A physics-based model to predict micro-pitting lives of lubricated point contacts

Sheng Li *, Ahmet Kahraman

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

ARTICLE INFO

Article history:
Received 29 May 2012
Received in revised form 5 September 2012
Accepted 10 September 2012
Available online 21 September 2012

Keywords:
Micro-pitting
Life prediction
Surface roughness
Stress
BEM

ABSTRACT

This paper proposes a physics-based model to predict the onset of micro-pit formation for lubricated point contacts of rough surfaces. A mixed elasto-hydrodynamic lubrication model is employed to predict the surface normal and tangential tractions for a stress prediction model to determine the resultant stress histories. The boundary element approach is used in the stress model to fully capture the measured three-dimensional topographies of the contacting rough surfaces, allowing an accurate prediction of the localized stress concentrations that dictate the occurrence of micro-pits. In the process, the method of coordinate transformation and the indirect approach of rigid body motion are devised to eliminate the kernel singularities. A novel numerical procedure is also developed to minimize any Gaussian quadrature numerical error in the integration of the near singular kernels. The fatigue damage is then evaluated employing a multi-axial fatigue criterion. The proposed micro-pitting life prediction methodology is demonstrated using an example ball-on-disk contact problem.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Micro-pitting is a progressive rolling contact fatigue phenomenon that occurs on micro-scale primarily due to severe localized stress concentrations at or very near the contacting surfaces. Considering two rough surfaces in lubricated contact under combined rolling and sliding, the lubrication conditions are often mixed or boundary type, allowing the roughness asperities to contact each other and produce large instantaneous peaks of both normal pressure and surface shear. These time-varying surface tractions result in large multi-axial stress amplitudes for the material points within a very shallow layer of material (typically less than 2 μm in depth). The localized cyclic loading, which can be several times higher than the corresponding Hertzian pressure, dictates the fatigue failure in the form of micron size pits. These small pits can either be clustered together or scattered over the surface, depending on the surface roughness texture. As the number of contact cycles increases, the amount of micro-pits grows. To some extent, the scattered light reflections from the numerous micro-pits display grey-colored appearance, and has often been referred as grey staining or frosting. Under the condition that the two surfaces are well separated by a layer of fluid film, the hydrodynamic pressure and viscous shear can still be raised to relatively high levels when the surfaces are relatively rough, leading to the occurrence of micro-pitting, such as observed in Ref. [1]. This study aims at investigating the micro-pitting fatigue phenomenon observed in a lubricated point contact through a novel predictive model.

Although the link between the surface roughness and the resultant micro-pitting performance has been well established experimentally, published fatigue models for rough surface contacts commonly used the potential theory based half-space formulations, which assume perfectly smooth surfaces [2] to calculate the stress fields [3–7]. This smooth surface assumption is reasonable for sub-surface initiated macro-pitting (spalling) failures, while it is not suitable for surface or near surface initiated micro-pitting failures. These half-space displacement and stress formulations [2] have also been widely used in the rough surface contact models such as [8–10], assuming the difference of roughness height between any two points on the surface is far smaller than the distance between them [8]. This is often not the case for machine components such as gears whose surface roughness due to the finishing processes such as grinding and shaving can be rather significant. Even for highly polished surfaces, this assumption is not necessarily valid for the neighboring nodes on a very fine mesh grid. The contact model of Ref. [11] used the finite element (FE) approach to determine the apparent contact area and the mean contact pressure between two mating rough surfaces. Because the detailed description of the surface topography requires a very fine mesh grid, FE based models demand extremely large number of finite elements which lead to rather unaffordable computational time. A semi-detailed approach was employed in Ref. [11], excluding the predictions of the localized stress concentrations.

The alternative to the finite element method is the boundary element method (BEM) which considers only the boundary instead of the entire contact volume, effectively reducing the three-dimensional problem to a two-dimensional problem, such that a fine mesh is possible. Some BEM studies [12,13] focused on the
contacts between smooth components using finite boundary elements while other [14–17] considered the smooth half-space contact problem employing infinite boundary elements. A contact model that fully captures the surface topography variations of typical engineered rough surfaces is not available in the literature.

With the aim of bringing a fundamental understanding to the micro-pitting phenomenon commonly observed in gear and bearing components of transmissions and gearboxes in diverse applications ranging from wind turbine gearboxes to rotocraft and automotive transmissions, this paper proposes a new physics-based contact fatigue model for micro-pitting, incorporating the local roughness geometry in the near surface stress predictions. The boundary element approach is used in the process for the detailed description of the surface topography with an affordable computational effort. Since the contact zone is usually small for the mechanical components such as bearings and gears, an infinite half-space is assumed. The boundary of the computational domain is divided into a finite field (containing the contact zone) that is rough and an infinite field that is considered to be perfectly smooth. Finite triangular boundary elements are used to mesh the former and infinite quadrilateral boundary elements are used to discretize the latter. For the surface displacement computation, the integrations of the singular kernels are determined through the coordinate transformation and the indirect approach of rigid body motion. The stress distributions along the surfaces are calculated from the local strains through the Hooke’s law such that the singularities of the integral kernels can be circumvented. For the near surface stress fields, a method that combines the coordinate transformation and the progressive element subdivision is developed to minimize the numerical error caused by the singular behavior of the kernel functions when the load and the field points are too close to each other. The surface traction distribution inputs for the boundary element based stress prediction model are predicted by using a mixed elastohydrodynamic lubrication (EHL) model [18]. The stress histories yielded from the stress model are used to evaluate the fatigue damage according to the multi-axial fatigue criterion of Ref. [19].

2. Micro-pitting model

The methodology of predicting micro-pitting failures of lubricated rough contacts under combined rolling and sliding must include three major components as shown in the flowchart of Fig. 1. The first component is a deterministic mixed EHL model that computes the transient surface normal pressure \( p \) and shear \( q \) fluctuations induced by the local surface asperities. With these predicted surface traction distributions, the next step is to compute the multi-axial stress fields in a very shallow surface layer of material using a stress prediction model. This stress formulation should be able to include the influence of the local surface roughness geometry on the near surface stress concentrations, since the crack nucleation has been experimentally observed to be closely related to the surface roughness profiles. From the stress field prediction that is carried out for a sufficiently long time period, the alternating and mean values of the stress components must be determined for every material point that passes through the computational volume. Lastly, a suitable multi-axial fatigue criterion is needed in order to predict the micro-pitting fatigue life from the stress histories.

In this study, the mixed EHL model of Li and Kahraman [18], the newly devised boundary element based stress prediction model for rough surface point contacts and the high cycle, multi-axial fatigue criterion proposed by Liu and Mahadevan [19] are combined to model the micro-pitting fatigue behavior. The life of crack propagation from the nucleation stage to the micro-sized pit is assumed to be negligible in comparison to the initiation life [20]. For micro-pitting failure, any thermal effect caused by the frictional heating between the surfaces are neglected. The roughness profiles measured after the run-in of the surfaces are used in the simulations. The changes to these roughness profiles during contact are considered to be negligible. It is also assumed that the hardness within the very shallow layer of material of interest is constant. Additionally, the stress fields are considered to be purely elastic for the loading range employed in this study. For simplicity reasons, the residual stress is not included in the stress field prediction formulations while it can easily be incorporated in the methodology of Fig. 1.

2.1. Point contact mixed EHL model

For hydrodynamically lubricated rough surfaces contacting under combined rolling and sliding, the fluid film might break down intermittently within the contact due to the local roughness features and the unfavorable operating conditions such as relatively low rolling velocity, high sliding (which causes shear thinning), high temperature (which reduces the fluid viscosity) and heavy loading. The existence of these isolated boundary lubrication spots (asperity contacts) and the surrounding fluid regions complicates the numerical solution of the transient surface pressure and shear distributions. Hu and Zhu [21] and later Zhu [22] proposed a unified numerical approach to solve for the asperity contact pressure and hydrodynamic pressure simultaneously with great numerical robustness. To improve the convergence rate and reduce the computational time, Li and Kahraman [18] introduced the asymmetric integrated control volume discretization scheme, which was also shown to reduce the dependence of the solution accuracy on grid density. This dependence was shown to be a potential shortcoming of the models of [21,22] especially when the rolling speed is low [23]. A brief description of the mixed EHL formulations is provided here with other details in regards to discretization and numerical solution can be found in Ref. [18].

Within the fluid region of the contact, the hydrodynamic fluid flow is governed by the transient Reynolds equation of

\[
\frac{\partial}{\partial x} \left( \rho v_x \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left( \rho v_y \frac{\partial p}{\partial y} \right) = \frac{\partial (\rho v_x h)}{\partial x} + \frac{\partial (\rho h)}{\partial y} \tag{1a}
\]

where \( x \) and \( y \) point into the direction of rolling and the direction to the rolling direction, respectively, and \( \theta \) is the time. The parameters \( p, h \) and \( \rho \) in Eq. (1a) represent the pressure, thickness and density of the fluid, respectively, which are all functions of \( x, y \) and \( \theta \). \( v_x \) is the rolling velocity defined as \( v_x = (v_1 + v_2)/2 \) with \( v_1 \) and \( v_2 \) denoting the surface velocities (in the \( x \) direction designated

---

**Fig. 1.** Flowchart of the micro-pitting contact fatigue modeling methodology.
as the direction of rolling and sliding) of the contacting bodies 1 and 2. The non-Newtonian effects are modeled by assuming an Eyring fluid such that the flow coefficients \( f_x \) and \( f_y \) in the x and y directions, respectively, are given as [18]

\[
f_x = \frac{\rho h}{12\omega_0} \cosh \left( \frac{\tau_m}{\tau_0} \right) \quad f_y = \frac{\rho h}{12\omega_0} \sinh \left( \frac{\tau_m}{\tau_0} \right)
\]

where \( \omega \) is the lubricant viscosity, \( \tau_m/\tau_0 \) is the ratio of the lubricant reference stress to the viscous shear stress, determined according to \( \tau_m/\tau_0 = \sinh^{-1}[\nu_{p0}(\tau_m/\tau_0)] \), and \( v_t = v_1 - v_2 \) is the sliding velocity. It is noted that the Eyring model can be substituted by the more accurate Ree-Eyring model if the coefficients of the latter are available [24].

For the local spots where the mating asperities contact each other, a reduced form of Reynolds equation \([18,21,22]\) applies as

\[
\frac{\partial (\rho \psi h)}{\partial x} + \frac{\partial h}{\partial y} = 0
\]

Eqs. (1a) and (2) describe the mixed lubrication behavior of the contact covering the fluid regions and the asperity contact regions simultaneously. Here, a smooth transition between the hydrodynamic and boundary lubrication is assumed.

The local film thickness at the position of \((x, y)\) and time \(\psi\) is defined as

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

where \(h_0(\psi)\) is the reference film thickness, \(R_1(x, y, \psi)\) and \(g_0(x, y)\) are the roughness heights of the two surfaces at time \(\psi\), and \(g_0(x, y)\) is the geometric separation between the two surfaces before any elastic deformation occurs. Assuming perfectly smooth surfaces, the normal deflections of surface 1 and 2 caused by the tangential shear offset each other (considering the same material for both surfaces). With the same assumption, the total normal elastic deformation \(V(x, y, \psi)\) induced by the normal loading can be calculated using the Boussinesq’s half space formulation as [22]

\[
V(x, y, \psi) = \int \Theta((x - x', y - y')/p(x', y', \psi) \, dx' \, dy'
\]

where \(I\) is the computational domain of the contact zone, and \(\Theta(x, y)\) is the influence function. Recognizing that Eq. (4) is the convolution operation between \(p(x, y, \psi)\) and \(\Theta(x, y)\), the Discrete Fourier Transform (DFT) convolution technique is used to reduce the computation time significantly.

The predicted contact pressure distribution (both hydrodynamic and asperity contact) over the entire contact zone must balance the normal load \(W\) applied, i.e.

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

If this equation is not satisfied, the reference film thickness \(h_0(\psi)\) in Eq. (3) must be adjusted within a load iteration loop. The Roeland’s viscosity–pressure relationship \([25]\) and the Dowson–Higginson density–pressure relationship \([26]\) are used in this study to describe the dependency of viscosity and density on pressure. It should be noted here that this viscosity–pressure relationship might be less accurate within the high-pressure ranges. More accurate relationships such as the free-volume model \([27]\) or actual measurements (if available) can be used in its place without any loss of generality.

With the converged solutions of \(p(x, y, \psi)\) and \(h(x, y, \psi)\), the transient surface shear \(q(x, y, \psi)\) can be determined. Assuming that there is no slippage at the interface between the lubricant and the surfaces, and considering both the Poiseuille and Couette flows, the viscous shear stress acting on surface 1 for hydrodynamically lubricated areas is written as

\[
q(x, y, \psi) = \frac{1}{2} h(x, y, \psi) \frac{\partial p(x, y, \psi)}{\partial x} - \omega_s^* \frac{v_t}{h(x, y, \psi)}
\]

where \(\omega_s^* = \omega/cosh(\tau_m/\tau_0)\) is the effective viscosity for an Eyring fluid. For the asperity contact regions where the fluid film breaks down, the shear stress is defined simply as \(q(x, y, \psi) = \mu p(x, y, \psi)\) where \(\mu\) is the boundary friction coefficient between the mating asperities (assumed to be \(\mu = 0.15\) in this study). The details of the discretization scheme and the companion numerical approach can be found in Ref. [18].

2.2. Stress prediction model for rough surface contacts

For most gear and bearing contacts, the contact zone is usually small and a half-space is assumed. In order to implement the boundary element formulation, the boundary of the infinite half-space is divided into (i) the finite field \(\Gamma_x\), (ii) the infinite field \(\Gamma_I\) and (iii) the infinite semi-sphere \(\Gamma_S\) [15] as shown in Fig. 2. Along \(\Gamma_I\) that is located at infinite, the displacements are considered to be zero. The integral over \(\Gamma_S\) due to the local loading was also shown to be zero [14]. For \(\Gamma_I\), the zero traction condition is applied. With these and assuming homogeneous, isotropic and elastic material and absence of body forces, the displacements at the load point \(P\) are given in terms of the displacements and tractions at the field point \(Q\) through the boundary integral equation (BIE) [28]

\[
C_{ij}(P)U_i(P) + \int_{\Gamma_T} T_{ij}(P, Q)U_j(Q) \, d\Gamma = \int_{\Gamma_T} U_j(P, Q)T_{ij}(Q) \, d\Gamma
\]

where \(u\) is the displacement and \(t\) is the traction. The subscripts \(i, j = x, y, z\) indicate the directions. The integral kernel functions \(U_j(P, Q)\) and \(T_{ij}(P, Q)\) are the three-dimensional Kelvin fundamental solutions for \(u\) and \(t\) at point \(Q\) induced by a unit load exerted at point \(P\). With the distance between \(P\) and \(Q\) denoted by \(d\), \(U_i\) and \(T_{ij}\) are given as \([28]\)

\[
U_i = \frac{1}{16\pi G(1 - v)} \frac{1}{d} \left[ 3 - 4v \delta_{ij} + \frac{\partial \delta_{ij}}{\partial \theta} \right]
\]

\[
T_{ij} = \frac{1}{8\pi (1 - v) d^2} \left[ \frac{\partial \delta_{ij}}{\partial n_i} - \frac{\partial \delta_{ij}}{\partial n_j} \right]
\]

where \(G\) is the shear modulus, \(v\) is the Poisson’s ratio, \(n = [n_x, n_y, n_z]\) is the unit outward normal vector and \(\delta_{ij}\) is the Kronecker delta (\(\delta_{ij} = 1\) when \(i = j\) and \(\delta_{ij} = 0\) when \(i \neq j\)). It is seen that \(U_i\) and \(T_{ij}\) are of the order of \(1/d\) and \(1/d^2\), respectively, and hence, become singular as \(Q\) approaches \(P\) (i.e. \(d \rightarrow 0\)).

The finite field consists of the loading region defined by the bold rectangle, which circumscribes the Hertzian zone (shaded circle), and the surrounding transition region as shown in Fig. 3. Within \(\Gamma_x\), triangular boundary elements (TBE) are used. The \(z\) coordinates of the TBE nodes represent the roughness heights. In order to
capture the local surface roughness geometry, the mesh size within the loading region is set to be on the order of microns (about the same as the typical roughness measurement resolution). Under this fine mesh condition, isoparametric linear boundary elements are assumed to be sufficient for the accurate description of the contact. In the transition region, the surface displacements and tractions are much smaller. Accordingly, the mesh size of the TBE is set to be larger and gradually increase along the radial direction (going outward from the loading region). Isoparametric linear shape functions are also used in this region. Utilizing the linear shape functions of \( N_i = \xi, N_2 = \eta \) and \( N_4 = 1 - \xi - \eta \), the integrals over the finite region in Eq. (7a) are discretized as

\[
\int_{\Gamma_f} T_y u_i d\Gamma = \sum_{k_f} \sum_{\ell = 1}^{3} \sum_{\ell' = 1}^{3} T_{ij} u_{ij}^\ell + \sum_{k_f} \sum_{\ell = 1}^{3} u_{ij}^\ell T_{ij}^\ell
\] (8a)

\[
\int_{\Gamma_f} U_{ij} t_i d\Gamma = \sum_{k_f} \sum_{\ell = 1}^{3} t_{ij}^\ell u_{ij}^{\ell'} + \sum_{k_f} \sum_{\ell = 1}^{3} t_{ij}^{\ell'} u_{ij}^\ell
\] (8b)

where \( u_i^\ell \) and \( t_i^\ell \) represent the displacements and tractions at the node \( \ell \) of the element, \( u_{ij}^\ell = \int_{\Gamma_f} T_{ij} N_i d\Gamma \) and \( t_{ij}^\ell = \int_{\Gamma_f} U_{ij} N_j d\Gamma \) are the TBEs that are away from the load point \( P \) and the TBEs that are connected to \( P \), respectively. In addition, \( \ell' \) is the node where \( P \) and \( Q \) coincide and \( S \) is the area of the element.

For the numerical integrations of Eq. (8), Gaussian Quadrature Technique can be directly used for the integrals over the \( K_f \) TBEs (first terms on the right hand side of Eqs. (8a) and (8b)). When \( P \) and \( Q \) are in the same element (\( K_f \) TBEs), however, \( U_{ij} \) and \( T_{ij} \) exhibit singular behavior due to the \( 1/d \) and \( 1/d^2 \) terms as \( Q \) approaches \( P \). In order to eliminate such singularity, the coordinate transformation from the area coordinate system of \( (\xi_1, \xi_2) \) to the unit square coordinate system of \( (\eta_1, \eta_2) \) is applied, as illustrated in Fig. 4, introducing the transformation Jacobian of \( J_f^b = d/4PQ \).

Additionally, the shape function \( N_i \) for \( \ell \neq \ell' \) is of the order \( d \) [28], such that both \( T_y N_i J_f^b \) and \( U_{ij} N_j J_f^b \) are bounded for the TBEs of \( K_f \) when \( \ell \neq \ell' \). For the case of \( \ell = \ell' \), \( u_{ij} N_j J_f^b \) is still finite, while \( T_y N_i J_f^b \) is on the order of \( 1/d \). Hence, all the integrals in Eq. (8) with the exception of \( u_{ij}^\ell \) can be evaluated using Gaussian quadrature without any difficulty.

The infinite field \( \Gamma_\infty \) is discretized into isoparametric linear infinite quadrilateral boundary elements (IQBE) whose two nodes are connected with \( \Gamma_f \) (solid circles on the bound of \( \Gamma_f \) in Fig. 3) and the other two are positioned at infinity. For an arbitrary point \( A(\xi_1, \xi_2) \) within the IQBE as shown in Fig. 5a, its displacements can be related to the ones at point \( B(\xi_1, -1) \) which is on the bound of \( \Gamma_\infty \) by using a decay function \( D = CB/CA \) (\( C \) is the decay origin) as \( u_i(\xi_1, \xi_2) = Du_i(\xi_1, -1) \) [15]. The mid-side nodes 5 and 6 in Fig. 5a are selected such that \( C \) is positioned at the center of the contact zone. Writing \( u_i(\xi_1, -1) \) in terms of the displacements at nodes 3 and 4 of the IQBE as (Fig. 5a)

\[
u_i(\xi_1, -1) = M_i u_3^i + M_4 u_4^i,
\] (9a)

the displacements at point \( A \) can be further expressed in the form of

\[
u_i(\xi_1, \xi_2) = D(M_3 u_3^i + M_4 u_4^i)
\] (9b)

such that the discretization of the integral over the infinite region in Eq. (7a) reads

\[
\int_{\Gamma_f} T_y u_i d\Gamma = \sum_{k_f} \sum_{\ell = 3}^{4} u_{ij}^\ell \int_{S} T_y D M_j dS + \sum_{k_f} \sum_{\ell = 3}^{4} u_{ij}^\ell \int_{S} T_y M_j dS
\]

\[
+ \sum_{k_f} u_{ij}^{\ell'} \int_{S} T_y (D - 1) M_j dS + \sum_{k_f} u_{ij}^\ell \int_{S} T_y M_j dS
\] (10)

Fig. 3. Computational domain and boundary element mesh for infinite half space.

Fig. 4. Mapping from area coordinate system \((\xi_1, \xi_2)\) to unit square coordinate system \((\eta_1, \eta_2)\).

Fig. 5. Mapping from (a) infinite quadrilateral element to (b) unit square element.
where $K_r$ and $K_r'$, respectively, refer to the IQBEs that are away from the load point $P$ and the IQBEs that are connected to $P$, and $M_3$ and $M_4$ are the shape functions defined as $M_3 = (1 - \zeta_3)/2$ and $M_4 = (1 + \zeta_4)/2$.

In order to be able to use Gaussian quadrature for the infinite field, the infinite element is mapped into the unit square element as shown in Fig. 5 [15].

$$\zeta_1 = \eta_1, \quad \zeta_2 = \frac{1 + 3\eta_2}{1 - \eta_2} \quad (11a,b)$$

The Jacobian of this transformation is $J^T_{f} = 4/(1 - \eta_2)^2$. For the $K_r'$ IQBEs, the quadrilateral element (Fig. 5b) is further divided into two TBEs. For each of them, the transformation of Fig. 4 is applied to introduce the Jacobian that is on the order of $d$. When $\ell = \ell'$, the shape function $M_i$ is of the order $d$. On the other hand, when $\ell = \ell'$, the decay function $D \rightarrow 1$ and $(D - 1) \rightarrow 0$ as $Q$ approaches $P$, i.e. the term $(D - 1)$ is on the order of $d$ as well. As a result, the singularities in the second and third terms on the right hand side of Eq. (10) are eliminated. However, the singularity in the integral of the last term remains. Substituting Eqs. (8) and (10) into Eq. (7a) and moving all the undetermined terms to the left, one obtains

$$\sum_{k_r} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(M_r \cdot dS)u_{y_r}' = \sum_{k_r} \sum_{\ell = 1}^{3} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(D_{r}-1)M_{r} \cdot dS \quad (12)$$

To evaluate the integrals with singular kernels and the parameter $C_0$ in Eq. (12), the indirect approach of rigid body motion is used. Under rigid body motion, $u_i (i= x, y, z)$ is constant and $t_i = 0$ along the entire boundary. Applying this condition and $u_i = 1$, the boundary integral equation reads

$$C_0(P) + \int_{T_1 \cup T_2 \cup T_3} \frac{1}{1 - \eta_2} d\Gamma' = 0 \quad (13)$$

Similar to Eqs. (8) and (10), and applying the analytical solution of $I_{r, s} d\Gamma' = -\frac{1}{1 - \eta_2} \delta_0$ [14], Eq. (13) is discretized as

$$C_0 + \sum_{k_r} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(M_r \cdot dS)u_{y_r}' = \sum_{k_r} \sum_{\ell = 1}^{3} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(D_{r}-1)M_{r} \cdot dS + \frac{\delta_0 y}{2} \quad (14)$$

It is noted that there is no decay function involved in Eq. (14) since the displacements in all the directions are constant for rigid body motion. The integrals on the right hand side of Eq. (14) can be readily evaluated using Gaussian quadrature as described above. Substituting Eq. (14) into Eq. (12), Eq. (7a) is written in its final discretized form as

$$\sum_{k_r} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(M_r \cdot dS)u_{y_r}' = \sum_{k_r} \sum_{\ell = 1}^{3} \frac{3}{1 - \eta_2} \int_{S_r} T_{y_r}(D_{r}-1)M_{r} \cdot dS$$

$$= \int_{S_r} T_{y_r}(D_{r}-1)M_{r} \cdot dS \quad (15)$$

The displacement distributions along $\Gamma_r'$ can then be solved from Eq. (15), provided the boundary traction distributions are available. Different from Ref. [15], the proposed boundary element formulation does not require any additional integration of the kernel function $T_{y_r}$ over $\Gamma_r$. In Ref. [15], the integral of $\sum_{k_r} \int_{S_r} T_{y_r}M_{r} \cdot dS$ was mistakenly doubled count.

For any boundary or interior point, its stress state is determined by the displacement and traction distributions along the entire boundary as [28]

$$\sigma_{ij}(P) = \int_{\Gamma_r} E_{ij}(P, Q)\epsilon_{ij}(Q)d\Gamma - \int_{T_{r, s}} H_{ij}(P, Q)u_{ij}(Q)d\Gamma \quad (16a)$$

where $k = x, y, z$ and the third order kernel functions read [28]

$$E_{ij} = \frac{1}{8\pi(1 - \nu)d^2} \left[ (1 - 2\nu) \left( \frac{\partial \delta_k}{\partial \eta_l} \frac{\partial \delta_l}{\partial \xi_k} + \frac{\partial \delta_k}{\partial \xi_l} \frac{\partial \delta_l}{\partial \eta_k} \right) + \frac{3}{2} \frac{\partial \delta_l}{\partial \xi_k} \frac{\partial \delta_l}{\partial \eta_k} \right] \quad (16b)$$

$$H_{ij} = \frac{G}{4\pi(1 - \nu)d^2} \epsilon_{ij} \left[ (1 - 2\nu) \delta_k \frac{\partial \delta_l}{\partial \eta_k} + \delta_k \frac{\partial \delta_l}{\partial \xi_k} + \frac{3}{2} \frac{\partial \delta_l}{\partial \xi_k} \frac{\partial \delta_l}{\partial \eta_k} \right] \quad (16c)$$

These kernel functions are on the order of $1/d^2$ and $1/d^3$, and hence, become singular when $P$ and $Q$ coincide. To circumvent such singularities, the stress components for the boundary points are computed using the strains and tractions according to the Hooke’s law [28] in a local orthogonal coordinate system $(x', y', z')$ as shown in Fig. 6a. Here, the $z'$ axis points out of the surface. The area coordinate system $(\zeta_1, \zeta_2)$ is related to $(x', y')$ through

$$\zeta_1 = \frac{1}{AC} \left( x' - \frac{y'}{\tan \theta} \right), \quad \zeta_2 = \frac{y'}{BC \sin \theta} \quad (17a)$$

Denoting the direction cosines of $x'$, $y'$ and $z'$ in the global $(x, y, z)$ system as $(\alpha_x, \alpha_y, \alpha_z)$, $(\beta_x, \beta_y, \beta_z)$, and $(\gamma_x, \gamma_y, \gamma_z)$, respectively, and interpolating the displacements within an element using the linear shape function as $u_i = u_i(N_i)$, the local systems of $u_{x'} = \alpha_x (u_i N_i)$ and $u_{y'} = \beta_y (u_i N_i)$. The local strains of $\epsilon_{xx'} = \partial u_{x'}/\partial x'$, $\epsilon_{yy'} = \partial u_{y'}/\partial y'$ and $\epsilon_{zz'} = \partial u_{z'}/\partial z'$ are determined accordingly as
However, since micro-pitting cracks are surface nucleated and which are then transformed back into the global (Fig. 6). (a) Transformation from (x, y, z) to (ξ, η, ζ) for boundary stress computation and (b) subdivision of a triangular boundary element for interior stress computation.

$\varepsilon_x = a_u (\frac{\partial N_1}{\partial x_1} + \frac{\partial N_2}{\partial x_2})$,  $\varepsilon_y = \beta_u (\frac{\partial N_1}{\partial y_1} + \frac{\partial N_2}{\partial y_2})$

$\varepsilon_{xy} = a_u (\frac{\partial N_1}{\partial x_1} \frac{\partial N_2}{\partial y_2} + \frac{\partial N_2}{\partial x_2} \frac{\partial N_1}{\partial y_1}) + \beta_u (\frac{\partial N_1}{\partial y_1} \frac{\partial N_2}{\partial x_2} + \frac{\partial N_2}{\partial y_2} \frac{\partial N_1}{\partial x_1})$

Likewise, the tractions in the local (x, y, z) system are defined as $t_x = a_t t_x$, $t_y = \beta_t t_z$ and $t_z = \gamma_t t_y$. With these, the stress components are computed from the strains and tractions in the local system as [28]

$\sigma_x = \frac{E}{1-v^2} (\varepsilon_x + \nu \varepsilon_y)$,  $\sigma_y = \frac{E}{1-v^2} (\varepsilon_y + \nu \varepsilon_x)$

$\sigma_{xy} = \frac{E}{2(1+v)} \varepsilon_{xy}$,  $\sigma_{xz} = t_x$,  $\sigma_{yz} = t_y$,  $\sigma_{x'z'} = t_{x'}$  (19a–f)

which are then transformed back into the global (x, y, z) system.

Eq. (16a) can be directly used for the stress field computation of the interior points, which are always away from the field points. However, since micro-pitting cracks are surface nucleated and the surface layer of material of interest has a very shallow depth, any small distance between P and Q introduces numerical errors into the Gaussian quadrature process due to the near singular behavior of $E_{xy}$ and $H_{xy}$. In this work, a numerical approach that combines a coordinate transformation and an element subdivision is devised to minimize such error.

Similar to the treatment for $U_0$ and $T_0$, the transformation of Fig. 4 is first applied to introduce the Jacobian of the order of O(d), weakening the kernel singularity of $E_{xy}$ and $H_{xy}$ from O(1/d^2) and O(1/d^2) to O(1/d) and O(1/d^2), respectively. Noticing the gradients of 1/d and 1/d^2 increase sharply as d → 0, a progressive subdivision method is developed here to sub-mesh the element in an efficient manner. This method introduces the transformation Jacobian that goes to zero at a faster pace than d does when Q approaches P. As illustrated in Fig. 6b, the size of the sub-elements along the local $\eta_1$ direction (column), denoted as $T_{m}$ (m is the sub-element index in the $\eta_1$ direction and 1 ≤ m ≤ k) is progressively reduced when approaching the load point P. In the local $\eta_2$ direction (row), however, a uniform sub-element size of $T$ is used since the singular behavior does not vary in this direction. The subdivision procedure can be summarized in the following three steps. (i) Determine the size for the column of sub-elements that is farthest away from P in the $\eta_1$ direction, i.e., $T_1 = 2/\Omega$, where the integer $\Omega$ can be selected using the length of the longest side of the TBE as the element size according to Ref. [29]. (ii) Create a progressive mesh in the $\eta_1$ direction according to $T_m = (1 - 1/\Omega)T_{m-1}$ (m > 1) such that $T_{m1}T_{m-11} = T_{m2}T_{m-12}$ where $T_{m1}$ and $T_{m2}$ are the sub-element sizes in the global coordinate system. (iii) Generate $\Omega$ rows of sub-elements in the $\eta_2$ direction. This progressive subdivision introduces the transformation of

$$\eta_1 = \frac{1}{2} (\eta_1^w + \eta_1^i) + \frac{m}{2} \eta_2, \quad \eta_2 = \frac{1}{2} (\eta_2^w + \eta_2^i) + \frac{L}{2} \eta_2$$

(20a, b)

where $\eta_1^w$ and $\eta_1^i$ are the $\eta_1$ coordinates of the west and east bounds of the sub-element, and $\eta_2^w$ and $\eta_2^i$ are the $\eta_2$ coordinates of the south and north bounds of the sub-element. The corresponding transformation Jacobian is $J = 1 - 1/\Omega G$. It is noted that $T_0 = 0$ as m → k such that the singular orders of the kernel functions are reduced further. Additionally, the sub-element that is closer to the load point P has smaller area, leading to higher density of Gaussian quadrature points and better accuracy. Eq. (16a) is discretized in a similar way as that of Eq. (7a).

The accuracy of the boundary element based surface displacement and stress prediction formulation presented above was assessed in a companion study [30] by comparing the BEM solutions with the closed-form solutions for Hertzian contacts. The maximum relative errors were found to be less than 2% for the surface displacements and less than 0.7% for the stress components. Furthermore, the maximum relative errors for the surface displacements were shown to occur where the displacement magnitudes approached zero and the absolute error magnitudes were observed to be negligibly small.

### 2.3. Multi-axial contact fatigue model

Various stress-based critical plane criteria are available for high-cycle multi-axial fatigue analysis [31–35]. These criteria evaluate the fatigue damage on a pre-defined critical plane. Some of them defined the critical plane as the plane of maximum shear stress amplitude [31–33]. Others included the normal stress component [34] or hydrostatic stress [35] in the definition of the critical plane. These different definitions are usually arrived from the experiments or the rolling contact fatigue damage to show that different criteria yielded similar results. The maximum damage was found to occur at the depth about 10 μm that is below the micro-pitting nucleation depths of less than 2 μm. Li and Kahraman [4] applied the critical plane criterion of Susmel and Lazzarini [33] to gear contacts under various loading and surface roughness condition to conclude that the predicted crack nucleation sites were much deeper than the experimental observations.

Another multi-axial fatigue criterion, the characteristic plane approach, was proposed by Liu and Mahadevan [19] for railroad wheel contacts. This approach assumes that the fatigue assessment
can be approximated by using a certain stress combination (damage parameter) on a certain plane, which is not necessarily the macro crack plane (fatigue fracture plane). The fatigue criterion is defined on this characteristic plane as [19]

\[
\frac{1}{\chi (1 - \frac{\sigma_{\text{m, max}}}{\sigma_{\text{m, ref}}})} \left[ \sigma_a^2 + \left( \frac{\sigma_a}{\sigma_{\text{m, ref}}} \right)^2 \right]^{\frac{1}{2}} = S_a
\]

(21)

where \(S_a\) and \(S_a^*\) are the uni-axial fully-reversed bending fatigue strength and the fully-reversed torsion fatigue strength corresponding to a finite fatigue life cycle of \(N\), \(\sigma_a\), \(\tau_{a1}\) and \(\tau_{a2}\) are the normal and shear stress amplitudes acting on the characteristic plane, and \(\chi\) is a material property related parameter. The term \((1 - \sigma_{\text{m, max}}/\sigma_{\text{m, ref}})\) in the denominator is incorporated to include the mean normal stress effect. Here \(\sigma_{\text{m, max}}\) is the mean normal stress on the macro crack plane (assumed as the plane with the maximum normal stress amplitude) and \(\sigma_{\text{m, ref}}\) is a reference stress that can be calibrated by using the uni-axial bending fatigue data with non-zero mean normal stress. In the absence of such material data, \(\sigma_{\text{m, ref}}\) can be approximated using the ultimate tensile strength or yield strength.

In order to determine the orientation of the characteristic plane, the macro crack plane which is assumed to be the plane experiencing the maximum normal stress amplitude is used as a reference plane. Defining the characteristic plane to be the one on which the damage introduced by the hydrostatic stress amplitude is minimum, the angle between the characteristic plane and the macro crack plane \(\theta\) and the material parameter \(\chi\) can be derived by applying Eq. (21) to the fully reversed bending and torsion fatigue problems as

\[
\theta = \frac{1}{2} \cos^{-1} \left[ \frac{s^2 - \sqrt{s^4 - (3s^2 - 1)/(4s^4 - 5s^2 + 1)}}{4s^4 - 5s^2 + 1} \right]
\]

(22a)

\[
\chi = \sqrt{s^2 \cos^2(2\theta) + \sin^2(2\theta)}
\]

(22b)

for non-extremely brittle materials with \(s = S_a^*/S_a < 1\). This multiaxial fatigue criterion was applied to predict the macro-pitting crack nucleation lives of point [3] and gear contacts [4], with both...
cases showing good correlation to experiments. For this reason, the same multi-axial fatigue criterion is applied to the micro-pitting crack initiation life prediction problem in hand. However, the methodology outlined in this work is general such that any other multi-axial fatigue criterion can be used in place of this characteristic plane criterion, if deemed appropriate.

3. A simulation example: a ball-on-disk contact

A ball-on disk contact problem is considered here as an example. The same contact was used in another recent study by these authors [36] on the prediction of scuffing failures of lubricated point contacts. In that study, a thermal version of the EHL model
described in Section 2.1 was used to predict the surface traction which was shown to agree well with the measurements. Both the ball and the disk are made of case hardened (surface hardness of 60 HRC) AISI 8620 steel. The contact between them operating under the Hertzian pressure of 1.12 GPa, Hertzian radius of 0.15 mm, rolling velocity of \(v_l = 5\) m/s and slide-to-roll ratio of \(SR = v_l/v_r = -0.25\) is simulated using the proposed model for the micro-pitting fatigue life prediction. The mesh size of the loading region (bold rectangle, which circumscribes the shaded Hertzian zone in Fig. 3) is set as \(\Delta x = \Delta y = 3\) \(\mu\)m which corresponds to that of the resolution of the three-dimensional roughness measurement as shown in Fig. 7. The time increment of \(\Delta t = \Delta x/v_l\) is used with a total of 550 time steps to cover about 1.5 mm travel distance. The synthetic fluid of Optigear Synthetic A320 is used as the lubricant whose inlet temperature is controlled at 95°C. At this temperature, the ambient viscosity and pressure-viscosity coefficient of this oil is 0.0276 Pas and 14.3 GPa respectively.

Fig. 8 shows the instantaneous surface normal and tangential traction (in the \(x\) direction that is the direction of rolling and sliding) distributions together with the fluid film thickness distribution within the Hertzian zone for two example time instants of \(n_s = 232\) and 528, as predicted by the mixed EHL model of Section 2.1. Multiple isolated islands where the film thickness breaks down are observed in the Hertzian zone. At these localized spots, the surface asperities interact and both the normal and tangential traction values become significantly larger. For the same boundary traction conditions, the surface is predicted to deform as shown in Fig. 9 by the boundary element formulation of Section 2.2, which includes the full description of the local roughness geometry. Comparing with Fig. 8, multiple concentrated deformations are seen in Fig. 9 to distribute correspondingly to the roughness interactions.

With both the traction distributions and the corresponding displacements in hand, the multi-axial stress fields within a very shallow layer of material (up to a depth of 7 \(\mu\)m) are computed for every time instant. For the same time instants as those of Figs. 8 and 9, Fig. 10 shows the instantaneous positions of the surface roughness profiles relative to each other together with the stress distributions for the normal components \((\sigma_n, \sigma_r, \sigma_z)\) and the orthogonal shear component \((\sigma_{xy})\) on the central vertical plane of \(y = 0\) (i.e. the \(xz\) plane). The \(\sigma_{xy}\) and \(\sigma_{xz}\) components are not shown in Fig. 10 since they are negligibly small on \(y = 0\). Localized severe stress concentrations are evident in Fig. 10 at locations where surface asperities interact. Although the maximum normal stresses take place on the surface, the orthogonal shear stress has the large amplitude positioned below the surface. In Fig. 11, the distributions for all the six stress components at the depth of 2 \(\mu\)m (typical micro-pit depth) are plotted for \(n_s = 528\). Although the magnitude of \(\sigma_{xz}\) is small on \(y = 0\), it is found to be comparable to \(\sigma_{xy}\) away from \(y = 0\) while \(\sigma_{xy}\) is seen to be relatively small compared to the other two shear components.

From the predicted stress fields of all six stress components within a time period, the histories of the stresses for every material point passing through the computational domain are found and used in the multi-axial fatigue criterion of Eq. (21) to predict the micro-pitting fatigue life. The fatigue fracture plane is assumed to be the plane with the maximum normal stress amplitude and is searched in the three-dimensional space with a \(2^\circ\) angle increment. With the orthogonal frame system of \(x'-y'-z'\) defined such...
that \( z \) is normal to the fracture plane, \( x' \) and \( y' \) are within the fracture plane with \( y' \) points into the direction of the maximum shear stress amplitude, the characteristic plane is reached by rotating the angle \( \theta \) from the fracture plane about the \( x' \) axis. The fatigue damage is then evaluated on this plane according to the fully reversed pure bending and pure torsion fatigue strength of the steel considered [3,19].

Fig. 12a shows the predicted micro-pit crack initiation life distribution along the central vertical plane of \( y = 0 \). Fig. 12b shows the same along the top surface layer of the contact. The fatigue lives of the material points are seen to range between \( 10^8 \) and \( 10^{14} \) (no failure) cycles. The crack nucleation sites are observed to be on the surface along the asperity peaks, which agree with the experimental observations [30]. From the distribution of Fig. 12b, the localized micro-pitted areas are determined and plotted in Fig. 13 at the contact loading cycles of 1 million, 10 million and 100 million (i.e. all the material points with the lives less than or equal to 1 million, 10 million and 100 million, respectively). Progression and spreading of the micro-pits with the contact cycles is evident in Fig. 13.

### 4. Conclusions

In this paper, a physics-based model was proposed to predict the micro-pitting life of lubricated point contacts of rough surfaces. A mixed EHL model was employed to predict the surface normal and tangential tractions. A new boundary element based rough surface stress prediction model was used to compute the multiaxial stress fields, capturing the influences of the surface roughness topography on the localized stress concentrations. The method of coordinate transformation and the indirect approach of rigid body motion were devised to eliminate the kernel singularities encountered in the surface deflection computation. A novel numerical procedure was developed to minimize any error in the numerical integration of the near singular kernels for the prediction of interior stress fields. A multi-axial fatigue criterion was implemented to evaluate the micro-pitting fatigue lives. In the end, the proposed micro-pitting life prediction method was applied to an example ball-on-disk contact problem with the measured rough surfaces to demonstrate the capabilities of the model.

Our current work focuses on the validation of the methodology presented above. For this purpose, a two-disk test machine is used to perform a number of micro-pitting experiments according to a test matrix constructed using the Fractional Factorial technique, including the potential factors of contact pressure, rolling velocity, slide-to-roll ratio and surface roughness [30]. The direct comparisons between the predictions of the proposed model and these experiments will be presented in a companion paper.

### Acknowledgements

This study was sponsored by Department of Energy (DOE) EERE – Wind & Water Power Program (Grant Number: DE-EE0002735).

### References