each temperature is calculated as follows:

$$K_{FS} = K_{FS,nom} \cdot \frac{\gamma_{FS}}{\gamma_{FS,nom}}; \tag{21}$$

where γ_{FS} and $\gamma_{FS,nom}$ are the ratio $\frac{E_{steel}}{\nu_{steel}+1}$ at actual and reference temperature.

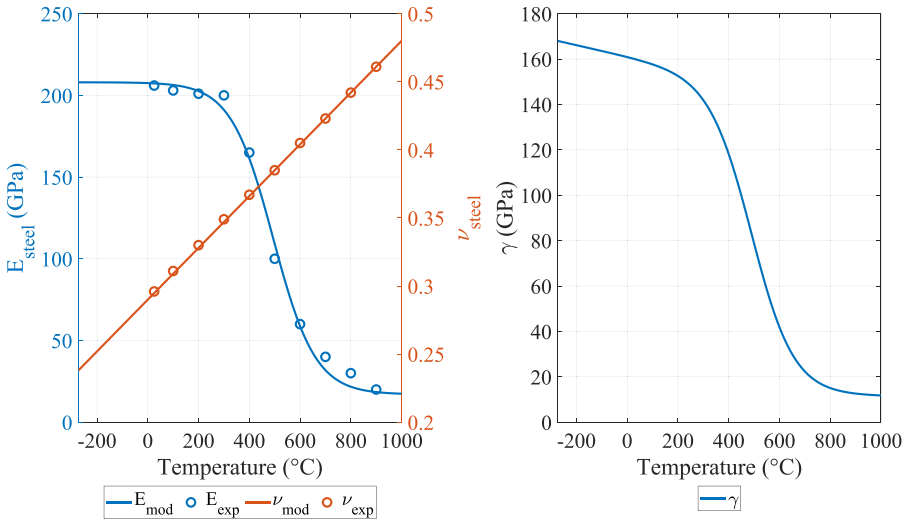


Fig. 7. Young's modulus and Poisson's ratio of steel and γ_{FS} respect to temperature [29].

It is possible to define the laws which link the two material parameters to temperature:

$$E_{steel} = E_{steel,0} + a_{E,0} \cdot \tanh \frac{b_{E,0} - T}{c_{E,0}}; \qquad \nu_{steel} = \nu_{steel,0} + a_{\nu,0} \cdot T \tag{22}$$

obtaining in this way the graphs in Fig. 7.

As aforementioned the flexspline modeling is fundamental to describe the hysteresis behavior of the reducer, in fact it is assumed that the hysteresis behavior of the gear is primarily linked to the flexspline cup, which experiences the most deformation during the reducer's motion.

To evaluate the torsional stiffness of the flexspline, the Bouc–Wen hysteresis model is introduced [30,31]. This model considers a restoring force derived from the elastic energy of the flexspline, given by:

$$F_{FS} = K_{FS} \cdot \left(\alpha_{BW} \cdot y_{FS} + \left(1 - \alpha_{BW}\right) \cdot z_{FS}\right) + c_{FS} \cdot \dot{y}_{FS};$$
$$\dot{z}_{FS} = \left(A_{BW} - \left(\beta_{BW} \cdot \mathrm{sgn}\left(z_{FS} \cdot \dot{y}_{FS}\right) + \gamma_{BW}\right) \cdot \left|\frac{z_{FS}}{u_{BW}}\right|^{n_{BW}}\right) \cdot \dot{y}_{FS}. \tag{23}$$

where:

- z_{FS} is the hysteresis model hidden state;
- α_{BW} is the stiffness repartition factor;
- A_{BW} is the initial hysteresis factor;
- n_{BW} is the transition factor;
- β_{BW} and γ_{BW} are the width and shape factors;
- u_{BW} is the yielding displacement.

By varying the parameters of the Bouc–Wen model, the shape and amplitude of the hysteresis cycle can be adjusted to fit experimental data [32,33]. This allows for precise modeling of the flexspline's dynamic behavior.

The graph shown in Fig. 8 illustrates the hysteretic behavior of the flexspline alone. Manufacturer data was utilized to ensure that the general hysteretic behavior of the entire reducer model is comparable, as shown in the results section.

A comparable outcome can be achieved through a polynomial estimation of the variable stiffness of the meshing, but by using the Bouc–Wen model, it will be possible to describe the degradation of hysteresis behavior over time, which is crucial for long-term reliability and performance analysis of the gear system.

2.4.2. Flexspline tangential dynamics

Analyzing the tangential translation equilibrium (and consequently the rotation) of the flexspline, several forces must be considered to obtain the tangential displacement (y_{FS}). These include the internal friction force within the bearing, the meshing force with the circular spline, the restoring elastic force opposing the tangential deformation of the flexspline, and the inertia torque. The governing equation for this equilibrium is:

$$F_m \cdot \cos\alpha_t - F_{FS} + F_{fb} + F_{fm} \cdot \sin\alpha_t = m_{FS} \cdot \ddot{y}_{FS}, \tag{24}$$

where F_{fb} is the bearing friction force.

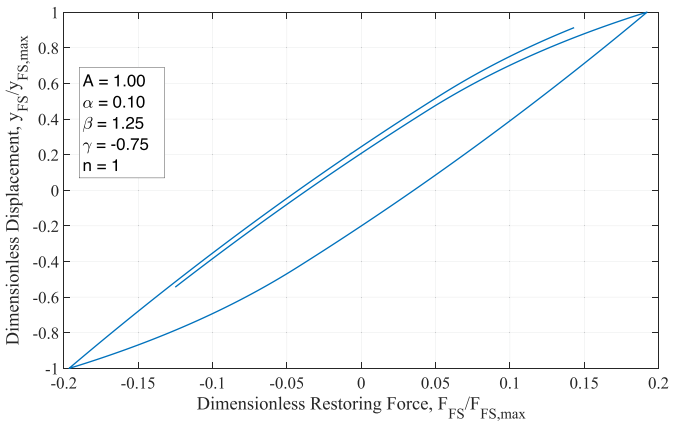


Fig. 8. Hysteresis behavior of the flexspline.

As aforementioned, the restoring elastic force is intrinsically linked to the material and geometric properties of the flexspline. This force is crucial for maintaining the structural integrity and proper functioning of the flexspline, as it ensures that the component can effectively oppose tangential deformations and return to its initial configuration when external forces are removed.

2.4.3. Flexspline radial dynamics

The radial dynamics equilibrium of the flexspline must consider several key forces and factors to compute the radial displacement of the flexspline (x_{FS}). These include the radial force transmitted through the bearing and caused by the motion of the wave generator, the interaction forces exchanged during the meshing with the circular spline, and the inertia of the flexspline itself.

$$F_{\mathrm{b}} - F_{\mathrm{m}} \cdot \sin\alpha_{\mathrm{t}} - F_{\mathrm{fm}} \cdot \cos\alpha_{\mathrm{t}} = m_{\mathrm{FS}} \cdot \ddot{x}_{\mathrm{FS}}; \tag{25}$$

where:

- F_{b} is the bearing radial force;
- F_{m} and F_{fm} are the meshing force and friction;
- $m_{\mathrm{FS}} = \frac{J_{\mathrm{FS}}}{r_{\mathrm{HR}}^2}$ is the flexspline equivalent mass linked to rotational inertia.

2.5. Meshing interaction

In this section, the interaction between the teeth of the flexspline and the circular spline is discussed. The teeth are approximated as trapezoidal in shape, and all the various meshing teeth along the load arc are treated as a unique equivalent contact of a single tooth pair with equivalent properties such as stiffness, damping, and backlash, representative of the contribution of all the considered meshing couples. This trapezoidal approximation simplifies the analysis and modeling of the mechanical interactions. Trapezoidal teeth provide a practical approximation that captures the essential geometric and mechanical characteristics necessary for calculating forces and deformations during meshing.

2.5.1. Meshing force

The total stiffness of the meshing process in harmonic reducers is crucial as it directly impacts the static and dynamic response and load-bearing capacity of the gear system. The total meshing force is calculated by means of a spring–damper equivalent model.

$$\begin{aligned} F_{\mathrm{m}} &= K_{\mathrm{m}} \cdot \Delta s_{\mathrm{m}} + c_{\mathrm{m}} \cdot \Delta v_{\mathrm{m}}; \\ \Delta s_{\mathrm{m}} &= x_{\mathrm{FS}} \cdot \sin\alpha_{\mathrm{t}} + y_{\mathrm{FS}} \cdot \cos\alpha_{\mathrm{t}} - \theta_{\mathrm{CS}} \cdot R_{\mathrm{HR}} \cdot \cos\alpha_{\mathrm{t}}; \\ \Delta v_{\mathrm{m}} &= \dot{x}_{\mathrm{FS}} \cdot \sin\alpha_{\mathrm{t}} + \dot{y}_{\mathrm{FS}} \cdot \cos\alpha_{\mathrm{t}} - \dot{\theta}_{\mathrm{CS}} \cdot R_{\mathrm{HR}} \cdot \cos\alpha_{\mathrm{t}}; \end{aligned} \tag{26}$$

where:

- K_{m} is the load arc equivalent meshing stiffness;
- c_{m} is the load arc equivalent damping coefficient.

As previously mentioned, the number of tooth pairs engaged in the meshing process is influenced by the external load, which dictates the effective load distribution across the gear teeth. In the following a load arc condition around the optimum value is considered: in this conditions the total meshing stiffness could be considered proportional to the load arc.

After the definition of the load arc (β) it is possible to obtain the total number of meshing teeth $\left(Z_{\mathrm{m}}\right)$ as:

$$Z_{\mathrm{m}} = Z_{\mathrm{FS}} \cdot \frac{\beta}{2\pi}; \tag{27}$$

Next, the engaging factor within the load arc can be defined as the ratio between the meshing stiffness of each teeth pair and the nominal, or maximum, stiffness of a fully engaged tooth pair [34]. In this paper, a specific skew asymmetric piece-wised Gaussian-like probability density function is proposed and utilized to characterize the engaging factor.

$$\mathrm{EF}(\delta) = \left(\mathrm{erf}\left(\lambda_{\mathrm{EF}}\right) + 1\right) \cdot \left(1 - c_{\mathrm{EF}}\right)^{n_{\mathrm{EF}}} \cdot \begin{cases} e^{a_{\mathrm{EF}} \cdot \left(\delta - \delta_0\right)^2} & \delta < \delta_0 \\ e^{b_{\mathrm{EF}} \cdot \left(\delta - \delta_0\right)^2} & \delta \geq \delta_0 \end{cases}; \tag{28}$$

where:

- δ is the dimensionless coordinate along the load arc;
- δ_0 is the maximum stiffness coordinate depending of the load. It can be considered as a limit or reference point for the stiffness of the system;
- λ_{EF} is the skewness factor;
- a_{EF} and b_{EF} are the engaging factor model parameters;
- c_{EF} is the module parameter;
- n_{EF} is the exponential parameter $\left(\text{equal to } \frac{1}{5}\right)$.

The equation is made up by three terms: and exponential one, a module one, and an error function one. The exponential term introduces a smooth transition, where the difference between δ and δ_0 drives the rate of change of the engaging factor. As δ approaches δ_0, the term tends to decrease the overall engaging factor. The absolute value term ensures that the distribution reaches zero at the boundaries of the considered interval. The error function components control the non-linearity of the stiffness with respect to the load position. It ensures a smooth, sigmoidal-like transition between different stiffness regions, where λ modulates how sharp or gradual this transition is. A higher λ makes the transition sharper.

The function provided has similarities to a skewed normal distribution [35], but it is not exactly a skewed normal distribution in the strictest sense. In a skewed normal distribution, the asymmetry comes from a term that skews a normal distribution in one direction. In the same way in the proposed equation, the asymmetry is introduced by the error function, which modulates the magnitude of the engaging factor.

By analyzing data from the literature [17–19], it becomes possible to tailor these parameters to the specific application under consideration, in this case, the harmonic reducer. This fitting process allows for the adjustment of model inputs to reflect the unique characteristics and operational demands of the harmonic drive, ensuring that the simulation aligns closely with real-world behavior. Such an approach not only improves the model's accuracy but also enhances its predictive capabilities when applied to practical engineering scenarios and fault description.

All the meshing teeth in the loaded arc act in parallel, hence the total equivalent stiffness is given by the sum of the single contributions. However, each tooth does not contribute equally to the total stiffness, but follows the engaging factor function, depending on its relative position within the load zone. The equivalent stiffness is therefore the value that, if assigned to every tooth equally, would lead to a total stiffness equal to the actual one obtained considering the engaging factor.

From the defined engaging factor it is possible to obtain a mean engage factor, exploiting the mean value theorem for integrals, in order to calculate the mean stiffness in the load arc. To do this the engage factor constant $\left(k_{\mathrm{EF}}\right)$ is defined:

$$k_{\mathrm{EF}} = \frac{1}{\delta_{\max} - \delta_{\min}} \cdot \int_{\delta_{\min}}^{\delta_{\max}} \mathrm{EF}\, \mathrm{d}\delta = \frac{1}{2} \int_{-1}^{1} \mathrm{EF}\, \mathrm{d}\delta; \tag{29}$$

By considering the experimental evolution of the total stiffness of one meshing area with respect to the number of meshing teeth, it is possible to derive a mathematical relationship between the number of meshing pairs, the load arc and the total stiffness of the meshing area.

$$K_{\mathrm{m}} = Z_{\mathrm{m}} \cdot k_{\mathrm{EF}} \cdot K_{\mathrm{mi}} = \frac{\beta}{2\pi} \cdot Z_{\mathrm{FS}} \cdot k_{\mathrm{EF}} \cdot K_{\mathrm{mi}}; \tag{30}$$

where:

- K_{mi} is the stiffness of a fully engaged single meshing tooth pair.

The stiffness of a fully engaged tooth pair can be computed based on the individual stiffness contributions from both the Hertzian contact and the bending and shear effects in the gear teeth [19]. The stiffness of a single tooth is modeled as a cantilever beam, excluding the flexspline rim's flexibility due to its direct contact with the wave generator. Partial tooth engagement in the load arc is accounted for using an engagement factor that adjusts the overall meshing stiffness.

In Fig. 9, a comparison is presented between the proposed engaging factor model and the dimensionless stiffness of each tooth within the loading arc, based on two different sets of literature data. This comparison highlights the variations in stiffness behavior under load and provides insights into the performance characteristics of the engaging factor model relative to established research.

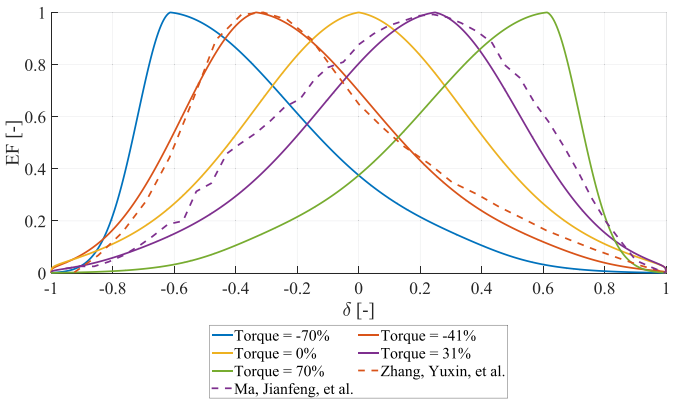


Fig. 9. Variation of the engaging factor shape with external load, compared to literature data [18,19].

Ensuring a sufficient number of tooth pairs are engaged, even under minimal load conditions, is vital for maintaining structural integrity and operational efficiency. The shape of the engaging factor is influenced by the external load. As shown in Eq. (2), the amplitude of the load arc is also influenced by the external load. Although δ is dimensionless the increase in external load leads to an increase in the total number of tooth pairs in contact.

$$\delta_0 = \tan\left(\frac{Q_{\text{red}}}{Q_{\max}} \cdot \frac{\pi}{4}\right);$$

$$a_{\text{EF}} = -\frac{1}{\left(\frac{Q_{\text{red}}}{Q_{\max}}+1\right)^3} - 2.4; \qquad b_{\text{EF}} = -\frac{1}{\left(-\frac{Q_{\text{red}}}{Q_{\max}}+1\right)^3} - 2.4; \tag{31}$$

$$c_{\text{EF}} = \frac{|\delta-\delta_0|}{1-\text{sign}(\delta-\delta_0)\cdot\delta_0} \qquad \lambda_{\text{EF}} = |\delta-\delta_0|^5;$$

2.5.2. Meshing friction

Meshing friction is a crucial factor in the dynamic response of gear systems. The frictional force in the meshing area can be modeled using a simple Amonton-Coulomb friction model, which takes into account the normal load and the coefficient of friction:

$$F_{\text{fm}} = f_{\text{m1}} \cdot F_{\text{m}}; \tag{32}$$

where:

- f_{m1} is the coefficient of meshing friction [9].

This model assumes a constant coefficient of friction and does not distinguish between static and dynamic friction, which could be justified by the specific operational conditions of harmonic drives. Given that these systems generally operate under continuous motion and often avoid the stick–slip transitions associated with static friction, the simplification helps reduce computational complexity. Including both static and dynamic friction coefficients would require more detailed modeling of stick–slip behavior, which is often negligible in continuously operating systems like harmonic drives.

The model also does not account for the relative speed dependency typically associated with dynamic friction. In many applications, the relative speed between mating surfaces can influence frictional forces. However, in harmonic drives, the relative motion is often small and consistent, leading to the choice of a constant coefficient of friction. Adding a speed-dependent term could be more accurate but would increase the model's complexity, which may not provide significant improvements for most practical purposes.

Lubrication effects are similarly not considered in this simplified model, despite their potential to reduce friction. The decision to neglect lubrication stems from a focus on worst-case scenarios or a desire to isolate mechanical behavior. In reality, lubrication plays a critical role in reducing wear and altering the frictional behavior, particularly in high-performance harmonic drives.

If the external load is null, the frictional force in the meshing area would also be null, as the normal force between the teeth would decrease significantly. This is because the normal load is the primary contributor to the frictional force in this model, and a lack of external loading would reduce the engagement forces between the teeth.

At the zero-crossing point (when the direction of meshing changes), the frictional force could momentarily drop or reverse its direction. The Amonton-Coulomb model does not inherently capture the behavior at zero-crossing, where other effects such as adhesion or transient changes in contact forces might come into play. Including such effects would require a more advanced friction model, possibly incorporating hysteresis or a micro-slip regime.

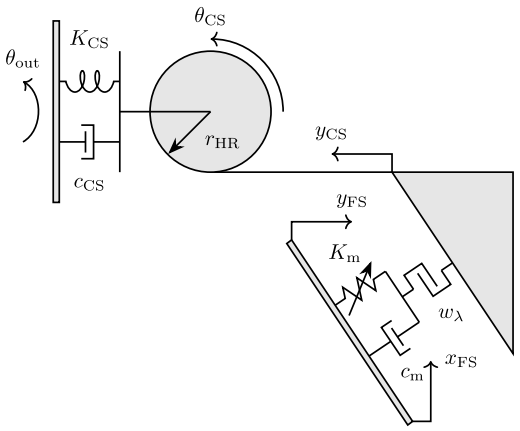


Fig. 10. Circular spline lumped model.

2.6. The circular spline: Rigid internal gear

The circular spline is the rigid internal gear that, in the considered configuration, serves as the output element connected to the output shaft of the reducer. The forces and torques considered include the meshing interaction with the flexspline, meshing friction, and the reducer's output torque (Fig. 10).

The reducer output torque $(Q_{red,out})$, which is the torque applied to the circular spline, is computed using a spring–damper system. This system takes into account both the angular position of the output shaft and the circular spline itself, as described in [21]:

$$Q_{red,out} = K_{CS} \cdot (\theta_{CS} - \theta_{out}) + c_{WG} \cdot (\dot{\theta}_{CS} - \dot{\theta}_{out}); \tag{33}$$

where:

- K_{CS} is the output shaft stiffness;
- c_{CS} is the output shaft damping;
- θ_{CS} is the CS angular position;
- θ_{out} is the output shaft angular position.

The stiffness of the output shaft coupling is calculated considering the stiffness of an hollow shaft, and the damping is directly linked to it.

The governing equation of the circular spline rotation dynamic is expressed in Eq. (34):

$$F_m \cdot r_{HR} \cdot \cos\alpha_t - F_{fm} \cdot r_{HR} \cdot \sin\alpha_t - Q_{red,out} = J_{CS} \cdot \ddot{\theta}_{CS} \tag{34}$$

where:

- J_{CS} is the CS moment of inertia.

2.7. Reducer thermal dynamics

A thermal dynamic model is introduced for the reducer to evaluate the evolution of its temperature and, consequently, how material properties change with it, particularly focusing on the stiffness of the flexspline, which is the most flexible element in the overall system. According to [36], various heat sources within the reducer, such as bearing friction and meshing friction, are identified (Fig. 11).

The thermal power generated by internal friction within the reducer can be approximately evaluated as:

$$P_{th,red} = F_b \cdot \Delta v_b + F_m \cdot \Delta v_m, \tag{35}$$

where:

- Δv_b is the relative speed in the bearing interaction;
- Δv_m is the relative speed in the meshing interaction.

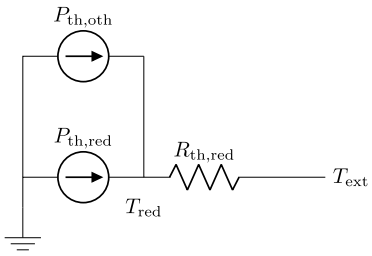


Fig. 11. Equivalent reducer thermal network.

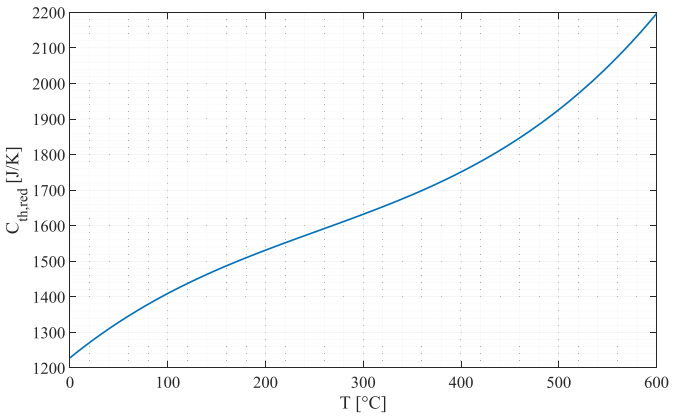


Fig. 12. Thermal capacity of the reducer respect to temperature [29].

This thermal power, combined with the thermal power exchanged with the other subsystems of the robotic joint $\left(P_{\text{th,oth}}\right)$, e.g. the electric motor, and the heat exchanged with the external environment through the case, is used to evaluate the mean temperature of the overall reducer; considered as a bulk mass:

$$P_{\text{th,red}} + P_{\text{th,oth}} - \frac{T_{\text{ext}} - T_{\text{red}}}{R_{\text{th,red}}} = C_{\text{th,red}}(T) \cdot \frac{dT_{\text{red}}}{dt}, \tag{36}$$

where:

- T_{ext} is the external temperature;
- T_{red} is the temperature of the reducer;
- $R_{\text{th,red}}$ is the thermal resistance of the reducer;
- $C_{\text{th,red}}(T)$ is the thermal capacity of the reducer [29], which varies with temperature, as shown in Fig. 12. It is obtained as the product between the steel specific heat [37,38] and the overall mass of the reducer.

3. Defects model of the harmonic reducer

The primary aim of this section is to explore the dynamic effects resulting from the presence of faults, the ones commonly encountered during the lifespan of most harmonic drives. The goal is to develop a robust model that can dynamically simulate the system's behavior in the presence of these faults, which can lead to significant alterations in system parameters and overall performance. By understanding these dynamics, it is possible to better predict and mitigate the impact of faults, ultimately enhancing the reliability and efficiency of harmonic drives.

3.1. Meshing teeth wear

The abrasive wear between spur gears, as described in [39], could be modeled using the Archard law [40]. This approach is adapted for wear in gear teeth with a double-arc profile in [41].

Wear, in this context, leads to two significant effects [42]: loss of torque transmission during changes in motion direction, i.e. backlash, and alteration of the teeth's pressure angle (Fig. 13). These two phenomena can be modeled independently, but both

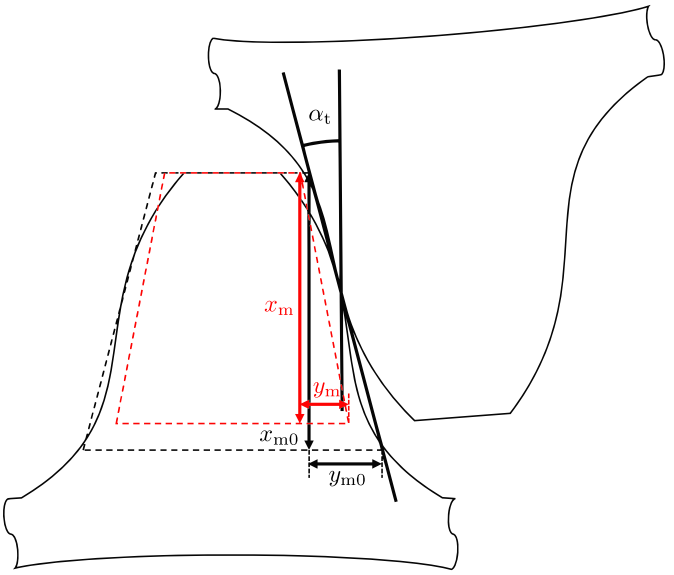


Fig. 13. Trapezoidal profile approximation and effect of wear on the pressure angle.

linked to the teeth thickness reduction (w_λ). The torque transmitted during meshing derived from a force acting between the teeth of the two gears. This can be formulated through the following equation for backlash:

$$F_m = \begin{cases} K_m \cdot \left(\Delta s_m - \frac{w_\lambda}{2}\right) + c_m \cdot \Delta v_m & \text{if } \Delta s_m \geq \frac{w_\lambda}{2} \\ K_m \cdot \left(\Delta s_m + \frac{w_\lambda}{2}\right) + c_m \cdot \Delta v_m & \text{if } \Delta s_m \leq -\frac{w_\lambda}{2} \\ 0 & \text{otherwise} \end{cases}; \tag{37}$$

The effect of wear on the pressure angle is modeled using the trapezoidal approximation of the teeth profile, as described in [8,43]. This approximation provides an analytical expression to describe the change in pressure angle:

$$\tan \alpha_t = \frac{y_m}{x_m} = \frac{y_{m0} - w_\lambda \cdot \cos \alpha_t}{x_{m0} - w_\lambda \cdot \sin \alpha_t}; \tag{38}$$

where:

- y_m and x_m are respectively the equivalent teeth width and height.
- y_{m0} and x_{m0} are respectively the initial equivalent teeth width and height.

The primary consequence of the variation in pressure angle is a proportional discrepancy between the actual output torque and the expected torque. This deviation results from a change in the transmission ratio, which is directly influenced by the wear-induced alteration of the teeth's geometry.

3.2. Flexspline tooth crack

In the event of a crack insurgence, the stiffness of a specific tooth of the flexspline interested by this fault is reduced. This localized stiffness reduction is associated with a specific angle of the flexspline, identifying the position of the absolute cracked tooth along the fixed flexspline. If the crack growth is such to cause a complete removal of the faulty teeth, the detached tooth would increase vibrations in the reducer by causing force oscillation within the wave generator or the flexspline.

Assuming that a tooth of the flexspline has a percentage stiffness reduction proportional to the crack depth [44], as expressed by the stiffness modification coefficient, the calculation of the gear's overall stiffness expressed in (30) modifies in:

$$\begin{aligned} K_m &= k_{EF} \cdot \left(\frac{\beta}{2\pi} \cdot Z_{FS} - w_{TC} \cdot EF_{tot}\right) \cdot K_{mi}; \\ EF_{tot} &= EF|_{\delta_{t,crack}} + EF|_{\delta_{t,crack} + \frac{4\pi}{\beta}}; \\ \delta_{t,crack} &= \frac{\theta_{WG} - \theta_{t,crack}}{\frac{\beta}{4}}; \end{aligned} \tag{39}$$

where:

- $\theta_{t,crack}$ is the angular position of the cracked tooth.
- w_{TC} is the tooth crack stiffness reduction factor.

The stiffness reduction as a function of crack width reflects the progressive degradation of the tooth's load-bearing capacity. Initially, minor cracks have a limited impact on the overall stiffness. However, as the crack widens, the tooth's ability to resist bending and deformation is significantly reduced, and once the tooth breaks, it no longer contributes to the harmonic drive's function, effectively having zero stiffness.

By representing this degradation as a quadratic function, the model captures the accelerating rate of stiffness loss as the crack propagates, which aligns with theoretical fracture mechanics [44].

$$w_{TC} = \left(\frac{w_{crack,tooth}}{w_{tooth}} \right)^2; \tag{40}$$

where:

- $w_{crack,tooth}$ is the crack width;
- w_{tooth} is the tooth width.

In real-world applications, the presence of a cracked tooth can significantly affect the performance and reliability of the harmonic reducer. The increased vibrations and potential jamming can lead to further mechanical issues, necessitating robust monitoring and maintenance strategies. Understanding and accurately modeling the impact of such faults is crucial for the development of effective PHM systems.

3.3. Flexspline rim crack

When a crack occurs in the flexspline rim in radial direction, its torsional stiffness decreases and can ultimately lead to complete failure under extreme conditions. The reduction in flexspline stiffness due to a crack can be quantified using a stiffness reduction factor, expressed as:

$$K_{FS} = w_{RC} \cdot K_{FS,nom}; \tag{41}$$

where:

- w_{RC} is the rim crack stiffness reduction factor.

The stiffness reduction is proportional to the crack size [44] and becomes zero when the component fails:

$$w_{RC} = \left(1 - \frac{w_{crack,rim}}{w_{rim}} \right)^2; \tag{42}$$

where:

- $w_{crack,rim}$ is the crack width;
- w_{rim} is the rim width.

3.4. Defects model parameters

As aforementioned the proposed model allows for the simulation of various fault types or defects by modifying specific parameters that represent the physical effects of these faults. The key parameters associated with each fault type are as follows:

- Misalignment of the Wave Generator:
  - Parameter: Small angular displacements $(\underline{\varepsilon} = [\varepsilon_x, \varepsilon_y])$ and translational offsets $(\underline{e} = [e_x, e_y])$ applied to the wave generator.
  - Description: These parameters introduce an eccentricity in the wave generator's position, simulating misalignment effects.
- Wear of Meshing Teeth:
  - Parameter: Wear level (w_λ).
  - Description: An increase in the wear level is used to simulate the loss of material on the teeth, altering the meshing dynamics and increasing backlash.
- Crack on Flexspline Teeth:
  - Parameter: Localized stiffness reduction in specific teeth (w_{TC}).

Table 1
Parameters of the selected harmonic drive reducer.

Variable	Symbol	Value	UoM
FS teeth number	Z_{FS}	200	[–]
CS spline teeth number	Z_{CS}	202	[–]
Equivalent radius	r_{HR}	55.5	[mm]
WG moment of inertia	J_{WG}	807.41	[kg mm^2]
FS moment of inertia	J_{FS}	723.47	[kg mm^2]
CS moment of inertia	J_{CS}	107.88	[kg mm^2]
Harmonic reducer mass	m_{HR}	2.89	[kg]
Limit torques	$[Q_1, Q_2, Q_3]$	[76,275,982]	[Nm]
Torsional stiffness	$[K_1, K_2, K_3]$	[18,29,33]$\times 10^4$	[Nm/rad]

  - Description: The stiffness of individual teeth is reduced based on the crack depth and location, affecting the load distribution and increasing vibrational responses.

- Crack on the Flexspline Rim:
  - Parameter: Rim stiffness (K_{FS}).
  - Description: A reduction in K_{FS} represents the structural weakening caused by a crack, influencing the overall compliance of the flexspline, it is linked to the crack propagation factor (w_{RC}).

By systematically varying these parameters, the model can replicate the dynamic responses of harmonic drives under various fault conditions, providing valuable insights for diagnostics and health management.

4. Results

It is possible to summarize the presented equations in a system of non-linear differential equations in which the unknown variable are the kinematic quantities of each body considered.

The set of equations could be split in two, the first one regards the dynamic equilibrium equations and the second one the interaction calculation. Starting from that it is possible to define the initial value of each internal kinematic quantity.

The parameters of the harmonic drive model, listed in Table 1, have been assumed coherent with those adopted in [45] in order to allow the comparison of the results.

The simulation results presented in this section have been performed replicating a possible real configuration of a harmonic reducer test bench, in which the input shaft is controlled in position imposing the motion of the wave generator, while the output shaft is subjected to the external load and the flexspline is connected to the ground.

Fig. 14 illustrates the evolution of key dynamic variables during the simulation of the harmonic reducer under nominal conditions, with the input speed varying from 0 to 2400 rpm in both directions. The top graphs present the angular positions and velocities of the input shaft $(\theta_{in}$ and $\omega_{in})$, wave generator, flexspline, and circular spline. These plots emphasize the nearly identical behavior of the input shaft and wave generator, as well as the circular spline and output shaft, confirming the transmission ratio, in the considered case equal to 101 according to Eq. (1) and the flexspline and circular spline number of teeth, listed in Table 1.

The middle graphs show the internal forces and displacements within the reducer. Specifically, the bearing radial force (F_b) and meshing force (F_m) are displayed alongside the radial (x_{FS}) and angular (θ_{FS}) displacements of the flexspline. These plots highlight the interaction dynamics between the flexspline and the other components, illustrating how the forces evolve as the input speed changes.

Upon analyzing the internal dynamics of the reducer, it is evident that the flexspline remains nearly stationary rotation, as it is rigidly fixed to the frame. In contrast, the coordinate representing virtual deformation closely follows the position trend of the wave generator.

The bottom plot provides an overview of the torque transmitted by the reducer, showing both the input (Q_{in}) and output torque (Q_{out}). The spikes observed during speed reversals are attributed to the dynamic transitions required to accelerate internal masses and overcome frictional forces. Under steady-state conditions, the transmitted torque stabilizes, reflecting the system's efficiency in converting input power into output torque. This plot underscores the model's ability to capture both transient dynamics and steady-state behavior, providing valuable insights into the reducer's performance.

As the wave generator begins to rotate, a bearing force arises that pushes the wedge, representing the flexspline teeth, towards the circular spline. This interaction results in an increase in the meshing force. The transmission of motion is ensured by the flexspline's rotational constraint, which can be observed through its angular displacement. Notably, aside from the moment of speed inversion, when dynamic transitions cause brief oscillations, the flexspline's steady-state rotation remains limited and constant, depending solely on its stiffness.

Additionally, due to the downward inertia and friction on the output shaft, which represents the external load, the transmitted torque experiences spikes during speed inversion. These spikes are necessary to accelerate the various masses within the reducer and

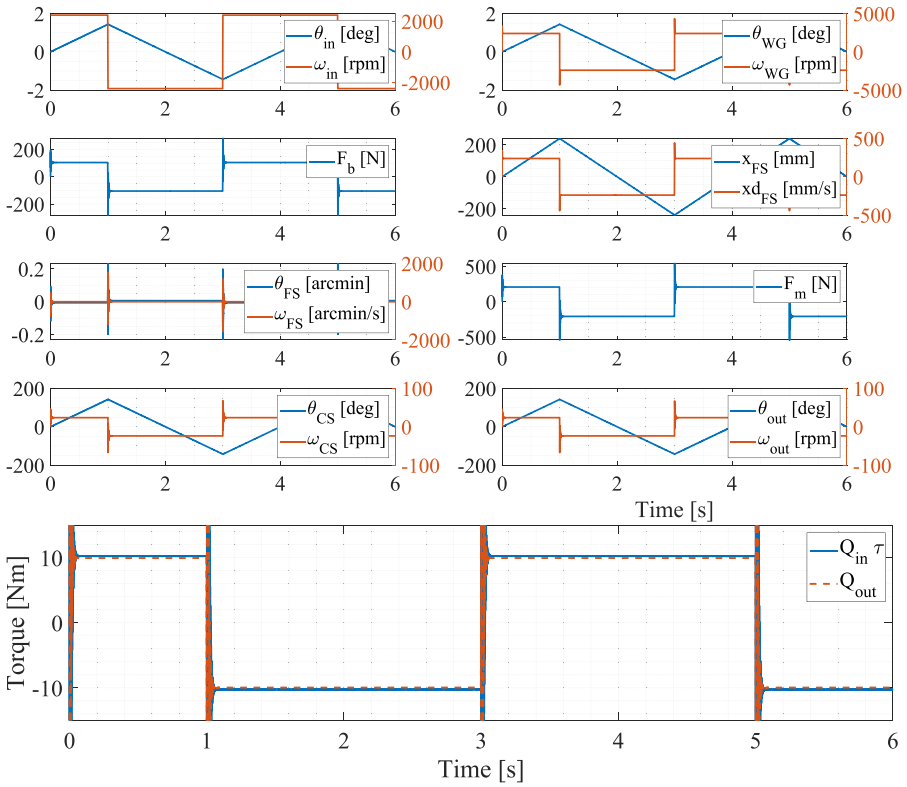


Fig. 14. Signals evolution in nominal conditions with a maximum input speed equal to 2400 rpm.

the external inertia. Under constant speed conditions, the torque significantly decreases to a level that corresponds to the frictional torque needed to overcome both internal and external resistance.

The proposed multibody approach achieves an effective balance between computational efficiency and model fidelity, making it well-suited for real-time or near-real-time applications in PHM. Unlike FEA, which provides high accuracy at the expense of substantial computational resources, this method employs simplified spring–damper systems and lumped parameters to minimize the degrees of freedom while preserving essential dynamic interactions. For instance, simulating a complete operational cycle with the proposed model takes less than 10 s on a standard computational platform, compared to several hours for an equivalent FEA model. This significant reduction in computation time does not compromise the model's ability to capture nonlinear behaviors, such as kinematic errors, hysteresis, and fault-induced dynamics, as validated by both literature and experimental data. By striking this balance, the model enables accurate and computationally feasible analyses, meeting the stringent requirements of PHM applications.

4.1. Analytical formulation of the harmonic reducer efficiency

A closed-form equation for the mechanical efficiency of the entire harmonic reducer can be derived by integrating all the constitutive equations outlined in Section 2, which pertain to the various components and their interaction forces. This resulting equation expresses efficiency as a function of several key parameters, including the transmission ratio, reducer dimensions, and operational conditions.

$$\eta_{\text{red}} = \frac{1}{\left(\frac{\tan(\alpha_t)+f_{\text{m1}}}{1+f_{\text{m1}}\cdot\tan(\alpha_t)}\cdot\left(\frac{1}{\tan(\alpha_t)}+\tau\cdot f_{\text{b1}}\right)+\tau\cdot f_{\text{b0}}\cdot\nu_0\cdot r_{\text{HR}}\cdot n_{\text{HR}}\right)}. \tag{43}$$

According to [46], also the efficiency value obtained by Eq. (43), always less the one, varies with the applied load and internal friction. The energy loss is influenced by operational speed (n_{HR}), reducer size (r_{HR}), friction parameters $(f_{\text{m1}}, f_{\text{b0}}, f_{\text{b1}})$, and used lubricant kind (ν_0). Efficiency is a critical metric, reflecting how much input power is converted into useful output power. Higher efficiency indicates lower energy losses due to friction and resistive forces.

The efficiency of the harmonic reducer η_{red} was evaluated at four different operational speeds (500 rpm, 1000 rpm, 2000 rpm, and 3500 rpm) and compared to the results presented in [47], shown in Fig. 15, tuning the friction coefficients to match the described trend.

As can be observed in Fig. 15 the highest efficiency is achieved at lower speeds (500 rpm), and as speed increases, the efficiency generally decreases. This is consistent with increased energy losses due to friction at higher speeds, which impact the performance of the harmonic reducer.

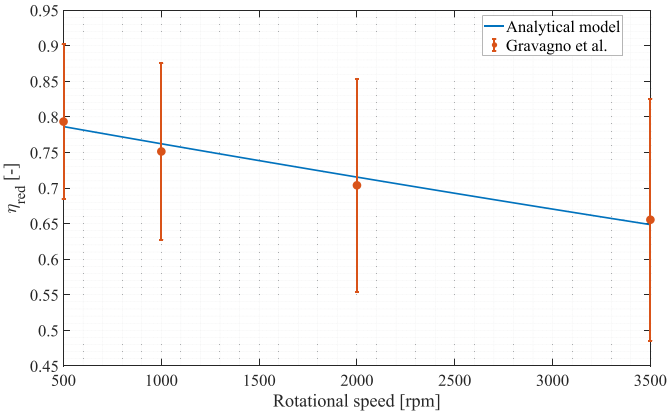


Fig. 15. Harmonic reducer efficiency as a function of different speeds with an output torque equal to 150 N m [47].

The equation proposed in (43) is independent from temperature, but future studies should focus on extending this analysis to different temperature conditions, as the friction parameters are known to be temperature-dependent [47].

4.2. Kinematic error due to wave generator misalignment

The kinematic error $(\tilde{\theta})$ is represented by the difference between the expected output value from the reducer and the actual value. This error can be expressed as:

$$\tilde{\theta} = \frac{\theta_{\text{in}}}{\tau} - \theta_{\text{out}}. \tag{44}$$

In the proposed model, by disregarding input–output shaft dynamics, it can be rearranged as:

$$\tilde{\theta} = \frac{\theta_{\text{WG}}}{\tau} - \left(\theta_{\text{CS}} + \theta_{\text{e}} + \Delta\theta\right) \approx \theta_{\text{e}} + \Delta\theta. \tag{45}$$

The kinematic error has several causes and can be decomposed into two main components [48]: a pure component (θ_{e}), related to manufacturing and assembly errors, and a component induced by the internal flexibilities (rigidities/compliances) of the system $(\Delta\theta)$.

The flexibility-dependent component is intrinsic to the model: the rigidity of the harmonic drive indeed creates the elastic delay component of the kinematic error. For the pure component, instead, it is necessary to inject its causes into the model. The main causes of the pure component of kinematic error can be traced to two categories: assembly errors (misalignments) and manufacturing errors in creating the gear profiles, which result in non-uniform motion.

The high precision of harmonic drives can be compromised by the presence of kinematic errors due, among other causes, to component misalignment. The parameter exploited in the bearing model (e_{HR}) in Eq. (13) represents the kinematic error caused by misalignment expressed as a radial displacement, and it is linked to the kinematic error (θ_{e}) expressed in radians by means of the harmonic reducer equivalent radius (r_{HR}).

The graphs in Fig. 16 show the effect of misalignment on the wave generator's kinematic profile. The left graph provides a circular plot representing the kinematic profile under both nominal (aligned) and misaligned conditions. The profiles under both conditions are almost identical, but there is a slight deviation for the misaligned condition, indicating a small perturbation in the wave propagation.

This slight deviation could be seen in the right plot, which zooms in on a very small section of the nominal and misaligned conditions. The profiles almost overlap, with only a small difference in this narrow range, highlighting that the misalignment causes a minimal but noticeable shift in performance.

The proposed model is calibrated against existing literature data [48] to assess its capability to replicate specific behaviors by adjusting the input parameters within realistic ranges. This fitting process not only enhances the model's reliability but also provides insights into how variations in these parameters influence the system's performance.

The graph in Fig. 17 compares the pure kinematic error component due to wave generator misalignment as calculated by the proposed model versus literature data [48]. The blue curve represents the proposed model, while the dashed red curve corresponds to the results from the literature. The comparison shows that the model successfully captures the kinematic error induced by WG misalignment, with values of the misalignment error expressed in milliradians over a full 360-degree angular range of the wave generator's rotation. This approach provides a comprehensive framework for understanding and quantifying the impact of WG misalignment on harmonic drive performance.

The error is plotted against the WG angular position from 0 to 360 degrees. The proposed model predicts a slightly higher error magnitude compared to the literature model, with both models showing periodic error patterns that align in shape but vary slightly in amplitude.

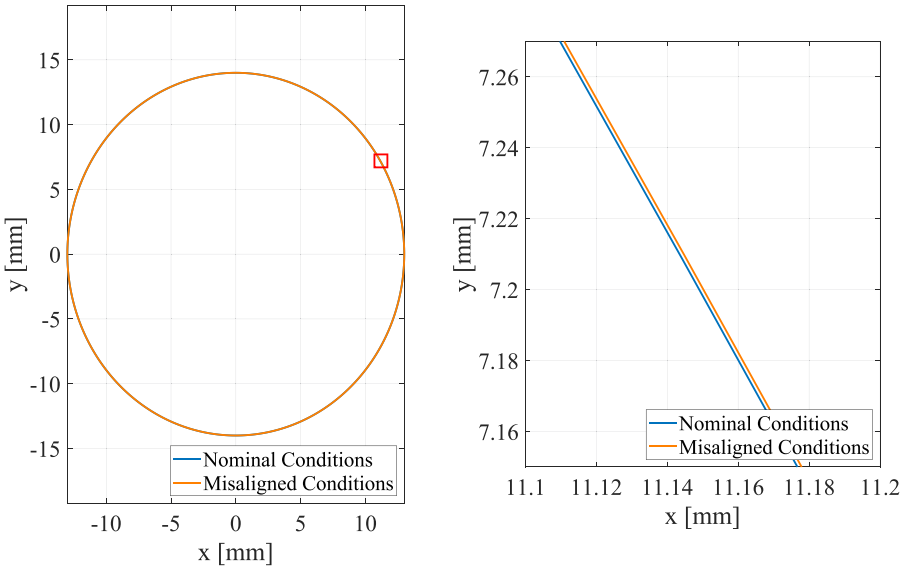


Fig. 16. Nominal and misaligned profiles calculated with the proposed model ($\underline{e} = [1, 10] \times 10^{-1}$ μm and $\underline{\varepsilon} = [1, 5] \times 10^{-1}$ deg).

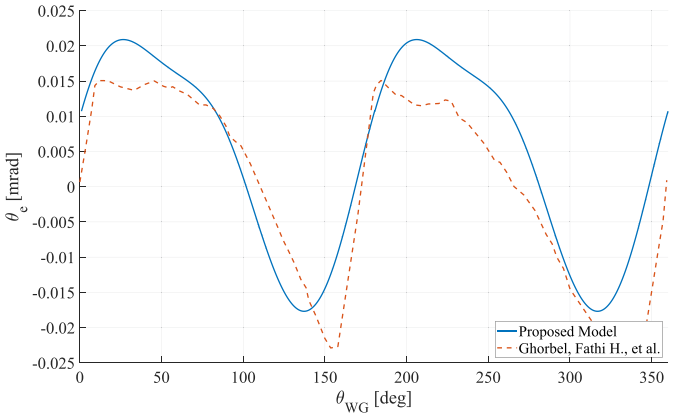


Fig. 17. Pure kinematic error component calculated with the proposed model ($\underline{e} = [1, 10] \times 10^{-1}$ μm and $\underline{\varepsilon} = [1, 5] \times 10^{-1}$ deg) compared to literature data [48].

4.3. Overall stiffness and hysteretic behavior

To analyze the overall stiffness of the reducer and its hysteretic behavior under nominal operating conditions, the rotational motion of both the input and output shafts is imposed, while the input shaft is moving the output shaft is fixed and a torquemeter is considered to evaluate the torque exerted on the fixed frame by the circular spline. This approach enables the evaluation of the torsional stiffness, which is represented as the angular displacement between the input and output shafts relative to the applied torque. The stiffness is expressed as the ratio between the transmitted torque and the resulting angular displacement (torsion).

Fig. 18 illustrates how the model accurately replicates the variation in stiffness relative to torsion, as well as the characteristic shape of the hysteresis loop. The results obtained from the simulation have been compared with the manufacturer's data provided in the catalog for the selected case study.

4.4. Meshing teeth wear

The initial investigation using the proposed faulted model focuses on analyzing the impact of tooth wear and the resulting onset of backlash. To achieve this, several simulations with increasing wear were performed, demonstrating that the model successfully replicates the experimental behavior reported in the literature [45] with a high degree of similarity. This validation confirms the model's ability to accurately capture the effects of wear on the system's performance.

The primary effect of backlash, as previously mentioned, is the introduction of a loss of motion between the driving and driven components. This loss manifests as a delay or gap in the transmission of torque, which has several significant consequences.

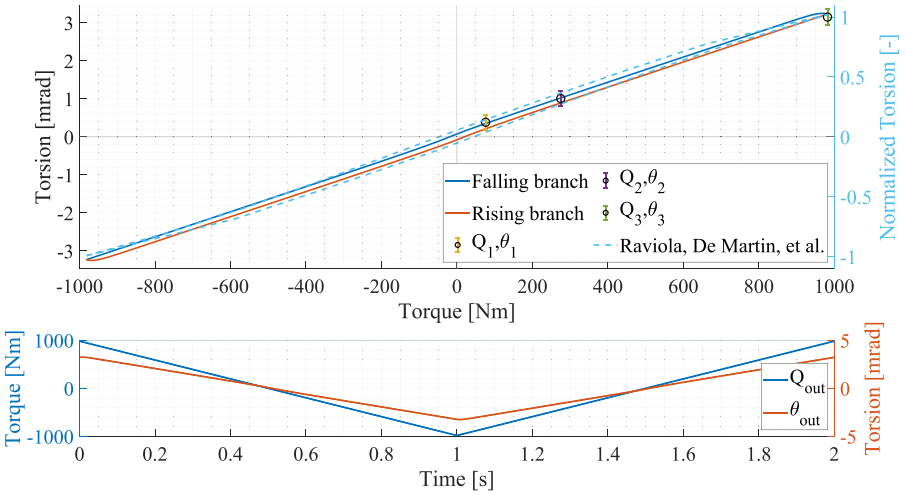


Fig. 18. Example of hysteresis curve obtained from the model simulations [45].

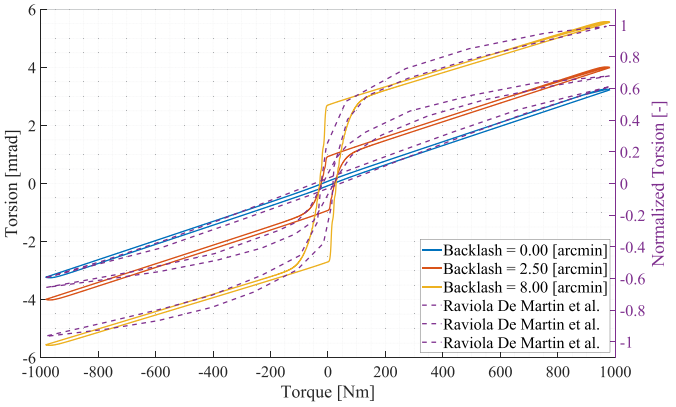


Fig. 19. Evolution of the hysteresis curve for different levels of backlash.

By comparing the obtained results with those reported in [45], it becomes evident that the effect of increasing wear is manifested primarily as a reduction in stiffness (Fig. 19). This reduction is noticeable within the system's response, particularly in the backlash region. This suggests that while wear leads to a noticeable degradation in stiffness, it does not significantly alter the slope of the curve in regions where backlash is not present.

Furthermore, the presence of backlash increases the area of the hysteresis loop, which represents the energy lost during each load-unload cycle. The larger hysteresis loop indicates greater energy dissipation due to the non-ideal, non-linear behavior of the system. This lost energy reduces the overall efficiency of the mechanical system and can exacerbate wear over time.

The backlash allows also for increased torsional deformation within the system. The gap between meshing teeth due to backlash leads to a reduction in effective tooth contact, meaning the mechanical components have more 'play' before full engagement. As a result, the maximum torsional deflection that the system can tolerate before the teeth engage is increased. This can lead to undesired oscillations, vibrations, or shock loads when the teeth finally make contact, further stressing the system.

By considering the same wear conditions, but a test in which there is no load on the CS, and it is free of motion, it is possible to analyze how the transmission ratio varies respect to wear level (Fig. 20). As previously mentioned, wear not only introduces backlash but also leads to a variation in the pressure angle. This change in the pressure angle directly affects the transmission efficiency and overall mechanical performance of the system, highlighting the critical role that wear plays in the long-term behavior of the mechanism. The direct link between the backlash level and the wear is by means of the equivalent radius $\left(bl = \frac{w_\lambda}{r_{\mathrm{HR}}}\right)$.

The progressive increase in wear does not lead to a linear variation in the transmission ratio; instead, the variation follows an exponential trend. As wear accumulates, the system deviates from nominal operating conditions at an accelerating rate, causing a more rapid deterioration in performance. This non-linear behavior highlights the criticality of early wear detection and maintenance to prevent significant functional decline.

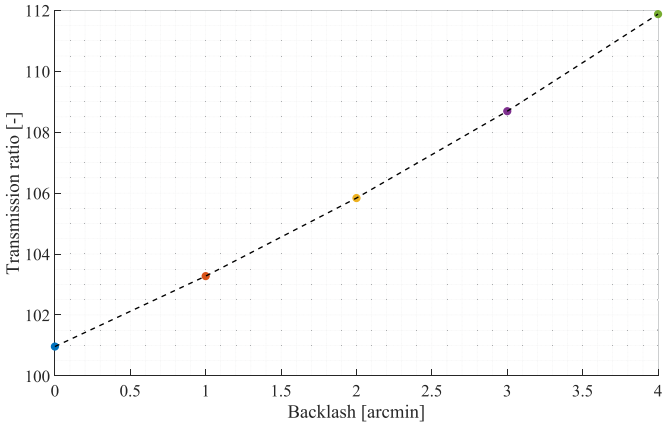


Fig. 20. Transmission ratio at different wear levels.

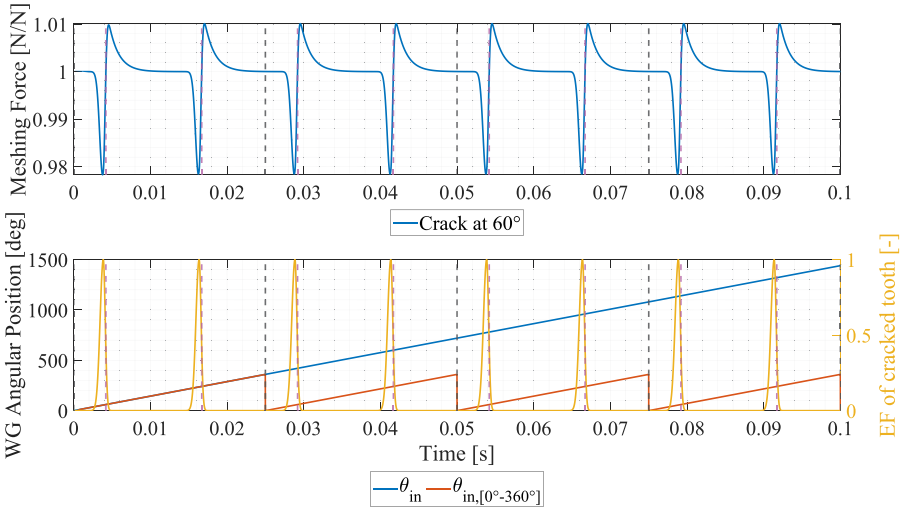


Fig. 21. Overall equivalent meshing force with and without tooth crack.

4.5. Flexspline tooth crack

The next failure mode considered is the presence of a crack in the flexspline tooth. Such a failure leads to a reduction in the stiffness of the affected tooth, eventually reducing it to zero as the crack propagates. To analyze this failure mode, a study of the meshing force is particularly useful. All the test have be conducted with a constant input speed equal to 40rps. Although this type of analysis cannot be directly assessed in a real applications, it provides valuable insights into the internal dynamics offering a deeper understanding of the effect of various sizes of the crack on the system's behavior.

By analyzing the mesh force, it becomes evident that the stiffness reduction occurs twice per revolution of the wave generator. In Fig. 21 each vertical dashed black line represents a 360° rotation of the wave generator. In fact, there are two symmetrical interaction zones in the flexspline, as shown in Fig. 3. In each rotation of the wave generator the cracked tooth passes through both of them, causing therefore an effect with a frequency twice the WG input frequency. The periodic nature of these interactions highlights the stiffness reduction due to the cracked tooth, contributing to the progressive degradation of the system's performance.

It can be observed that as a tooth passes through the meshing point, the flexspline experiences variations in the meshing force compared to the nominal operating condition, manifesting in both positive and negative directions. This phenomenon occurs because the tooth exceeds the peak of the engagement factor curve, which represents the optimal contact conditions for force transmission. When the tooth surpasses this maximum point, it triggers a rebound effect that disrupts the smooth transfer of force through the system. This rebound can lead to fluctuations in torque and can adversely affect the overall performance of the harmonic reducer.

The observed frequency of 80 Hz (the crack-beating frequency) in Fig. 22 is directly related to the rotational speed of the input shaft and the interaction of the flexspline with the crack. In the simulated scenario, the input shaft operates at a rotational frequency of 40 Hz. However, due to the system dynamics, the flexspline encounters the crack twice per revolution, resulting in a crack-induced frequency that is double the input rotational frequency [49].

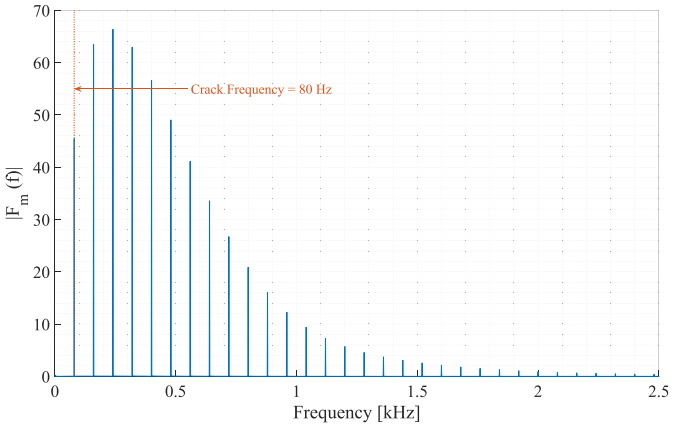


Fig. 22. Fast Fourier transform analysis of the meshing force.

This phenomenon can be explained by the periodic engagement of the flexspline with the circular spline, where the crack's position introduces a localized stiffness variation. The resulting frequency of 80 Hz reflects the system's response to this periodic interaction. This relationship is consistent with the harmonic drive configuration and the modeled fault dynamics, confirming the validity of the simulated results.

This frequency, characteristic of the periodic impact between the wave generator and the cracked tooth, is a key indicator of crack propagation within the system. While directly monitoring this frequency under real operating conditions is not feasible, it is possible to assess its effects indirectly by monitoring the speed of the output shaft and, consequently, the circular spline. This indirect monitoring allows for the identification of anomalies that may indicate fault conditions or inefficiencies within the harmonic reducer.

The amplitudes observed at other frequencies in Fig. 22 are attributed to harmonic responses and secondary vibrational modes of the system. These frequencies arise from nonlinear interactions between the flexspline, wave generator, and circular spline. These harmonics, combined with the one at 80 Hz, provide additional insight into the system's dynamic response and the redistribution of forces caused by the crack. Such information is critical for identifying secondary fault indicators and validating the robustness of the proposed model.

4.6. Flexspline rim crack

The analysis of cracks on the flexspline rim was conducted by incorporating localized stiffness reductions in the model to represent the physical effects of cracks on the rim's structural integrity. A crack was introduced as a localized region with reduced stiffness along the flexspline rim. The reduction factor depended on the crack depth and its influence on the flexspline's bending and torsional rigidity, which were calculated using standard fracture mechanics formulas.

To account for the impact of the crack, the time-varying stiffness of the system was modified, incorporating the reduced stiffness at the crack location. This modification allowed the model to capture the altered load distribution across the flexspline and its interaction with the wave generator. The simulation was then performed under steady-state input conditions, with rotational speed and load reflecting typical operating scenarios.

As the crack grows in size, it leads to a progressive reduction in the stiffness of the flexspline cup. This loss of stiffness ultimately results in improper or failed transmission of motion. To better understand the severity of this failure mode, we analyze the evolution of the angular position of the output shaft over time. The study of this progression provides insights into how the crack compromises the system's ability to maintain proper motion transmission, signaling the need for timely intervention before complete mechanical failure occurs.

As shown in Fig. 23 as the severity of the failure increases, the ability to transmit motion becomes progressively more compromised. This results in an exponential phase delay between the input shaft speed and the speed of the output shaft, reflecting the growing mechanical inefficiency. As the failure worsens, the system's capacity to maintain motion transmission deteriorates rapidly, eventually leading to complete breakage, at which point the transmission of motion ceases entirely.

5. Conclusions

This study presents a novel computationally efficient model for capturing the non-linear dynamics of harmonic drive reducers, validated against both experimental and literature data. The proposed model uniquely highlights variations in dynamic parameters that align with defect propagation trends observed in the literature, particularly in fault-related behaviors such as frequency shifts associated with crack and wear progression. By offering a high level of correlation with established data, the model successfully represents how specific faults influence the reducer performance, thus aligning well with real-world mechanical expectations.

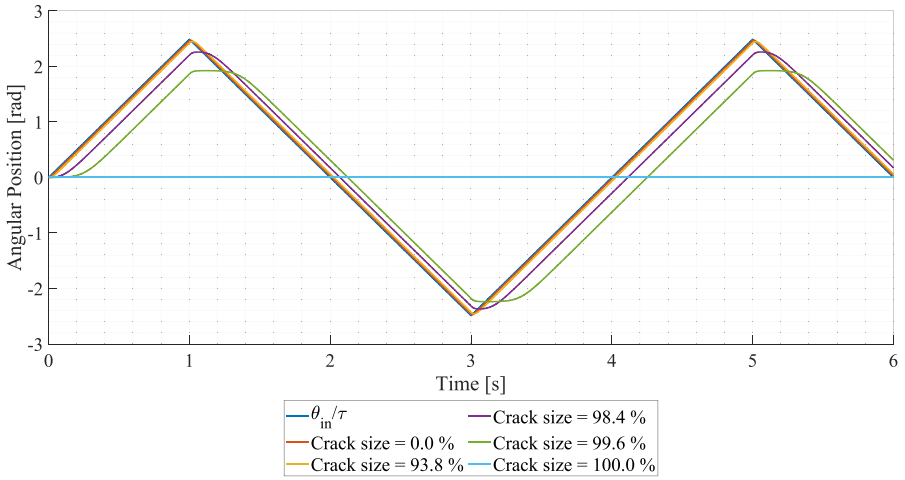


Fig. 23. Angular position of the output shaft.

A key contribution of this work is the model's adaptability across a range of applications, notably as a practical tool for PHM tasks. Its computational simplicity allows it to be integrated into applications that rely on large, sustained datasets, making it especially suitable for digital twin frameworks, where real-time performance analysis and design optimizations are essential. Furthermore, by enabling comprehensive yet manageable simulations of both nominal and fault conditions, this model supports the predictive maintenance of harmonic drives and can be used effectively in initial design phases, as well as in ongoing performance assessments. Through this balance of accuracy and efficiency, the model provides an invaluable foundation for systems requiring efficient, data-intensive simulation without the computational demands of more complex modeling techniques.

Future work could enhance the model by expanding its scope to include temperature effects, as variations in operating temperature can significantly influence material properties, particularly friction and stiffness. Incorporating a thermal model that simulates temperature-dependent changes in material behavior would enable more accurate predictions under a broader range of conditions. Additionally, further studies could explore the model's integration with advanced lubrication dynamics, capturing how lubrication degradation affects meshing efficiency and wear progression over time.

CRediT authorship contribution statement

Roberto Guida: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Antonio Carlo Bertolino:** Writing – review & editing, Visualization, Validation, Supervision, Conceptualization. **Andrea De Martin:** Writing – review & editing, Validation, Supervision, Conceptualization. **Massimo Sorli:** Supervision, Project administration, Funding acquisition.

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

No data was used for the research described in the article.

References

[1] T.D. Tuttle, Understanding and Modeling the Behavior of a Harmonic Drive Gear Transmission (Ph.D. thesis), Massachusetts Institute of Technology, 1992.
[2] C. Preissner, T.J. Royston, D. Shu, A high-fidelity harmonic drive model, 2012.
[3] K.L. Tsui, N. Chen, Q. Zhou, Y. Hai, W. Wang, Prognostics and health management: A review on data driven approaches, Math. Probl. Eng. 2015 (1) (2015) 793161.
[4] P.E. Leser, Probabilistic Prognostics and Health Management for Fatigue-Critical Components Using High-Fidelity Models, North Carolina State University, 2017.
[5] L.A. Grosso, A. De Martin, G. Jacazio, M. Sorli, Development of data-driven PHM solutions for robot hemming in automotive production lines, Int. J. Progn. Heal. Manag. 11 (1) (2020).
[6] T.D. Tuttle, W.P. Seering, A nonlinear model of a harmonic drive gear transmission, IEEE Trans. Robot. Autom. 12 (3) (1996) 368–374.

[7] F. Rhéaume, H. Champliaud, Z. Liu, Understanding and modelling the torsional stiffness of harmonic drives through finite-element method, Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. 223 (2) (2009) 515–524.
[8] C. Zou, T. Tao, G. Jiang, X. Mei, J. Wu, A harmonic drive model considering geometry and internal interaction, Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. 231 (4) (2017) 728–743.
[9] X. Zhang, T. Tao, G. Jiang, X. Mei, C. Zou, et al., A refined dynamic model of harmonic drive and its dynamic response analysis, Shock. Vib. 2020 (2020).
[10] E. Clochiatti, L. Scalera, P. Boscariol, A. Gasparetto, Electro-mechanical modeling and identification of the UR5 e-series robot, Robot. (2024) 1–23.
[11] P. Zhang, Z. Shi, B. Yu, H. Qi, Research on the EMA control method based on transmission error compensation, Energies 17 (11) (2024) 2528.
[12] D. Shendre, V. Bhole, S. Ghunage, Simulink exploration of design and control techniques for BLDC-driven compliant actuators in the legged robotics, in: 2024 Control Instrumentation System Conference, CISCON, IEEE, 2024, pp. 1–6.
[13] R. Guida, A.C. Bertolino, A. De Martin, S. Mauro, M. Sorli, Effect of wave generator misalignment on the strain wave gear hysteretic behavior in a rotary electro-mechanical actuator, in: IFToMM World Congress on Mechanism and Machine Science, Springer, 2023, pp. 154–162.
[14] R. Guida, A.C. Bertolino, A. De Martin, M. Sorli, Comprehensive analysis of major fault-to-failure mechanisms in harmonic drives, Mach. (2024).
[15] D. Ma, S. Yan, Z. Yin, Y. Yang, Investigation of the friction behavior of harmonic drive gears at low speed operation, in: 2018 IEEE International Conference on Mechatronics and Automation, ICMA, IEEE, 2018, pp. 1382–1388.
[16] H.-L. Zhu, H.-n. Wu, M. Zou, X.-p. Qin, P.-y. Song, J. Pan, Minimal tooth number of flexspline in harmonic gear drive with external wave generator, Gear Technol. 10 (2013) 58–62.
[17] H. Dong, Z. Zhu, W. Zhou, Z. Chen, Dynamic simulation of harmonic gear drives considering tooth profiles parameters optimization., J. Comput. 7 (6) (2012) 1429–1436.
[18] J. Ma, C. Li, Y. Luo, L. Cui, Simulation of meshing characteristics of harmonic reducer and experimental verification, Adv. Mech. Eng. 10 (3) (2018) 1687814018767494.
[19] Y. Zhang, X. Pan, Y. Li, G. Wang, G. Wu, Meshing stiffness calculation of disposable harmonic drive under full load, Mach. 10 (4) (2022) 271.
[20] C. Song, F. Zhu, X. Li, X. Du, Three-dimensional conjugate tooth surface design and contact analysis of harmonic drive with double-circular-arc tooth profile, Chin. J. Mech. Eng. 36 (1) (2023) 83.
[21] P. Folęga, G. Wojnar, R. Burdzik, Ł. Konieczny, Dynamic model of a harmonic drive in a toothed gear transmission system, J. Vibroengineering 16 (6) (2014) 3096–3104.
[22] P. Guay, A. Frikha, Ball bearing stiffness: A new approach offering analytical expressions, in: European Space Mechanisms and Tribology Symposium, 2015, pp. 23–25.
[23] B.J. Hamrock, W.J. Anderson, Rolling-Element Bearings, Technical Report, 1983.
[24] H. Jia, J. Li, G. Xiang, J. Wang, K. Xiao, Y. Han, Modeling and analysis of pure kinematic error in harmonic drive, Mech. Mach. Theory 155 (2021) 104122.
[25] T.A. Harris, M.N. Kotzalas, Essential Concepts of Bearing Technology, CRC Press, 2006.
[26] S. Abdan, N. Stosic, A. Kovacevic, I. Smith, N. Asati, Analysis of rolling bearing power loss models for twin screw oil injected compressor, in: IOP Conference Series: Materials Science and Engineering, Vol. 604, IOP Publishing, 2019, 012013.
[27] H. Zhang, S. Ahmad, G. Liu, Modeling of torsional compliance and hysteresis behaviors in harmonic drives, IEEE/ASME Trans. Mechatronics 20 (1) (2014) 178–185.
[28] H. Liu, P. Yan, P. Gao, Effects of temperature on the time-varying mesh stiffness, vibration response, and support force of a multi-stage planetary gear, J. Vib. Acoust. 142 (5) (2020) 051110.
[29] M. Adak, C. Guedes Soares, Effects of different restraints on the weld-induced residual deformations and stresses in a steel plate, Int. J. Adv. Manuf. Technol. 71 (2014) 699–710.
[30] F. Ikhouane, J. Rodellar, On the hysteretic bouc-wen model, Nonlinear Dynam. 42 (1) (2005) 79–95.
[31] M. Ismail, F. Ikhouane, J. Rodellar, The hysteresis bouc-wen model, a survey, Arch. Comput. Methods Eng. 16 (2009) 161–188.
[32] A.E. Charalampakis, Parameters of bouc-wen hysteretic model revisited, in: Proceedings of the 9h HSTAM International Congress on Mechanics, 2010.
[33] X. Zhu, X. Lu, Parametric identification of bouc-wen model and its application in mild steel damper modeling, Procedia Eng. 14 (2011) 318–324.
[34] F. Andary, C. Heinzel, S. Wischmann, J. Berroth, G. Jacobs, Calculation of tooth pair stiffness by finite element analysis for the multibody simulation of flexible gear pairs with helical teeth and flank modifications, Multibody Syst. Dyn. 59 (4) (2023) 395–428.
[35] A. Azzalini, The skew-normal distribution and related multivariate families, Scand. J. Stat. 32 (2) (2005) 159–188.
[36] Y. Li, G. Zhang, Y. Zhang, Thermal–mechanical coupling deformation difference analysis for the flexspline of a harmonic drive, Rev. Adv. Mater. Sci. 61 (1) (2022) 698–710.
[37] Y.C. Wang, Steel and Composite Structures: Behaviour and Design for Fire Safety, Crc Press, 2002.
[38] V.A. Narang, Heat Transfer Analysis in Steel Structures (Ph.D. thesis), Worcester Polytechnic Institute, 2005.
[39] J. Ren, H. Yuan, A dynamic wear prediction model for studying the interactions between surface wear and dynamic response of spur gears, Coat. 12 (9) (2022) 1250.
[40] A. Zmitrowicz, Wear patterns and laws of wear–a review, J. Theoret. Appl. Mech. 44 (2) (2006) 219–253.
[41] Q. Zhao, Z. Xing, J. Yuan, Z. Zhang, J. Zhu, H. Jiang, An improved modeling and numerical analysis method for tooth surface wear of double-arc harmonic gears, Mater. 15 (24) (2022) 8869.
[42] J. Wojnarowski, V. Onishchenko, Tooth wear effects on spur gear dynamics, Mech. Mach. Theory 38 (2) (2003) 161–178.
[43] W.-K. Nam, S.-H. Oh, A design of speed reducer with trapezoidal tooth profile for robot manipulator, J. Mech. Sci. Technol. 25 (2011) 171–176.
[44] L.-t. Yang, Y.-m. Shao, W.-w. Jiang, L.-k. Zhang, L.-m. Wang, J. Xu, Effects of tooth surface crack propagation on meshing stiffness and vibration characteristic of spur gear system, Appl. Sci. 11 (4) (2021) 1968.
[45] A. Raviola, A. De Martin, M. Sorli, A preliminary experimental study on the effects of wear on the torsional stiffness of strain wave gears, in: Actuators, Vol. 11 (2022) 305.
[46] I. Schafer, P. Bourlier, F. Hantschack, E. Roberts, S. Lewis, D. Forster, C. John, Space lubrication and performance of harmonic drive gears, in: 11th European Space Mechanisms and Tribology Symposium, ESMATS 2005, Vol. 591, 2005, pp. 65–72.
[47] F. Gravagno, V.H. Mucino, E. Pennestrì, The mechanical efficiency of harmonic drives: A simplified model, J. Mech. Des. 143 (6) (2021) 063302.
[48] F.H. Ghorbel, P.S. Gandhi, F. Alpeter, On the kinematic error in harmonic drive gears, J. Mech. Des. 123 (1) (2001) 90–97.
[49] O.D. Mohammed, M. Rantatalo, Dynamic response and time-frequency analysis for gear tooth crack detection, Mech. Syst. Signal Process. 66 (2016) 612–624.