A fatigue model for contacts under mixed elastohydrodynamic lubrication condition

Sheng Li*, Ahmet Kahraman

Department of Mechanical Engineering, The Ohio State University, 201 W. 19th Avenue, Columbus, OH 43210, USA

A R T I C L E   I N F O

Article history:
Received 3 June 2010
Received in revised form 3 September 2010
Accepted 30 September 2010

Keywords:
High cycle fatigue
Pitting
Life prediction
Surface roughness
Mixed lubrication

A B S T R A C T

A fatigue methodology for pitting crack nucleation is proposed for non-conformal point contacts under mixed lubrication condition where asperity contacts are prominent. The modeling methodology includes: (i) a mixed elastohydrodynamic lubrication model to determine the instantaneous surface contact pressure and shear traction distributions, (ii) a potential theory based stress formulation to compute the three-dimensional, transient, elastic stress fields into the contacting body induced by the surface stresses within the contact zone, and (iii) a multi-axial fatigue criterion to predict the crack nucleation life corresponding to the stress time histories. Given the random nature of the surface roughness profiles of the mating surfaces, a new Lagrangian–Eulerian approach is proposed to quantify the surface roughness effect on the crack nucleation position and life statistically from the deterministic predictions. Additionally, an experimental rolling contact fatigue study is performed and the results are compared to the model predictions to assess the accuracy of the proposed contact fatigue methodology.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Pitting is one of the most common fatigue failure modes observed on gears and other rolling machine elements that are subject to cyclic contact loads. These contacts generally involve combined rolling and relative sliding actions. Based on the depth of the crack nucleation location from the surface, a pitting failure can be characterized as a surface or a subsurface initiated failure. The full-film elastohydrodynamic lubrication (EHL) is often not achievable in the contacts of heavily loaded components operating under low to moderate speeds. As the contacting surfaces have roughness profiles (caused by the manufacturing finishing process) with the amplitudes comparable to the minimum film thickness, asperity interactions (metal-to-metal contacts) are commonly observed. Under such mixed EHL condition when the asperity contacts and the EHL film share the load, the roughness irregularities can increase the local surface stresses to the levels that are much higher than the maximum Hertzian pressure. These localized extreme stress concentrations appear to accelerate the occurrence of the contact fatigue failure, in the form of surface or near surface nucleated pit. For surfaces that are relatively smooth or when rolling speeds are high, continuous fluid film is more likely to form between the mating surfaces such that the metal-to-metal contacts are avoided. Under such favorable full-film EHL condition, the deviation of the contact pressure distribution from the correspond-

* Corresponding author. Tel.: +1 614 247 8688; fax: +1 614 292 3163.
E-mail address: li.600@osu.edu (S. Li).


Contents lists available at ScienceDirect

International Journal of Fatigue

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

A fatigue model for contacts under mixed elastohydrodynamic lubrication condition

Sheng Li*, Ahmet Kahraman

Department of Mechanical Engineering, The Ohio State University, 201 W. 19th Avenue, Columbus, OH 43210, USA

A R T I C L E   I N F O

Article history:
Received 3 June 2010
Received in revised form 3 September 2010
Accepted 30 September 2010

Keywords:
High cycle fatigue
Pitting
Life prediction
Surface roughness
Mixed lubrication

A B S T R A C T

A fatigue methodology for pitting crack nucleation is proposed for non-conformal point contacts under mixed lubrication condition where asperity contacts are prominent. The modeling methodology includes: (i) a mixed elastohydrodynamic lubrication model to determine the instantaneous surface contact pressure and shear traction distributions, (ii) a potential theory based stress formulation to compute the three-dimensional, transient, elastic stress fields into the contacting body induced by the surface stresses within the contact zone, and (iii) a multi-axial fatigue criterion to predict the crack nucleation life corresponding to the stress time histories. Given the random nature of the surface roughness profiles of the mating surfaces, a new Lagrangian–Eulerian approach is proposed to quantify the surface roughness effect on the crack nucleation position and life statistically from the deterministic predictions. Additionally, an experimental rolling contact fatigue study is performed and the results are compared to the model predictions to assess the accuracy of the proposed contact fatigue methodology.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Pitting is one of the most common fatigue failure modes observed on gears and other rolling machine elements that are subject to cyclic contact loads. These contacts generally involve combined rolling and relative sliding actions. Based on the depth of the crack nucleation location from the surface, a pitting failure can be characterized as a surface or a subsurface initiated failure. The full-film elastohydrodynamic lubrication (EHL) is often not achievable in the contacts of heavily loaded components operating under low to moderate speeds. As the contacting surfaces have roughness profiles (caused by the manufacturing finishing process) with the amplitudes comparable to the minimum film thickness, asperity interactions (metal-to-metal contacts) are commonly observed. Under such mixed EHL condition when the asperity contacts and the EHL film share the load, the roughness irregularities can increase the local surface stresses to the levels that are much higher than the maximum Hertzian pressure. These localized extreme stress concentrations appear to accelerate the occurrence of the contact fatigue failure, in the form of surface or near surface nucleated pit. For surfaces that are relatively smooth or when rolling speeds are high, continuous fluid film is more likely to form between the mating surfaces such that the metal-to-metal contacts are avoided. Under such favorable full-film EHL condition, the deviation of the contact pressure distribution from the correspond-

* Corresponding author. Tel.: +1 614 247 8688; fax: +1 614 292 3163.
E-mail address: li.600@osu.edu (S. Li).


Contents lists available at ScienceDirect

International Journal of Fatigue

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

A fatigue model for contacts under mixed elastohydrodynamic lubrication condition

Sheng Li*, Ahmet Kahraman

Department of Mechanical Engineering, The Ohio State University, 201 W. 19th Avenue, Columbus, OH 43210, USA

A R T I C L E   I N F A

A B S T R A C T

A fatigue methodology for pitting crack nucleation is proposed for non-conformal point contacts under mixed lubrication condition where asperity contacts are prominent. The modeling methodology includes: (i) a mixed elastohydrodynamic lubrication model to determine the instantaneous surface contact pressure and shear traction distributions, (ii) a potential theory based stress formulation to compute the three-dimensional, transient, elastic stress fields into the contacting body induced by the surface stresses within the contact zone, and (iii) a multi-axial fatigue criterion to predict the crack nucleation life corresponding to the stress time histories. Given the random nature of the surface roughness profiles of the mating surfaces, a new Lagrangian–Eulerian approach is proposed to quantify the surface roughness effect on the crack nucleation position and life statistically from the deterministic predictions. Additionally, an experimental rolling contact fatigue study is performed and the results are compared to the model predictions to assess the accuracy of the proposed contact fatigue methodology.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Pitting is one of the most common fatigue failure modes observed on gears and other rolling machine elements that are subject to cyclic contact loads. These contacts generally involve combined rolling and relative sliding actions. Based on the depth of the crack nucleation location from the surface, a pitting failure can be characterized as a surface or a subsurface initiated failure. The full-film elastohydrodynamic lubrication (EHL) is often not achievable in the contacts of heavily loaded components operating under low to moderate speeds. As the contacting surfaces have roughness profiles (caused by the manufacturing finishing process) with the amplitudes comparable to the minimum film thickness, asperity interactions (metal-to-metal contacts) are commonly observed. Under such mixed EHL condition when the asperity contacts and the EHL film share the load, the roughness irregularities can increase the local surface stresses to the levels that are much higher than the maximum Hertzian pressure. These localized extreme stress concentrations appear to accelerate the occurrence of the contact fatigue failure, in the form of surface or near surface nucleated pit. For surfaces that are relatively smooth or when rolling speeds are high, continuous fluid film is more likely to form between the mating surfaces such that the metal-to-metal contacts are avoided. Under such favorable full-film EHL condition, the deviation of the contact pressure distribution from the correspond-
dislocation pile-up theory and proposed a semi-analytical approach. Glocz et al. [7,8] included the crack propagation into the RCF modeling while considering only subsurface cracks under smooth contact condition with a user-defined friction coefficient. Flåker et al. [9] studied the surface crack propagation including the EHL effects in the form of the empirical smooth surface lubrication formulas. 

The model of Qiao et al. [10] provided improvements over those earlier studies by including the lubrication of the contacting rough surfaces forming a line contact. In the process, full-film (micro) EHL condition with no asperity contacts was considered. They compared a variety of multi-axial critical plane fatigue criteria to claim that all yielded similar predictions. This study also showed that the roughness effect on the location of the crack nucleation site and the corresponding contact fatigue life in a statistical way, considering sufficiently long measured surface roughness profile segments. As the second objective of this work, the results of an extensive RCF experimental study are presented and compared to the model predictions to demonstrate the capability of the proposed model under different contact conditions.

### 2. Modeling methodology

The overall rolling contact fatigue modeling methodology consists of three major components: (i) a mixed-EHL model for the prediction of the transient normal pressure \( p(x, y, t) \) and shear stress \( q(x, y, t) \) distributions along the contacting surface, (ii) stress distribution prediction formulation for the determination the time histories of the transient stress components in the vicinity of the contact induced by \( p(x, y, t) \) and \( q(x, y, t) \), and (iii) a multi-axial fatigue model to compute the crack nucleation lives of all points into the material. The proposed methodology combines the mixed-EHL model of Li and Kahraman [14] with a high cycle, multi-axial fatigue criterion proposed by Liu and Mahadevan [15] through a three-dimensional (3D) stress formulation. It is assumed in this study that the thermal effects on the EHL behavior are negligible and the contact is lubricated properly preventing other failure modes such as scuffing and wear. The changes to the surface roughness profiles during the run-in stage are
excluded by using the after run-in surface roughness profiles and assuming that the profiles remain the same after the run-in takes place [16,17]. Further, only the crack nucleation life is included here as the experimental studies such as Rahman et al. [18] and Hoffmann et al. [19] indicated that the crack propagation life in high cycle RCF was rather small compared to the crack nucleation life. Meanwhile, pure elastic stress fields are assumed in this work, since the high cycle RCF problem is of primary interest with the contact fatigue lives in excess of (10)⁶ cycles. The measured residual stress fields are superimposed onto the load induced stress components prior to the multi-axial fatigue evaluation. Any potential effect of the variation of the material hardness with depth is not taken into account in the present study.

2.1. Point contact mixed-EHL model

Accurate description of the tribological behavior under mixed EHL condition is one of the most challenging tasks in the contact fatigue modeling due to the numerical instability encountered under severe asperity contact condition and the computational time required. Robust mixed-EHL models of Hu and Zhu [20], Zhu [21] and Ren et al. [22] were shown to yield converged pressure and film thickness distributions for lubrication conditions ranging from full film to boundary condition. One potential shortcoming of these models was identified as the dependence of the accuracy of their solutions on the grid density especially when the rolling speed is low [23]. Recently, Li and Kahraman [14] proposed a point contact transient mixed-EHL model using an asymmetric integrated control volume discretization scheme to effectively reduce the dependence of the solution accuracy on the grid mesh size, in the process increasing the numerical accuracy and the computational efficiency. The same point contact mixed-EHL model is employed here to predict the transient contact pressure, film thickness and shear stress distributions along the contacting surfaces of the rolling components. Only the governing equations of this model are provided here as the other details of the model including the discretization scheme can be found in Ref. [14].

In the presence of asperity contacts, the transient Reynolds equation governs the fluid flow that circumvents the islands of asperity contacts. With the spatial coordinates x and y in the direction of rolling and in the direction normal to the rolling direction, respectively, the transient Reynolds equation reads

\[
\frac{\partial}{\partial x} \left( f_x \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left( f_y \frac{\partial p}{\partial y} \right) = \frac{\partial (u_t \rho h)}{\partial x} + \frac{\partial (\rho h)}{\partial t}
\]

(1a)

where t is the time. The parameters p, h and \( \rho \) denote the pressure, thickness and density of the fluid, respectively, which are all functions of x, y and t. The rolling velocity is defined as \( u_t = (u_1 + u_2)/2 \), where \( u_1 \) and \( u_2 \) are the surface velocities of the contacting body 1 and 2 in the x direction. \( f_x \) and \( f_y \) are the flow coefficients in the x and y directions, that are defined for a Re-EErying fluid as [14]

\[
f_x = \rho h^3 \frac{\cosh (\tau_m/\tau_0)}{12 \eta f}
\]

(1b)

\[
f_y = \rho h^3 \frac{\sinh (\tau_m/\tau_0)}{12 \eta f \tau_m/\tau_0}
\]

(1c)

Here, \( \eta \) is the lubricant viscosity, \( \tau_0 \) is the lubricant reference stress, and \( \tau_m \) is the viscous shear stress determined by \( \tau_m/\tau_0 = \sinh^{-1}[\eta u_t/(\tau_0 h)] \), \( u_t = u_1 - u_2 \) is the sliding velocity.

For the regions where the asperities of the mating surfaces contact each other, a reduced from of the Reynolds equation [11,12,14,20–22,24] takes effect as

\[
\frac{\partial (u_t \rho h)}{\partial x} + \frac{\partial (\rho h)}{\partial t} = 0
\]

(2)

Assuming a smooth transition between the fluid and asperity contact areas, this unified Reynolds equation system defined by Eqs. (1a) and (2) governs the mixed EHL behavior of the contact, considering both the asperity contact and fluid regions simultaneously.

The local film thickness of any contact point (x, y) at time t is defined as

\[
h(x, y, t) = h_0(t) + g_0(x, y) + V(x, y, t) - R_1(x, y, t) - R_2(x, y, t)
\]

(3)

where \( h_0(t) \) is the reference film thickness, \( R_1(x, y, t) \) and \( R_2(x, y, t) \) are the roughness heights of the two surfaces at time t, and \( g_0(x, y) \) is the gap between the two surfaces before any elastic deformations occur. Additionally, \( V(x, y, t) \) represents the sum of the elastic deformations of the two contacting bodies due to the normal load applied, and is defined by Boussinesq’s half space formulation as [25]

\[
V(x, y, t) = \int \int K(x - x', y - y') p(x', y', t) dx' dy'
\]

(4)

where S is the computational domain of the contact zone, and \( K(x,y) \) is the influence function. Recognizing that Eq. (4) is the convolution operation between \( p(x, y, t) \) and \( K(x,y) \), the Discrete Fourier Transform (DFT) convolution technique is applied here to improve the computation efficiency substantially.

A load balance equation is written by equating the total contact force due to the pressure distribution (both hydrodynamic and asperity contact) over the entire contact area to the normal load applied as

\[
W = \int \int p(x, y, t) dx dy
\]

(5)

where W is the applied normal load. While not directly required for the solution of Eqs. (1a) and (2), this equation is required in the determination of the film thickness. The value of \( h_0(t) \) in Eq. (3) must be adjusted within a load iteration loop until the predicted \( p(x,y, t) \) is such that Eq. (5) is satisfied. The viscosity–pressure and density–pressure relationships required by the EHL simulation are given in Ref. [14].

Once the converged \( p(x, y, t) \) and \( h(x, y, t) \) are obtained, the transient shear traction between the contact surfaces, which consists of: (i) the viscous shear within the hydrodynamically lubricated areas, and (ii) the contact friction due to the direct asperity interactions, can be determined. Assuming no slippage at the lubricant and solid surface interfaces and considering both the Poiseuille and Couette flows, the viscous shear stress acting on one of the surfaces (body 1) is written as

\[
q(x, y, t) = -\frac{1}{2} h(x, y, t) \frac{\partial p(x, y, t)}{\partial x} - \eta_i^* \frac{u_t}{h(x, y, t)}
\]

(6)

where \( \eta_i^* = \eta / \cosh(\tau_m/\tau_0) \) is the effective viscosity for a Re-Erying fluid. Meanwhile, within the contact regions where the film thickness breaks down, the shear stress is defined as \( q(x, y, t) = \mu p(x, y, t) \) where \( \mu \) is the friction coefficient between the asperity contacts. A typical value of \( \mu = 0.1 \) under the boundary lubrication conditions [26–28] is used in this study.

2.2. Stress prediction model

Fig. 1a shows the three-dimensional computational grid containing the contact zone. This Eulerian (fixed in space) 3D computational domain contains \( N_x \times N_y \times N_z \) grid elements, and its dimensions are dictated by the half Hertzian widths a and b in the x and y directions such that \(-1.875a \leq x \leq 1.125a, -1.5b \leq y \leq 1.5b \) and \( 0 < z < a \), where the z axis points down into the material, representing the depth. The grid increment \( \Delta z \) is designed to be variable to allow a finer grid in the regions near the surface to capture the local variations induced by the surface.
roughness, while the $\Delta x$ and $\Delta y$ increments in Fig. 1a are kept constant and sufficiently small to capture the surface roughness profiles in both directions.

When a certain material point enters the computational domain of Fig. 1a as the contacting bodies roll, its stress components due to the instantaneous surface normal pressure and shear distributions must be predicted. By modeling the contact at any time instant $t$ as an elastic half space subjected to both $p(x, y, t)$ and $q(x, y, t)$, the transient stress states below the contact surface can be determined by using the classical potential theory [25, 29, 30] as

$$
\bar{\sigma}_{ijk,t} = \sum_{m=1}^{N_x} \sum_{n=1}^{N_y} (\bar{S}_{ij,m} p_{mnt} + \bar{T}_{ij,m} q_{mnt})
$$

where $\bar{\sigma}_{ijk,t}$ is the stress tensor at an arbitrary grid point $ijk$ with coordinates $(x, y, z)$ at time $t$. Here $p_{mnt}$ and $q_{mnt}$ are the normal pressure and shear traction acting on the surface grid node $mn1$, and $\bar{S}_{ij,m}$ and $\bar{T}_{ij,m}$ are the influence coefficient tensors corresponding to the normal and tangential loadings, respectively, representing the stress states under the surface node $ij1$ induced by the unit normal pressure and shear stress exerted on the surface node $mn1$. The closed-form formulations of $\bar{S}_{ij,m}$ and $\bar{T}_{ij,m}$ can be found in Ref. [30]. Eq. (7) also has the form of discrete convolution and can be evaluated in a way as that of the surface elastic deformation of Eq. (4) using the DFT convolution technique.

Regardless of which multi-axial fatigue criterion is used [4, 15, 31–38], the mean and alternating values of the shear and normal stresses of each material point are required to assess its fatigue damage. At each time instant $t$, the stress components $\bar{\sigma}(x, y, z, t)$ within the Eulerian control volume (Fig. 1a) are recorded according to Eq. (7). Introducing a Lagrangian reference frame $XYZ$ that is attached to the moving contacting body $t$ ($t = 1, 2$) and stating the relationship between the two frames as $x = X + u \cdot t, y = Y$ and $z = Z$, the time history of the stress tensor for the material point $(X, Y, Z)$ is found as

$$
\bar{\Sigma}(X, Y, Z, t) = \bar{\sigma}(X + u \cdot t, Y, Z, t)
$$

Any residual stresses caused by the surface machining and heat treatment processes (measured along the $z$ axis) can be superimposed onto the predicted elastic stress fields, which alters the mean values while leaving the alternating stress amplitudes unchanged.

2.3. Multi-axial contact fatigue life model

The fatigue literature contains numerous multi-axial fatigue criteria, most of which employ a critical plane based method, perhaps because of reasonable correlation between the predictions and the experimental fatigue data [31–37]. However, these fatigue criteria are usually limited to narrow ranges of materials and loading conditions according to the different definitions of the critical plane and the damage parameter. No widely accepted high cycle multi-axial fatigue criterion is available for the rolling contact fatigue problem considered in this work. Liu [4] and Liu and Mahadevan [15] proposed a characteristic plane fatigue criterion for railroad wheel contacts that was claimed to be applicable to relatively wide ranges of materials and loading conditions through comparisons to the measured fatigue data for a variety of metals under both proportional and non-proportional loads. This approach assumes the fatigue evaluation can be approximated by using a certain stress combination (damage parameter) on a certain plane, namely characteristic plane, which is not necessarily the macro crack plane (fatigue fracture plane). The fatigue criterion applied on this characteristic plane is [4]

$$
\frac{1}{\bar{b}} \left[ \bar{\sigma}_{1}^{2} + \frac{\bar{S}_{1y}^{2}}{\bar{S}_{2y}^{2}} (\bar{\tau}_{1}^{2} + \bar{\tau}_{2}^{2}) + \kappa \bar{\sigma}_{3}^{2} \right]^{\frac{1}{2}} = S_{N}^{0}
$$

where $S_{N}^{0}$ and $S_{N}$ are the uni-axial fully reversed bending and fully reversed torsion fatigue strength of the material corresponding to the finite fatigue life cycles of $N$, $\sigma_{1}$, $\tau_{1}$ and $\tau_{2}$, and $\sigma_{3}$ are the normal stress amplitude, amplitudes of the shear stress components and the hydrostatic stress amplitude acting on the characteristic plane, respectively, and $\bar{b}$ and $\kappa$ are the material property related

---

Fig. 1. (a) The computational domain and grid mesh used in this study and (b) the definition of an Eulerian $yz$ line with the consecutive material points passing through the computational domain.
parameters. In order to determine the position of this characteristic plane, the macro crack plane which was assumed to be the plane experiencing the maximum normal stress amplitude was used as a reference plane. Defining the characteristic plane to be the plane on which the damage introduced by the hydrostatic stress amplitude is minimum, i.e. $\kappa = 0$, the angle between the characteristic plane and the macro crack plane $\beta$ and the material parameter $\beta$ were derived in Refs. [4,15] by implementing Eq. (9) to the fully reversed bending and torsion fatigue problems as

$$\alpha = \frac{1}{2} \cos^{-1} \left[ -\frac{s^2 + \sqrt{s^4 - (1 - 3s^2)(1 - 5s^2 - 4s^4)}}{1 - 5s^2 - 4s^4} \right] \quad \text{(10a)}$$

$$\beta = \sqrt{s^2 \cos^2(2\alpha)} + \sin^2(2\alpha) \quad \text{(10b)}$$

for non-extremely brittle materials with $s = S'_b/S'_n < 1$. As the loading characteristic of the RCF problem in hand is rather similar to that of the railroad wheel contacts considered by Liu [4] and Liu and Mahadevan [15], this fatigue criterion is adapted here. On the other hand, the critical plane criteria [31–37] involve the computation of the shear stress amplitude, which is defined as the radius of the minimum circle circumscribing the path described by the tip of the shear stress vector on a certain plane [38]. One efficient and robust algorithm for the search of the minimum circumscribed circle (MCC) was recommended by Bernasconi and Papadopoulos [38], called the randomized algorithm. Considering that a sufficiently large number of points (say more than 200) along a shear stress path must be included, and acknowledging that the MCC search is required for several thousands of shear stress time histories on every possible material plane in the 3D space at every time instant (say a total of 1000 time steps), this algorithm is still computationally unaffordable for the lubricated rough surface rolling point contact problem. The fatigue criterion of Liu [4] decomposed the shear vector on the characteristic plane into two perpendicular components, whose amplitudes are simply defined as the half of the shear range in each of the two directions, eliminating the time consuming task of MCC search.

Due to the above reasons, the criterion of Liu [4] in its modified form [15] is adopted here to evaluate the crack nucleation fatigue lives of the material points within the contacting component. However, the methodology outlined in this work is general such that any other multi-axial fatigue criterion can be used in its place if deemed appropriate. Introducing the correction term $(1 - \sigma_{m,\text{max}}/S'_n)$ to include the mean normal stress effect, the fatigue criterion defined on the characteristic plane ($\kappa = 0$) is rewritten as [15]

$$\frac{1}{\beta(1 - \sigma_{m,\text{max}}/S'_n)} \left[ \sigma'_2 + \left( \frac{S'_N}{S'_n} \right)^2 (\tau^2_{z1} + \tau^2_{z2}) \right]^{1/2} = S'_n \quad \text{(11)}$$

Here $\sigma_{m,\text{max}}$ is the mean normal stress on the macro crack plane (assumed as the plane experiences the maximum normal stress amplitude) and $S'_n$ is a reference stress defined by the uni-axial material fatigue data (in the absence of such material data, $S'_n$ can be approximated using the ultimate tensile strength value). In this work, the static yield strength of the material is used as the reference stress such that $S'_n = 2$ GPa. According to the characteristic plane approach [15], the macro crack plane passing through the material point of interest is searched in the three-dimensional space using an Euler angle based axis transformation [4]. Once the macro crack plane is located, the orthogonal coordinate system $x' - y' - z'$ is defined such that the $z'$ direction corresponds to the normal vector of the macro crack plane and $y'$ points into the direction in which the shear stress amplitude on this crack plane is maximum. The characteristic plane is then arrived at through a rotation about the $x'$ axis by an angle $\alpha$ from the macro crack plane [4,15], forming a new frame of $x'' - y'' - z''$, where $x''$ is the normal vector of the characteristic plane and $y''$ is on the characteristic plane perpendicular to $x''$. $\tau_{z1}$ and $\tau_{z2}$ in Eq. (11) are the shear stress amplitudes in the $x''$ and $y''$ directions, respectively. Finally, the fatigue lives are determined by solving Eq. (11) numerically.

As illustrated in Fig. 1b for the contacting body 1, the Lagrangian material points $a_m (m = 1, \ldots, M)$ passing through the computational domain along an Eulerian line $jk$ having $y = y_j$ and $z = z_k$ (i.e. at the same depth and the same axial coordinate) would have identical fatigue lives if the contacting surfaces were perfectly smooth such that $p(x, y, t) = p(x, y)$ and $q(x, y, t) = q(x, y)$. For the general case of rough surfaces, however, the variation of the fatigue lives of the population of the material points $a_m$ traveling in the $x$ direction is dictated by the statistical characteristics of the roughness profiles $R(x, y, t)$ and $h(x, y, t)$ of the mating surfaces (moving at unequal speeds of $u_1$ and $u_2$, respectively) as defined in Eq. (3). In order to capture the variations in the fatigue life caused by the instantaneous matching of the surface roughness profiles, the life predictions are performed for a period of time (large $M$) to allow sufficiently long segments of the measured surface profiles (8 mm for the slower component and 10 mm for the faster component in this work) to travel through the contact zone. The fatigue lives of all the points $a_m$ passing through the computational domain at the same $y$ and $z$ coordinates are predicted using the methodology presented above and used to establish the probability density distribution of the crack nucleation life at $(y, z)$. Repeating the same procedure for the other Eulerian lines, the life probability distributions along the $yz$ plane are established, allowing a statistical description of the crack nucleation site and the corresponding life.

### 3. Rolling contact fatigue experiments

The rolling contact test machine used to carry out the RCF experiments to generate the fatigue data for the assessment of the accuracy of the proposed modeling methodology consists of a test bed that rotates the main rotor at a user-defined speed and test fixtures connected to the rotor that operate the rollers under certain load and speed conditions. A schematic layout of the loading scheme is provided in Fig. 2a to describe the configuration. Here, two rollers, one on the left hand side (left roller) and one on the right hand side (right roller) are pushed against two disks stacked on the center shaft (that is connected to the main rotor of the test bed) by a pair of levers that are powered by pneumatic cylinders at the free ends. As shown in Fig. 2b, the left roller comes into contact with one of the disks on the center shaft while the other disk is in contact with the right roller. The shafts of the left and right rollers are also connected to the center shaft via two gear pairs as shown in the same schematic (this arrangement is often referred as the geared roller contact set-up). By choosing the ratio of these gear pairs intentionally different from the ratio of the radii of the roller-disk pair, a fixed value of the desired slide-to-roll ratio ($SR = u_i/u_o$) can be introduced to the roller-disk contact. This test set-up allows the left and right roller contacts to have exactly the same test conditions so that two independent tests can be performed simultaneously.

The roller and disk specimens used in this work are made of a typical gear steel (SAE 4620M) and designed to have the radii of 9.525 and 28.575 mm, respectively. The undesirable edge loading condition at the axial edges of the roller-disk contact zone is a main concern in the RCF tests, as many earlier tests of this kind used simple cylindrical specimens (flat profiles in the axial direction) leading to results subject to severe edge loading conditions [39–46]. Use of a circular crown in the axial direction is neither advantageous since it was reported to cause excessive surface wear with the same set-up used in this study [47]. In order to achieve a
uniform axial load distribution profile within the entire loading range and avoid the potential wear problem that hampers the efforts to collect pitting data [48], an exponential correction similar to those applied to cylindrical roller bearings [49] is implemented to the disk in the y direction while the roller specimens have purely cylindrical shapes. The exponential axial modification shown in Fig. 3a is rather flat at the center of the contact, increasing exponentially near the edges of the disk and is defined mathematically as

\[ g_y(y) = C(e^{-Dy} + e^{Dy} - 2), \quad y \in [-2.8, 2.8] \text{ mm} \]  

(12)

where \( C = 0.15e^{-Dy/2} \text{ mm} \) and \( D = 2.2 \text{ mm}^{-1} \) with the disk width \( F = 5.6 \text{ mm} \). Under the smooth and dry contact conditions, the pressure distributions between the cylindrical roller and the mating disk having this type of lead modification are illustrated in Fig. 3b at both a low loading level and a high loading level to demonstrate that no edge loading conditions are present and uniform pressure profiles are achieved along the y direction.

When manufacturing the roller and disk specimens, the textures of the roller and disk surface roughness, both the amplitude and the lay direction, are of primary importance. Cylindrically ground specimens [39–45] are not considered here since the resultant circumferential machining marks (grooves) leads to the maximum roughness in the axial direction. It is desirable to have such grooves in the axial direction to be able to represent the contacts of the components such as gear applications since the orientation of the machining marks was shown to have substantial impact on RCF failure [45]. While the axial roller grinding techniques were reported by Patching et al. [50] and Alanou et al. [51], an alternate finishing method is devised here to simulate the shaved gear surface finishes on the roller and disk surfaces. Fig. 4a shows the close-up views of a pair of roller and disk specimens, showing the directionality of the surface marks.

All the tests are performed using an automatic transmission fluid with its inlet temperature controlled at one of the two levels of 90 and 60 °C. The ratio of the gears in Fig. 2b is such that the roller-disk pair operates at \( SR = -0.25 \). The disk speed is maintained at 2500 rpm to achieve a rolling speed of \( u_r = 6.6 \text{ m/s} \). The surface hardness of the specimens is 59 HRC. A set of typical surface roughness profiles in the direction of rolling and sliding (x direction) are shown in Fig. 4b. These measurements were obtained using a
surface roughness profiler (Talysurf, Taylor-Hobson). The upper and lower cut-off wavelength values of the surface profiler were set at 2.5 μm and 0.8 μm. A composite surface roughness value of about $R_q = \sqrt{R_{qr}^2 + R_{qd}^2} \approx 0.6 \mu m$ is achieved for each roller-disk pair where $R_{qr}$ and $R_{qd}$ are the root-mean-square roughness amplitudes of the roller and disk in the rolling direction, respectively.

Before each test, a 1-h (350,000 roller cycles) run-in period is carried out at the half of the intended load and the same speed. Based on the life expectancy under a certain test condition, 5–7 interim inspections are performed to monitor any changes of the contact surfaces during the test. Interim inspections include roller lead trace measurements on a coordinate measurement machine to quantify the amount of wear depth accumulated and the surface roughness profile measurements on a surface roughness profiler to monitor the changes of the surface profiles. A pit size of 1 mm² is used as the failure criterion. Any test where excessive wear occurs (i.e. the wear depth exceeds 5 μm) is considered to be a wear failure and excluded from the pitting database. The great majority of the pits including the ones shown in Fig. 5 formed before the last interim inspection. Therefore for practical purposes, these cycles can be deemed very close to crack initiation lives. Diagnostics experiments of Refs. [18,19] on high cycle RCF also support this observation that crack propagation lives are very short.

Fig. 5a shows the example pits formed at three different load levels with the maximum Hertzian pressure $p_h = 2.72$, 2.41, 2.24 GPa, respectively. The oil inlet temperature is 90°C and the other operating parameters are as specified above. Two SEM images of the representative pits, as shown in Fig. 5b, indicate a 30° growth of the cracks into the material, which turn parallel to the contact surface at about 0.3 mm depth and then propagate back to the surface, forming the pits, indicative of a typical surface nucleated pitting failure mechanism [19].

4. Simulation of rolling contact fatigue experiments

The rolling contact fatigue tests described in the previous section is simulated here using the numerical methodology presented in Section 2. When comparing the predictions with the experimental measurements, it is assumed that the crack propagation life to grow a pit to the size specified above is negligible compared to the crack nucleation life under the high cycle contact fatigue condition [18,19]. The baseline contact condition is defined as $u_r = 6.6 \text{ m/s}$, SR = 0.25, lubricant inlet temperature of 90°C and $R_q = 0.6 \mu m$ with four loading levels of $p_h = 2.72, 2.41, 2.24$ and 1.90 GPa. The simulations are repeated at an oil temperature of 60°C as well. The computational domain of Fig. 1a is discretized into over half a million mesh elements ($N_x = 256$, $N_y = 128$ and $N_z = 20$). The simulations cover around 8 and 10 mm long surface roughness segments for surface 1 and 2, respectively, which are sufficiently long to represent the surface irregularity characteristics. The grid resolution employed in the y direction is somewhat coarser, since both the contact surfaces are textured axially along the y direction resulting in limited variations along the y axis [14]. The fully reversed bending fatigue strength of the gear steel used (SAE 4620M) was provided by the project sponsor as $S_{fb} = 2500 N/\text{mm}^2$.
MPa, and the fully reversed torsion fatigue strength is assumed to be 65% of the corresponding bending strength.

It is found the shear stress components $\sigma_{xy}$ and $\sigma_{yz}$ are very small (on the order of $10^{-4}$ GPa) compared to the other components, and hence, the plane strain condition is used to reduce the computation. Fig. 6 shows the transient solutions of the baseline contact for $p_h = 2.24$ GPa, including $p(x, 0, t_1)$, $h(x, 0, t_1)$ and $q(x, 0, t_1)$ along the mid-plane of the computational domain ($y = 0$) as well as the corresponding components of the transient stress fields on the vertical $xz$ plane at $y = 0$ at a time instant of $t = t_1$. It is observed in Fig. 6a that $h(x, 0, t_1)$ goes down to zero at several x values, indicating asperity contacts with sharp spikes in $p(x, 0, t_1)$ and $q(x, 0, t_1)$ in Fig. 6a and in $q(x, 0, t_1)$ in Fig. 6b. Defining the top 15 mm thick layer as the surface layer ($z < a/20$) and the subsurface region as $z > a/20$, severe stress concentrations are shown to exist in Fig. 6c–f within the surface layer, induced directly by the surface roughness. Especially, for the orthogonal shear stress component $\sigma_{xz}$ whose distribution is shown in Fig. 6f, its amplitude in the surface layer is elevated to the level comparable to its maximum alternating amplitude within the subsurface region. The earlier literature [39–45] showed that the subsurface crack nucleation was related to the maximum amplitude of the orthogonal shear stress. Thus, the possibility of a surface or near surface nucleated failure is increased with the presence of surface irregularities. It is also noted in Fig. 6c–e that $\sigma_x$, $\sigma_y$ and $\sigma_z$ all have negative values since the contacting bodies are under compression. The measured compressive residual stresses are added to these elastic stress fields before the fatigue analysis [39–45].

In order to show the transient nature of the contact fatigue analysis of lubricated rough surface contacts, the mixed EHL solutions at three different time instants are shown in Fig. 7 under the same contact conditions as in Fig. 6. Significant variations in $p(x, 0, t)$ and $h(x, 0, t)$ are observed with time $t$. Approximately, 21% of the Hertzian contact area is subject to asperity contacts in Fig. 7a while this ratio is about 10% and 4% for the other two time instants in Fig. 7b and c. This implies considerable variations of the resultant stress fields with $t$ as well.

Referring to Fig. 1b, the histograms of the fatigue lives of the populations of the material points traveling through the computational domain along the same $yz$ lines are constructed. For the baseline condition with $p_h = 2.41$ GPa for example, the histograms...
of the crack nucleation lives at different depth of $(y, z) = (0, 0.5), (0, 0.32), (0, 0.71), (0.0129)$ and $(0, 0.213) \mu m$ are compared in Fig. 8. Here, the probability distributions of the fatigue lives at different $z$ obey the lognormal distribution. As the depth increases, the crack nucleation life probability distribution moves to the right, indicating longer fatigue life. It is also found that the standard deviation of the logarithms of the fatigue lives become smaller when the depth increases, due to the diminishing effects of the surface roughness into the material along the $z$ direction. These trends are valid for the other $y$ positions as well.

Using the probability density distribution for every Eulerian $yz$ line, the lower 10th percentile, median and upper 10th percentile of these distributions are computed and plotted in Fig. 9a–c, respectively, for the baseline condition with $p_h = 2.41$ GPa. The corresponding life distribution contours for the case of lower oil inlet temperature ($60 ^\circ C$) are presented in Fig. 10a–c. In Fig. 9, the roller specimen is predicted to fail through surface initiated cracks at all three percentiles under the baseline condition ($90 ^\circ C$). In Fig. 10 with the $60 ^\circ C$ oil, the increased film thickness and reduced amount of asperity contacts result in an increase in the possibility of subsurface nucleated failure. The contact fatigue lives are predicted to increase with the reduced oil temperature as well. Under perfectly smooth contact condition (with no surface stress concentration), the maximum amplitude of the alternating orthogonal shear stress occur far below the surface, leading to subsurface initiated failure. With the introduction of the surface irregularities, the extreme stress concentrations on or near surface shortens the fatigue lives of that area. Thus, the contour plots in Figs. 9 and 10 show two critical depths (near surface and subsurface) that compete with each other, leaving a transition in between. Lubrication and surface roughness conditions, therefore, dictate whether surface or subsurface nucleated pits should form.

In Fig. 11a, the median fatigue life predictions of the proposed model are compared with the experimental results in a stress-life curve format. The measurements at four load levels with the maximum Hertzian pressure of $p_h = 2.72, 2.41, 2.24$ and $1.90$ GPa are represented by 22 data points. These relatively high contact pressures represent the contact stress range for certain automotive transmission gears. For the highest three load levels, the fatigue data are reasonably well-populated, while the same cannot be said for the lowest load of $1.90$ GPa that is represented by only two data points. This is because the life of the roller reaches the level of $(10)^8$ cycles, which requires nearly 13 days of testing time at the specified operating speed. The agreement between the predicted and measured results is rather good in Fig. 11a. Furthermore, all failures are predicted to nucleate on the surface, matching the experimental observations such as the pits shown in Fig. 5.

Another comparison between the model and the experiment is provided in Fig. 11b for the case of $60 ^\circ C$ oil temperature. Here, the number of the experimental data points is rather limited due to the longer lives observed. It can also be stated that the surface wear is
more pronounced in some of the tests of this group, which caused the suspension of several tests due to the excessive wear. Only the data points that produce pits without excessive wear are included in Fig. 11b. The model predictions are again in good agreement with the limited data available under this condition. The comparison of the results between Fig. 11a and b shows that the crack nucleation fatigue lives of the roller specimens are more than doubled when the oil inlet temperature is reduced from 90 to 60 °C. This is a direct result of the increase in the lubricant viscosity from 6.61 to 15.39 Cst that impacts the lubrication conditions positively.

5. Conclusion

A rolling contact fatigue model was proposed in this study for non-conformal point contacts under mixed EHL conditions where the metal-to-metal surface interactions are commonplace. A mixed-EHL model with improved accuracy and efficiency [14] was employed to predict the instantaneous surface contact pressure and shear traction distributions, from which the transient three-dimensional stress fields within the contacting component were evaluated by using the potential theory assuming purely elastic condition. The multi-axial fatigue lives of the material points were determined using the fatigue criterion that employed a characteristic plane approach [15]. In order to determine the surface roughness effects, a new Lagrangian–Eulerian approach was proposed to quantify the location of the crack nucleation site as well as the corresponding life distributions. The measurements of a companion rolling contact experimental study were compared to the predicted stress-life data and shown to support the proposed model well.

Acknowledgment

Authors thank General Motors Powertrain for sponsoring this research activity.

References


Klein M. An experimental investigation of the engineered surface treatments on gear pitting life. M.S. Thesis, The Ohio State University; 2009.