Engineering Fracture Mechanics 76 (2009) 1320–1335

Contents lists available at ScienceDirect

Engineering Fracture Mechanics

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

Fatigue crack initiation and propagation under cyclic contact loading

G. Fajdiga [a], M. Sraml [b,*]

[a] University of Ljubljana, Faculty of Mechanical Engineering, Askerceva 6, 1000 Ljubljana, Slovenia
[b] University of Maribor, Faculty of Mechanical Engineering, Smetanova ul. 17, 2000 Maribor, Slovenia

ARTICLE INFO

Article history:
Received 10 January 2008
Received in revised form 23 December 2008
Accepted 8 February 2009
Available online 20 February 2009

Keywords:
Contact fatigue
Crack initiation
Crack propagation
Fracture mechanics
Numerical modelling
Gear pitting

ABSTRACT

A computational model for contact fatigue damage analysis of gear teeth flanks is presented in this paper. The model considers the conditions required for the surface fatigue crack initiation and then allows for proper simulation of the fatigue crack propagation that leads to the appearance of small pits on the contact surface. The fatigue process leading to pitting is divided into crack initiation and a crack propagation period.

The model for prediction of identification of critical material areas and the number of loading cycles, required for the initial fatigue crack to appear, is based on Coffin–Manson relations between deformations and loading cycles, and comprises characteristic material fatigue parameters. The computational approach is based on continuum mechanics, where a homogenous and elastic material model is assumed and results of cyclic loading conditions are obtained using the finite element method analysis.

The short crack theory together with the finite element method is then used for simulation of the fatigue crack growth. The virtual crack extension (VCE) method, implemented in the finite element method, is used for simulating the fatigue crack growth from the initial crack up to the formation of the surface pit. The relationship between the stress intensity factor K and crack length a, which is needed for determination of the required number of loading cycles N_p for a crack propagation from the initial to the critical length, is shown.

1. Introduction

Mechanical behaviour of various machine elements, such as gears, brakes, clutches, rolling bearings, wheels, rails, and screw and riveted joints, are influenced by interaction between contact elements and surfaces. Surfaces in rolling and/or sliding contact are exposed to material contact fatigue. Contact fatigue can be defined as a kind of damage caused by changes in the material microstructure which results in crack initiation followed by crack propagation, under the influence of time-dependent rolling and/or sliding contact loads. The contact fatigue process can in general be divided into two main parts:

- initiation of micro-cracks due to local accumulation of dislocations, high stresses at local points, plastic deformation around inhomogeneous inclusions or other imperfections in or under the contact surface;
- crack propagation, which causes permanent damage to a mechanical element.

In this paper the pitting phenomenon of gears is addressed and the developed numerical model is used for determination of pitting resistance, i.e. the service life of gear teeth flanks. The initiation of fatigue cracks represents one of the most important stages in the pitting process. The position and mode of fatigue crack initiation depends on the microstructure of the material, the type of the applied stress and micro- and macro-geometry of the specimen [1]. The crack initiation periods

* Corresponding author. Tel.: +386 2 2207722; fax: +386 2 2207729.
E-mail address: saml.matjaz@uni-mb.si (M. Sraml).

0013-7944/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engfracmech.2009.02.005

Nomenclature

a	crack length
da	crack extension
$A_{fracture}$	cross-section of the specimen at fracture
A_0	initial cross-section of the specimen
b	strength exponent
b	half-width of the contact area
c	fatigue ductility exponent
C_0	material constant
C	material constant
D	ductility
D	grain size
E	Young's modulus
E^*	equivalent Young's modulus
F_N	normal force per unit length
G	shear modulus
G	energy release rate
K	stress intensity factor
$K_{I,II,III}$	stress intensity factor due to different crack opening
K'	material strength coefficient
K_{th}	threshold stress intensity factor
ΔK	stress intensity range
m	material constant
m_0	material constant
n	describes the relative position of the crack tip to the grain boundary
n'	material hardening exponent
N	stress cycles required for pitting
N_i	stress cycles required for the appearance of the initial crack
N_p	stress cycles required for a crack to propagate from the initial to the critical crack length
N_f	number of load cycles to failure
$p(x)$	normal loading
p_0	maximum contact pressure
$q(x)$	tangential loading
R	curvature radius
R^*	equivalent curvature radius
dV	released strain energy
z	number of grains transverse by the crack
α_0	initial inclination angle towards the contact surface
ε'_f	fatigue ductility coefficient
ε_a	strain amplitude
$\varepsilon_{fracture}$	specific deformation of the specimen at fracture
$\Delta\varepsilon$	specific deformation increment
$\Delta\varepsilon_p$	incremental value of plastic strain
$\Delta\varepsilon_e$	incremental value of elastic strain
$\Delta\varepsilon_1$	amplitude of the maximum principal deformation
Δ_{pl}	crack tip plastic displacement
μ	coefficient of friction
ν	Poisson ratio
σ_a	stress amplitude
σ'_f	fatigue strength coefficient
σ_m	mean stress
σ_{UTS}	ultimate tensile stress
σ_1^{max}	maximum value of principal stresses in the direction of maximum principal deformation
σ_y	yield stress
$\Delta\sigma$	incremental values of stress

can be very different and cracks can be initiated either on or under the surface, depending on a different combination of rolling and sliding contact conditions [2].

In general, gear teeth pitting may be initiated on sub-surface or surface. Sub-surface pitting initiation can be observed in high quality gears made of alloy steel, with smooth contact surfaces and good lubrication, where the large shearing sub-sur-

face stresses due to contact loading initiate substantial dislocation motion, which governs the crack initiation process [3]. Surface pitting initiation is common in industrial gears, which have rougher surfaces and are made of ordinary construction steels. The surface pitting is strongly influenced by surface roughness and other surface defects, like machining marks, large notches and inclusions. The surface cracks may also appear as a consequence of thermal treatment of the material due to residual stresses.

This paper considers only the second mechanism of gear pitting, i.e. surface pitting. The process of surface pitting can be visualised as the formations of small surface initial cracks grow under repeated contact loading. Eventually, the crack becomes large enough for unstable growth to occur, which causes the material surface layer to break away. The resulting void is a surface pit (Fig. 1).

The number of stress cycles N required for pitting of a gear teeth flank to occur can be determined from the number of stress cycles N_i required for the appearance of the initial crack in the material and the number of stress cycles N_p required for a crack to propagate from the initial to the critical crack length, when the final failure can be expected to occur:

$$N = N_i + N_p \tag{1}$$

This paper describes a computational model for contact fatigue crack initiation and crack propagation in the contact area of gear teeth flanks. The purpose of the present study is to present, firstly, a model for prediction of contact fatigue initiation, which is based on continuum mechanics, real cyclic contact loading and specific material fatigue parameters. The material model is assumed as homogeneous, without imperfections such as inclusions, asperities, roughness and residual stresses, as often occur in mechanical elements. A moving contact load is often used for simulation of the cyclic loading in fatigue crack initiation and propagation analyses on mechanical elements (simulation of meshing of gears) [4]. The second part of the computational model is a crack propagation model based on appropriate short fatigue crack growth theories [5]. The model attempts to account for the different parameters influencing the crack propagation process (Hertzian contact pressure, friction between contacting surfaces, fluid trapped in the crack, meshing of gears, etc.) leading to pitting, starting from the initial surface fatigue crack to the critical crack length, when the occurrence of a surface pit is expected. The results of the computations can be used to predict the time required for the development of pits, i.e. the service life of gears with regard to pitting [4].

Based on the computational results and considering some particular material parameters, the service life N of gear flanks can then be determined.

2. Simulation of contact problems in gears

Initial surface cracks leading to pitting of gears are generally observed to appear in the contact area where high normal contact pressure is combined with significant sliding velocity, which results in additional frictional loading of the surface layer. The most critical contact loading conditions for initial crack formation and propagation are identified for rolling contact with sliding, where contact sliding, and with that the effect of friction, is opposite to the direction of the rolling contact motion [6]. The worst contact loading conditions appear when the gears teeth are in contact at the inner point of single teeth pair engagement (point B, Fig. 2), where the surface-breaking initial cracks are expected to develop first. This is also the case considered in all computations reported in this study.

2.1. Equivalent contact model

For computational simulations of surface pitting it is advantageous to use an equivalent model of two contacting cylinders instead of simulating the actual contact of mechanical elements, Fig. 2. Using the equivalent contact model of two cylinders, the contact stresses can be determined by utilising the Hertzian contact theory [7]. The equivalent cylinders have the same radii as the curvature radii of the treated gear pairs at the point of the actual contact.

According to the Hertzian theory, the distribution of normal contact pressure in the contact area of two cylinders can be analytically determined by [7]

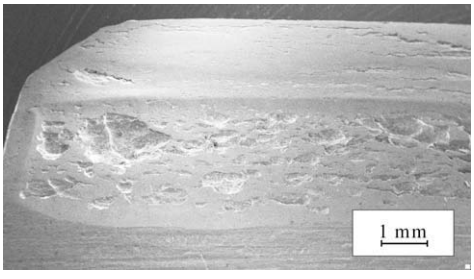

Fig. 1. Typical surface pits on gear tooth flank.

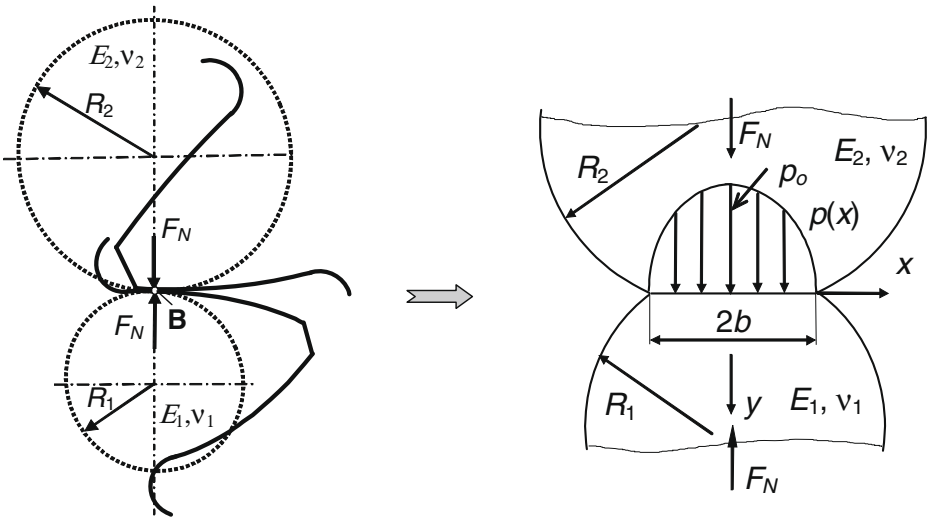

Fig. 2. Equivalent contact model of two cylinders.

$$p(x) = \frac{2F_N}{\pi b^2}\sqrt{b^2 - x^2}, \tag{2}$$

in which F_N is the normal force per unit width of equivalent cylinders and b is the half-length of the contact area, which can be determined from [7]

$$b = \sqrt{\frac{8F_N R^*}{\pi E^*}}, \tag{3}$$

where R^* and E^* are the equivalent radius and the equivalent Young's modulus, respectively, defined as [7]

$$R^* = \frac{R_1 \cdot R_2}{R_1 + R_2}, \tag{4}$$

$$E^* = \frac{2E_1E_2}{E_2(1 - v_1^2) + E_1(1 - v_2^2)}, \tag{5}$$

where R_1, E_1, v_1 and R_2, E_2, v_2 are the curvature radii, Young's modulus and Poisons ratio of contacting cylinders, see Fig. 2. The maximum contact pressure $p_0 = p(x = 0)$ can then be determined as [7]

$$p_0 = \sqrt{\frac{F_N \cdot E^*}{2\pi R^*}}. \tag{6}$$

The distribution of frictional contact loading $q(x)$ due to the relative sliding of the contacting cylinders is here determined by utilising the Coulomb friction law [7]

$$q(x) = \mu p(x), \tag{7}$$

where μ is the coefficient of friction between contacting cylinders.

For the general case of elastic contact between two deformable bodies in a standing contact situation, the analytical solutions are well known. However, using general Hertzian equations [7], it is hard to provide the loading cycle history and/or simulation of a contact pressure distribution of the moving contact in the analytical manner. Therefore, the finite element method is used for simulating two-dimensional friction contact loading in this case and the same procedure is usually used when dealing with complex contact loading conditions (e.g. in case of gears analysis).

The equivalent contact model is spatially discretised in the region of interest, where a finer mesh is used around material points (x_i, y_i) on and under the contact region. The computational model for evaluating contact stresses is a two-dimensional rectangle, with assumed plain strain conditions (Fig. 3).

3. Fatigue crack initiation

The initiation of fatigue cracks represents one of the most important stages in the fatigue process. The position and mode of the fatigue crack initiation depends on the microstructure of a material, the type of the applied stress and micro- and macro-geometry of the specimen.

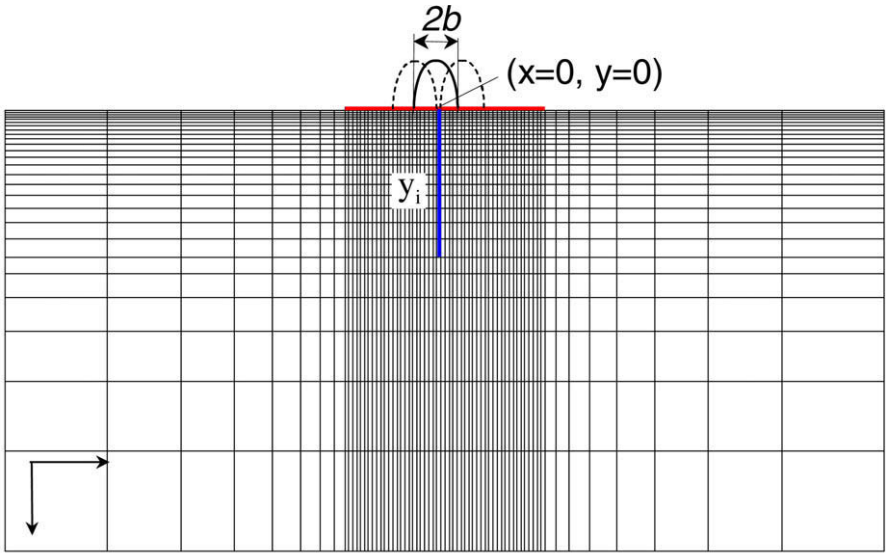

Fig. 3. Finite element model for determination of contact loading cycles.

The model for the fatigue crack initiation presented in this study is based on the continuum mechanics approach, where it is assumed that the material is homogeneous and isotropic, i.e. without imperfections or damage. A comprehensive model for the contact fatigue life prediction of mechanical elements should consider the time history of applied contact loads, regarding their range of variation. The rolling–sliding contact loads, typical of mechanical elements such as gears, wheels and rails and rolling bearings, are generally stochastic in a certain range. For the description of a general case of contact loading, one has to estimate the average normal and tangential contact forces for computational determination of surface and subsurface contact stresses. Normal contact forces can be appropriately determined by the Hertzian contact theory [7], while the tangential frictional forces can be considered using the Coulomb friction law (Eq. (7)) (see Chapter 2.1 in this paper). By applying these contact loading conditions in a moving fashion along the contact area of the computational model (simulation of gear meshing), one can computationally determine the stress loading cycle for each observed material point (x_i,y_i) on or under the contact surface (Fig. 3). That way, a realistic description of a stress cyclic loading in the time domain due to rolling and/or sliding contact conditions is achieved. A finite or boundary element method can be used for this purpose [8]. In this study, the FE analysis program MSC/Nastran has been used for the computational estimation of the loading cycles [9].

The complete procedure for computational determination of the material point stress loading cycle can be described as follows in the next subchapters.

3.1. Determination of conditions for fatigue crack initiation analysis

The stress analysis of the generalised contact model is performed in the framework of the finite element method. An appropriate stress state for each observed material point is computed for each position of the moving contact loading (Fig. 3). This procedure results in the generation of real stress loading cycles of material points in one pass of the rolling–sliding contact, which are necessary for the following contact fatigue analysis.

When the stress loading cycles are determined, the fatigue analysis for each observed material point could be performed. The methods for fatigue analysis are most frequently based on relation between deformations, stresses and number of loading cycles and are usually modified to fit the nature of the stress cycle, e.g. a repeated or reversed stress cycle [10–14]. The number of stress cycles required for a fatigue crack to appear can be determined iteratively with the strain-life method ε–N, where the relationship between the specific deformation increment $\Delta\varepsilon$ and the number of loading cycles N_f is fully characterised by the following equation [11–14]:

$$\frac{\Delta\varepsilon}{2}=\frac{\sigma_a}{E}+\frac{\Delta\varepsilon_p}{2}=\frac{\sigma'_f}{E}(2N_f)^b+\varepsilon'_f(2N_f)^c, \tag{8}$$

where E is the elastic modulus, σ'_f is the fatigue strength coefficient, b is the strength exponent, ε'_f is the fatigue ductility coefficient and c is the fatigue ductility exponent. Material parameters σ'_f, b, ε'_f, and c are experimentally determined. According to Fig. 4 it can be seen that σ'_f is determined as the stress where elastic line crossing the ordinate axis. Exponent b is determining the slope of elastic component line. The fatigue ductility coefficient can be determined as strain value at intersection of plastic component line and ordinate axis and the c as the slope of the plastic strain component line. Generally, the following modified approaches of the strain-life method are most often used for fatigue calculations: Coffin–Manson's

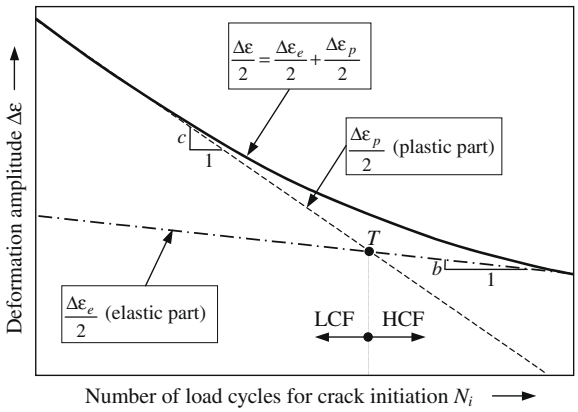

Fig. 4. Strain-life method for the fatigue crack initiation.

hypothesis (ε–N method), Morrow's analysis, and Smith–Watson–Topper (SWT) method [11–14]. According to Morrow, the relationship between strain amplitude, ε_a, and the pertinent number of load cycles to failure, N_f, can be written as [11]

$$\varepsilon_a = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c, \tag{9}$$

where E is the elastic modulus, σ_f' is the fatigue strength coefficient, ε_f' is the fatigue ductility coefficient, b is the exponent of strength and c is the fatigue ductility exponent, respectively.

Morrow's equation with mean stress, σ_m, correction (with a static mean stress not equal to 0) [11] is

$$\varepsilon_a = \frac{(\sigma_f' - \sigma_m)}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c. \tag{10}$$

According to Coffin–Manson [12–14], this relationship can be simplified as

$$\varepsilon_a = 1.75\frac{\sigma_{UTS}}{E}N_f^{-0.12} + 0.5D^{0.6}N_f^{-0.6}, \tag{11}$$

where σ_{UTS} is the ultimate tensile stress and D is the ductility, defined as

$$D = \ln\frac{A_0}{A_{fracture}} \cong \varepsilon_{fracture}. \tag{12}$$

The Smith–Watson–Topper method [10] considers the influence of the mean stress value and can be defined as

$$\sigma_a\frac{\Delta\varepsilon_1}{2} = \sigma_f'\varepsilon_f'(2N_f)^{b+c} + \frac{\sigma_f'^2}{E}(2N_f)^{2b}, \tag{13}$$

where $\Delta\varepsilon_1$ is the amplitude of the maximum principal deformation and $\sigma_1^{\max}$ is the maximum value of principal stresses in the direction of maximum principal deformation.

All the described fatigue life models used in this work are actually based on the strain-life method. Once a local stress or strain-time history has been established, a fatigue analysis method must be applied. The strain-life approach can be used to advantage over the S–N method, even in high cycle applications, due to its less scatter-prone materials data. Assumption that have been made in fatigue life models are that the models are uni-axial, which means that just one principal stress occurs in one direction (single stress vector). However, for more complex (multiaxial) loading conditions, multiaxial critical plane analysis should be applied. Before damage can be determined and summed for each cycle, certain corrections need to take place, the main correction being the conversion of purely elastic stresses and strains to elasto-plastic stresses and strains. In this work, plasticity is accounted for in the crack initiation method by the Neuber method [15]. The elastic stresses and strains are looked up on the elastic line and then corrected to fall onto the cyclic stress strain curve to determine the elastic–plastic stresses and strains. It is then this elastic–plastic strain that is used to look up damage on the strain-life damage curve. Neuber's elastic–plastic correction is based on the simple principle that the product of the elastic stress and strain should be equal to the product of the elastic–plastic stress and strain from the cyclic stress–strain curve (Eq. (14)) [15]

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma}{2E} + \left[\frac{\Delta\sigma}{2K'}\right]^{1/n'}, \tag{14}$$
$$\Delta\sigma\cdot\Delta\varepsilon = E\Delta\varepsilon_e^2$$

where $\Delta\varepsilon$, $\Delta\sigma$ are incremental values of strain and stress, E is the elastic modulus, $\Delta\varepsilon_e$ is the incremental value of elastic strain, n' is the material hardening exponent and K' is the material strength coefficient. Those parameters are obtained using experiments. Through an iterative method, the elastic–plastic stress and strain can be determined.

The material curve (Fig. 4) can be fully characterised by knowing the previously described material parameters σ'_f, b, ε'_f, c as shown in Eq. (13) and depends on which method of analysis is used. This curve is divided into an elastic component and a plastic component, which can also be plotted separately.

The transition point T (Fig. 4) presents the limit between the high cycle fatigue (HCF) and low cycle fatigue (LCF) regimes. Generally, there is no irreversible deformation in the HCF regime at the macroscopic level, while the LCF regime implies significant macroscopic deformation including irreversible deformation at this level [12–14]. However, the first micro-cracks in the persistent slip bands (PSB) appear quite early in the life of the mechanical component, regardless on the fatigue regime [10]. The strain and the plastic strain are no longer related to the stress through a simple relation. Although, the strain-life method ε–N is not an ideal model for fatigue damage initiation analysis at the microstructural level, since the micro-crack initiation in crystal grains and dislocation theory are not taken into account, it is most often model used for this purpose. In the work of Suresh [16] and Bhattacharya and Ellingwood [5] has been established that fatigue damage initiation is represented by applying of a certain number of loading cycles when the first fatigue damage occurs, on the basis of the assumed initial homogenous state of material. Thus, the ε–N method represents a very useful method for determining the initial fatigue damage, which is located at the material point with the largest stresses, where fatigue damage initiation is most likely to occur in the time domain.

4. Fatigue crack propagation

By considering initiation of small crack lengths in the contact area of the meshing gears, the short crack theory can be used for describing crack propagation from the initial to the critical crack length, when the pits occur on the surface. The short crack growth is characterised by a successive blocking of persistent slip bands (PSB) with grain boundaries, which implies the discontinuous character of the crack growth process. Navarro and Rios [18] and Sun et al. [19] proposed the model where the crack growth rate $\mathrm{d}a/\mathrm{d}N$ is assumed to be proportional to the crack tip plastic displacement δ_{pl}

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C_0(\delta_{pl})^{m_0}, \tag{15}$$

where C_0 and m_0 are material constants that are determined experimentally [20]. It is known that short cracks do not behave in accordance with LEFM. However, in view of the numerical simulation, it is beneficial to express the plastic displacement δ_{pl} in terms of the stress intensity factor K. This relationship has been provided in the form [18]

$$\Delta\delta_{pl} = \frac{2\kappa}{G\sqrt{\pi}} \cdot \frac{\sqrt{1-n^2}}{n} \Delta K\sqrt{a}, \tag{16}$$

where G is the shear modulus, and $\kappa = 1$ or 1ν depending on whether screw or edge dislocations are being considered, with ν being Poisson's ratio. Parameter n describes the relative position of the crack tip to the grain boundary (see Fig. 5). The number of stress cycles required for a crack to propagate through each crystal grain is obtained with integration of Eq. (15)

$$N_j = \int_{a_{j-1}}^{a_j} \frac{\mathrm{d}a}{C_0(\delta_{pl})^{m_0}}; \quad j = 1, 2, 3, \ldots, z \tag{17}$$

in which z is the number of grains transverse by the crack ($z = a/D$).

Integration limits a_{j-1} and a_j in Eq. (17) are in each grain determined in relation to the critical value of parameter n. When n reaches the critical value $n = n_c$, the slip band extends to the next grain. The critical parameter n_c is given by [18]

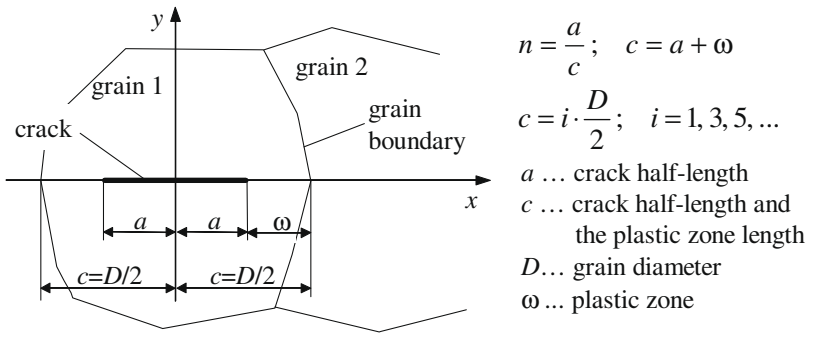

Fig. 5. Schematic representation of the short crack growth.

$$n_c = \cos\left[\frac{K}{2\sigma_y}\sqrt{\frac{\pi}{c\cdot n_c}}\left(1-\frac{K_{th}}{K}\sqrt{n_c}\right)\right], \tag{18}$$

where σ_y is the yield stress and K_{th} is the threshold stress intensity factor. This equation can be solved iteratively by Newton's method in three or four iterations by taking $n_c = 1$ as the initial value. The total number of stress cycles N required for a short crack to propagate from the initial crack length a_0 to any crack length a can then be determined as

$$N = \sum_{j=1}^{z} N_j. \tag{19}$$

As the crack extends through ten or more grains, the influence of the material structure on the crack growth becomes negligible and linear elastic fracture mechanic theory can be applied thereafter [20]. In the framework of the LEFM the propagation rate of long cracks can be described by the Paris' equation [21]

$$\frac{da}{dN} = C\cdot(\Delta K)^m, \tag{20}$$

where C and m are experimentally determined material parameters. However, this equation holds only before the crack reaches some critical crack length, when the crack growth rapidly increases until it becomes uncontrollable.

Expressing the plastic displacement $_{pl}$ in Eq. (16) in terms of the stress intensity range K enables treatment of short and long cracks in a similar fashion. If the relationship between the stress intensity range and the crack length $K = f(a)$ can be derived in some way, the remaining service life of a mechanical element with the crack can be estimated with appropriate integration of rate Eq. (15) (see [22]). Considering small crack lengths observed during pitting in the contact area of mechanical elements, only the theory of short cracks is usually needed for describing crack propagation from the initial to the critical crack length.

For simple cases, the relationship between the stress intensity factor and the crack length $K = f(a)$ is available in the technical literature in analytical form [23]. However, for cases with complicated geometry and boundary conditions, it is necessary to use alternative methods for its determination. Here, the numerical procedure based on the virtual crack extension method in the framework of the finite element analysis is used for that purpose [24].

4.1. Prediction of the crack growth with the virtual crack extension method

The virtual crack extension method (VCE), originally proposed by Hellen [25], is based on the criteria of released strain energy dV per crack extension da

$$G = -\frac{dV}{da}, \tag{21}$$

where G is the energy release rate, which serves as a basis for determining the combined stress intensity factor around the crack tip

$$K = \begin{cases} \sqrt{E\cdot G} & \text{plane stress,} \\ \sqrt{\frac{E\cdot G}{(1-\nu^2)}} & \text{plane strain} \end{cases} \tag{22}$$

with K further being equal to

$$K = \begin{cases} \sqrt{(K_{\mathrm{I}}^2 + K_{\mathrm{II}}^2) + (1+\nu)\cdot K_{\mathrm{III}}^2} & \text{plane stress,} \\ \sqrt{(K_{\mathrm{I}}^2 + K_{\mathrm{II}}^2)\cdot(1-\nu^2) + (1+\nu)\cdot K_{\mathrm{III}}^2} & \text{plane strain.} \end{cases} \tag{23}$$

If V_C is the strain energy obtained for all degrees of freedom not present in the crack tip elements, and V_N is the energy in the crack tip elements when the tip is not extended, while V_D is the energy in these elements when the tip is extended, Fig. 6, then the total energies of the initial and altered bodies, V_N^T and V_D^T, respectively, are equal to

$$V_N^T = V_C + V_N \quad \text{and} \quad V_D^T = V_C + V_D. \tag{24}$$

Thus for a virtual crack extension δa it follows that:

$$\frac{dV}{da} = \frac{V_D^T - V_N^T}{\delta a} = \frac{V_D - V_N}{\delta a}, \tag{25}$$

which is clearly independent of V_C. It follows that only strain energies V_N and V_D in the crack tip elements need to be calculated for every possible crack extension. This results in a very efficient method for determination of the instantaneous energy release rate and thus the stress intensity factor for any given crack extension.

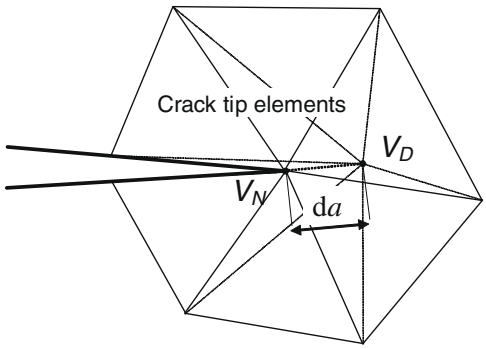

Fig. 6. Initial and extended crack tip configuration.

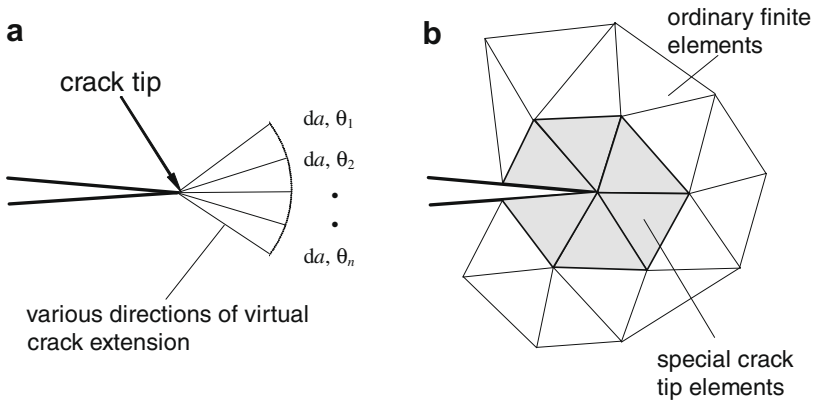

Fig. 7. Virtual crack extensions of the crack tip (a) and special crack elements (b).

Following the same argument, the energy release rate G and the stress intensity factor K can be easily determined for several different possible crack extension directions for a cluster of points on an arc around the initial crack tip with radius da, see Fig. 7a

$$\left(\frac{\mathrm{d}V}{\mathrm{d}a}\right)^j = \frac{V_D^j - V_N}{\mathrm{d}a^j}. \tag{26}$$

Assuming the validity of the maximum energy release criterion, the crack will propagate in the direction corresponding to the maximum value of $(\mathrm{d}V/\mathrm{d}a)^j$, i.e. in the direction of the maximum stress intensity factor K^j. The computational procedure is based on incremental crack extensions, where the size of the crack increment is prescribed in advance. In practise [25], it has been found that the virtual crack increment should be around 1/3 of the size of crack tip finite elements. For each crack extension increment the stress intensity factor is determined in several different possible crack propagation directions and the crack is actually extended in the direction of the maximum stress intensity factor, which requires local re-meshing around the new crack tip. The incremental procedure is repeated until full fracture occurs or until the stress intensity factor reaches the critical value K_c, when full fracture is imminent. For improved numerical results, special fracture finite elements are used in the first circle of elements around the crack tip, with ordinary elements elsewhere, Fig. 7b. In these special fracture finite elements, the displacements are proportional to the square root of the distance from the tip. Since the tip stresses are singular, they are not calculated at the crack tip node.

Following the above procedure, one can numerically determine the functional relationship $K = f(a)$ and the critical crack length a_c at $K = K_c$ from the computed values of K at discrete crack extensions a.

4.2. Parameters influencing the surface crack propagation

4.2.1. Moving contact

For a more realistic simulation of fatigue crack growth, it is necessary to consider the influence of moving contact in the vicinity of initial crack. The moving contact can be simulated with different loading configurations as shown in Fig. 8. Five contact loading configurations have been considered, each with the same normal $p(x)$ and tangential $q(x)$ contact loading distributions, but acting at different positions in respect to the initial crack.

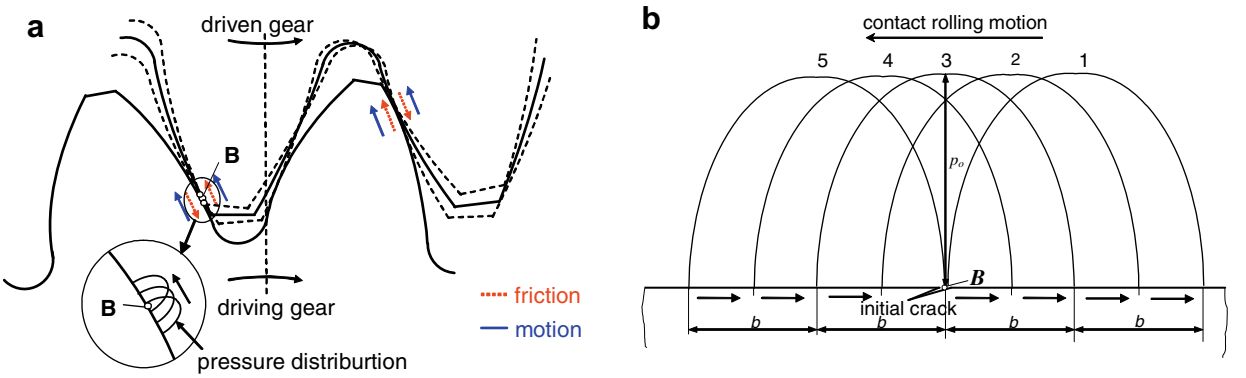

Fig. 8. Meshing of gear pair (a) and moving contact loading configurations in respect to the initial crack position (b).

4.2.2. The influence of fluid lubricant

The simulation of the surface initiated crack propagation should also consider the influence of the lubricant on crack propagation. Several mechanisms and models have been proposed in the past, and they can be basically divided into three categories [26]:

1. The hydraulic pressure mechanism, where the lubricant is driven into the crack by contact loading pressure and keeps the crack faces separated by its pressure, thus implying increased modes I and II separation at the crack tip due to additional normal pressure and lack of friction between crack faces.
2. Mode II crack propagation due to cyclic shear stresses along lubricated contacting crack surfaces, resulting in higher K_{II} values in comparison to dry crack surfaces.
3. The mechanism of trapped fluid, where the crack mouth closes under the contact pressure and the trapped fluid is driven toward the crack tip; this results in mode I separation at the crack tip, reduction of the crack face friction and increase in the mode II influence on crack propagation.

There is no clear consensus on which model is the most appropriate for surface initiated crack propagation simulation, since each model has some clear advantages and disadvantages and they mostly vary in the complexity of their implementation.

In the present computational procedure the hydraulic pressure mechanism has been adopted, where the lubricant pressure inside the crack is simply approximated with a uniform pressure distribution along the crack faces [26], Fig. 9.

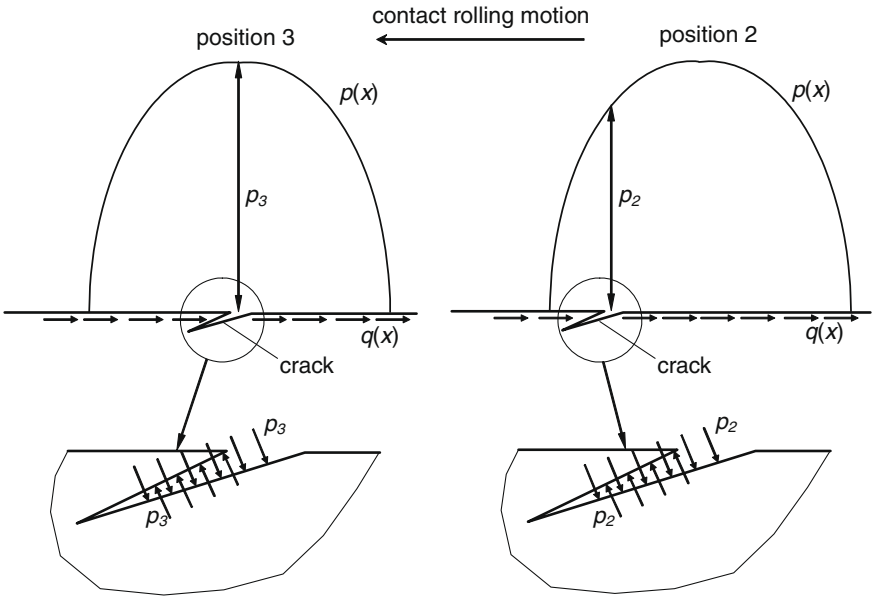

Fig. 9. Modelling of the lubricant driven in the crack by the hydraulic pressure mechanism.

The level of lubricant pressure is equal to the current pressure determined at the crack mouth, i.e. the pressure depends on the contact pressure distribution position in respect to the crack mouth. Fig. 9 illustrates crack face pressure determination and distribution for two consecutive contact loading configurations.

5. Practical example

The developed model has been used for the simulation of crack initiation and crack growth on a spur gear pair. The gear pair is made of high strength alloy steel 42CrMo4 with Young's modulus $E = 2.1 \times 10^5$ N/mm^2 and Poisson ratio $\nu = 0.3$. The maximum contact pressure $p_0 = 1550$ N/mm^2 acts at the inner point of single teeth pair engagement (point B, see Fig. 2), with the equivalent radius of gear teeth flanks $R^* = 10$ mm. Using Eqs. (2)–(6), the half-length of the contact area is equal to $b = 0.274$ mm. The Hertzian normal loading distribution $p(x)$ along the entire contact width of the gear flanks has then been determined using Eq. (2).

For all computations, the coefficient of friction $\mu = 0.04$ has been used, which is the average value for well-lubricated gears [27]. Therefore, the tangential loading $q(x)$ has been determined using Eq. (7).

5.1. Fatigue crack initiation

Contact loading cycles have been determined using a substitutional finite element model [8], presented in Fig. 3. The determination of contact loading cycles is presented, then further calculations of critical loading cycles for crack initiation are provided. The loading is moving along the contact surface of the generalised computational model (Fig. 3). Appropriate stress state for each observed material point is computed for each position of the moving contact loading (Fig. 3). This procedure results in generation of real stress loading cycles of material points in one pass of the rolling–sliding contact, which are necessary for the following contact fatigue analysis.

Shear stress and maximum principal stress based loading cycles are presented in Fig. 10, where the material point B (on y_1 and under y_2–y_{10} the surface) is considered, where y_i represents coordinate component of material point in vertical direction (i.e. depth under the surface), and y_1 represents material point on the surface, respectively.

Those cycles are further used for the fatigue analysis of damage initiation. The influence of position (on and under the observed inner point of single teeth pair engagement – point B, see Fig. 2) on the course of loading cycles was examined. The values of normalised stress cycles are higher under the contact surface, which is influenced particularly with such a low coefficient of friction.

The strain-life method (ε–N_i) and their three modified forms, also present in work [17], have been used for the determination of the fatigue crack initiation life, N_i, in the framework of the FEM program package MSC/FATIGUE [28]. The material parameters $\sigma'_f = 1820$ MPa, $\varepsilon'_f = 0.65$, $b = 0.08$ and $c = 0.76$, according to Eq. (13), have been taken from the material 42CrMo4 database available in [28]. Results of the number of loading cycles for crack initiation are presented in Fig. 11.

Considering maximum principal stress σ_1 based loading cycles, initial damage appears at (2.3–9.77 $\times$ 10^4) loading cycles on the contact surface; while considering shear stress τ_{xy} based loading cycles, initial damage appears at (5524–6457) loading cycles under the contact surface (at approximately 0.12 mm under the observed inner point of single teeth pair engagement – point B).

However, the determination of fatigue crack initiation life is, in this case, defined by the continuum mechanics approach and results are valid for approximately determined fatigue damage initiation, since the micro-mechanics effect was not considered within used material model.

Furthermore, the analysis of crack growth and appearance of fatigue damage on tooth flank, known as pitting formation, will be presented.

5.2. Fatigue crack propagation

The finite element mesh shown in Fig. 12, and the boundary conditions as described above, have been used in subsequent analyses. For the configuration of the initial crack on the surface, located at point B, it was assumed that the initial length of the crack is equal to $a_o = 15$ μm, with the initial inclination angle towards the contact surface equal to $\alpha_0 = 22°$. It is recognised, that the predicted crack growth heavily depends on the size and orientation of the initial crack. However, the used configuration follows from metallographic investigations of initial cracks appearing on similar gears [29]. A thorough investigation of the crack growth dependence on position and size of various initial defects is to be performed in future investigations with parametric simulations based on the proposed model.

In this study, the FE analysis program BERSAFE [30], based on VCE method, has been used for computational estimation of the stress intensity factor K and subsequent incremental crack growth simulation. In numerical computations, the crack increment was of size $\Delta a = 1.5$ μm. The stress intensity factor K was estimated in each crack increment for 30 different virtual crack tip extensions (see Fig. 7a). Five different loading configurations have been considered in each computation for the purpose of simulating the effect of the moving contact of the gear flanks (see Fig. 8). For each crack increment, the crack was actually extended in the direction of the recorded K_{max} from all calculated load cases. Fig. 13 shows the relationship between stress intensity factor K and the crack length a, and also the shape and magnitude of the surface pit. It can be seen that the

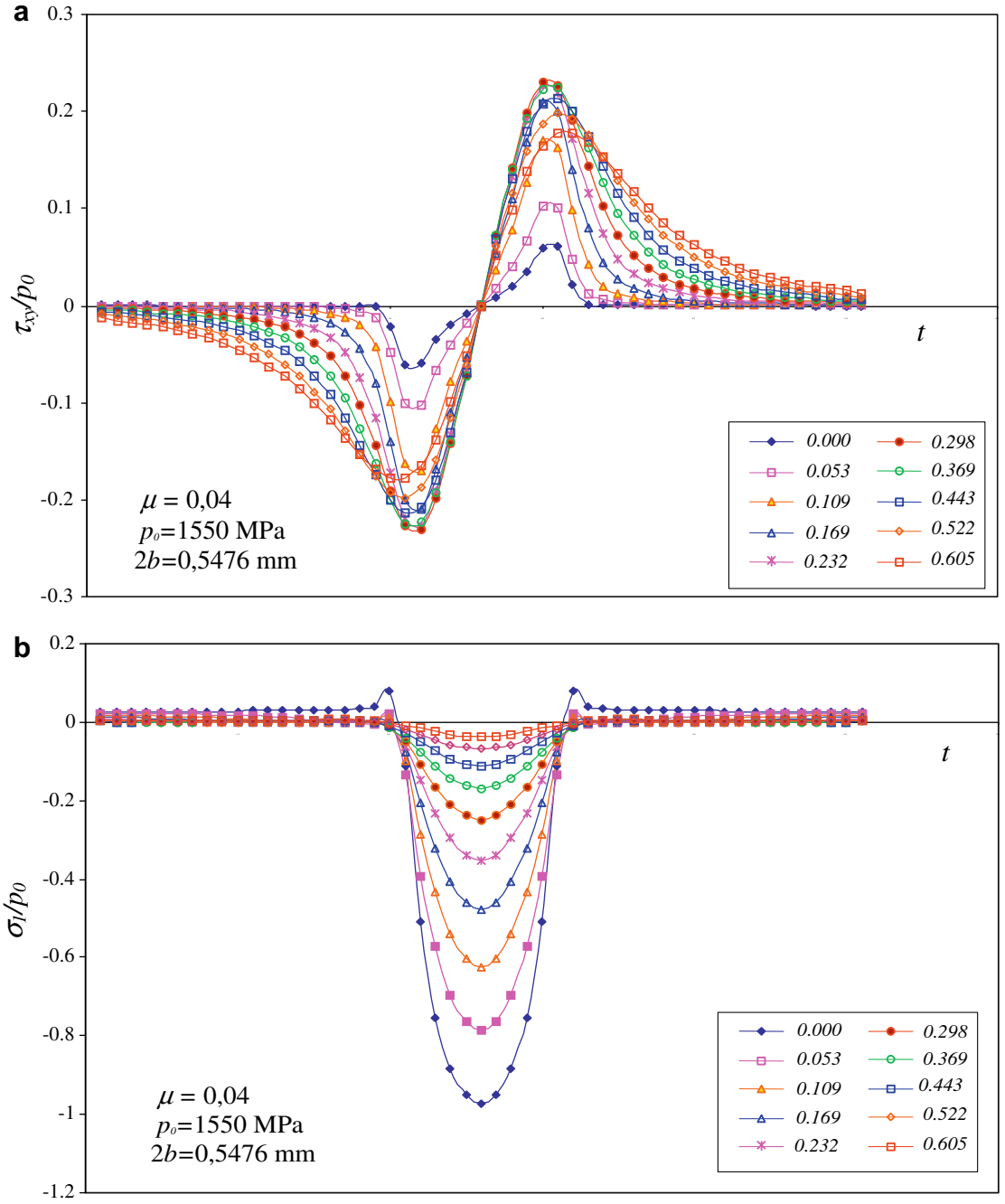

Fig. 10. Course of normalised contact loading cycles, (a) shear stress τ_{xy} based, (b) maximum principal stress σ_1 based; depths are in mm.

computed stress intensity factor K is very small at the beginning, but later increases as the crack propagates towards the contact surface. Numerical simulations have shown that at the moment when the crack reaches the vicinity of the contact surface, the stress intensity factor is extremely high. At that moment it can be expected that the material surface layer breaks away and the pit occurs on the surface. Because of the very small dimensions of surface pits, they can be termed "micro-pitting".

Micro-pitting as shown in Fig. 13 is not the final and most critical surface failure. Further operation of the gear pair results in the formation of larger pits, and consequently progressive pitting. In this respect, similar numerical simulations to those described above have been continued for two more steps, as shown in Fig. 14. In these calculations, it was assumed that the initial crack of length a_0 = 7.5 μm started from the bottom of the existing surface pit. For all numerical computations, the crack increment was of size Δa = 2 μm.

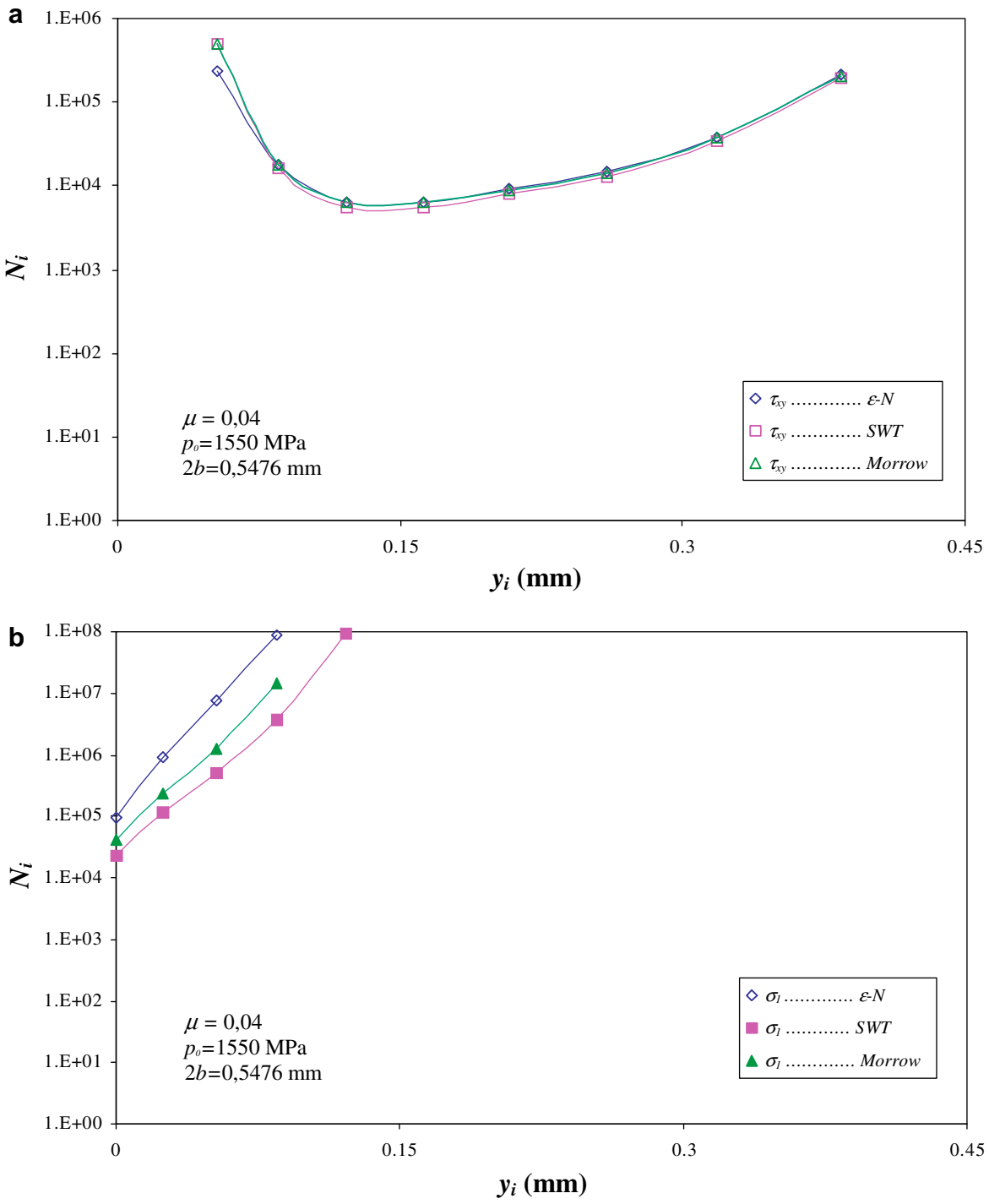

Fig. 11. Number of loading cycles for crack initiation using different ε–N initiation models (ε–N, *SWT* and *Morrow*), (a) shear stress τ_{xy} based, (b) maximum principal stress σ_1 based.

Following the above procedure, one can numerically determine the functional relationship $K = f(a)$, which is needed for determining the required number of loading cycles N_p for a crack propagating from the initial to the critical length (see Section 4 in this paper).

Fig. 15 shows that the shape and magnitude of the numerically determined pits correspond well with available experimental data, as determined by experimental testing of a spur gear pair using the FZG-pitting test machine according to DIN

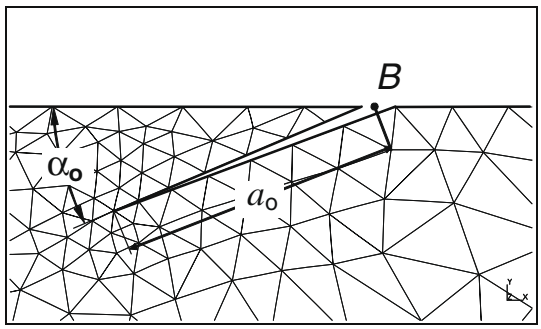

Fig. 12. FE discretisation and configuration of the initial crack.

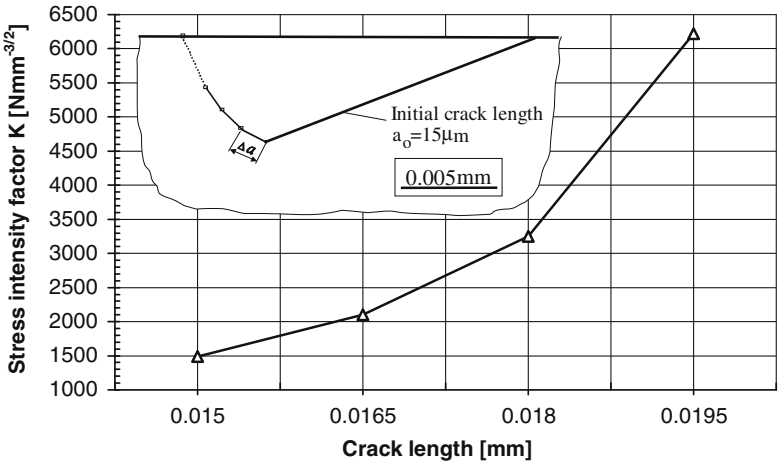

Fig. 13. Stress intensity factor for initial surface micro-crack propagation.

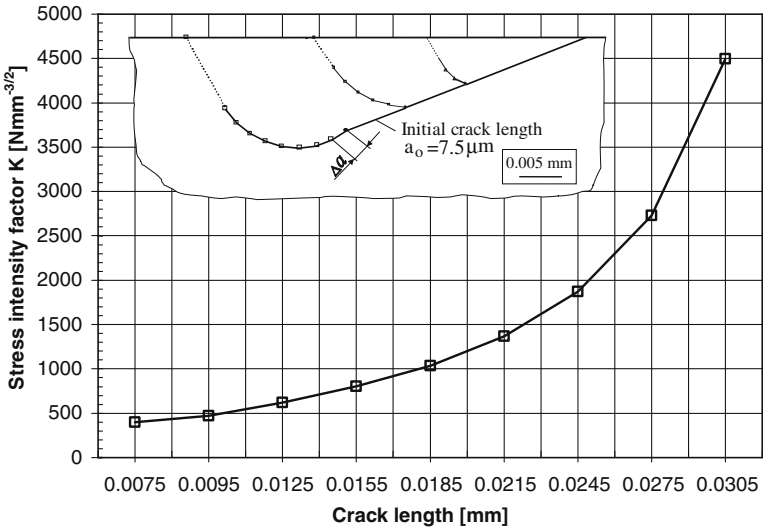

Fig. 14. Stress intensity factor for crack propagation from the existing micro-pit.

51354 standard [31]. The tested gears have been subjected to the same operating conditions and loading parameters as used in the numerical computations.

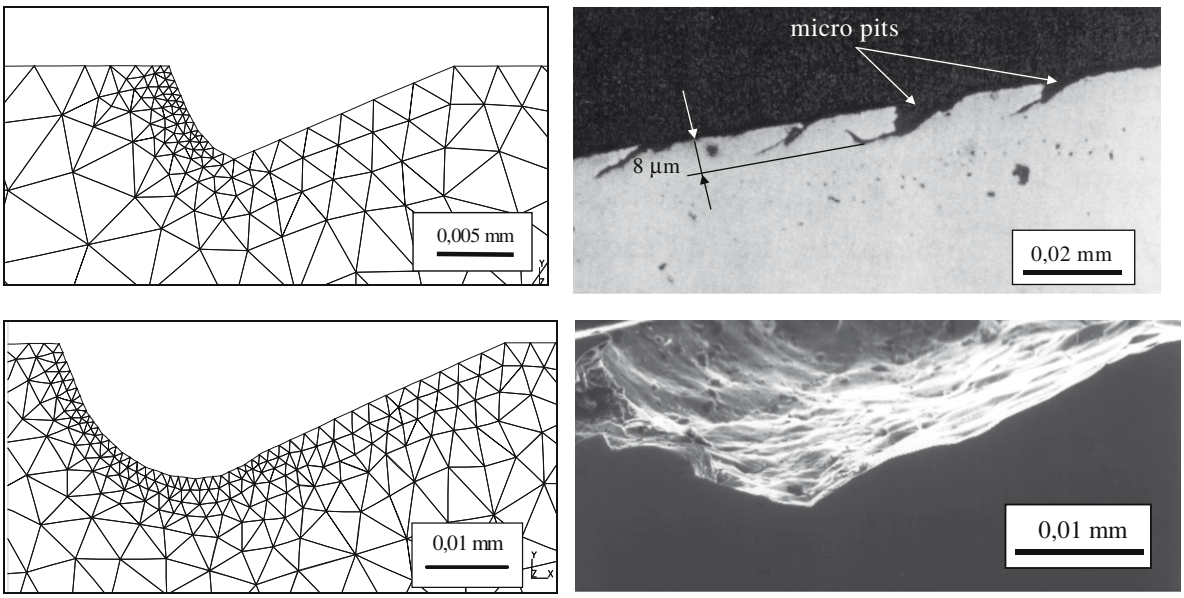

Fig. 15. Numerically (left column) and experimentally (right column) determined pit shape.

6. Concluding remarks

The paper presents a computational model for determining the service life of gears with regard to pitting on the gear teeth flanks. The fatigue process leading to pitting on these flanks is divided into crack initiation (N_i) and crack propagation (N_p) periods, which enables the determination of total service life as $N = N_i + N_p$.

A simple equivalent model of two cylinders is used for numerical simulation of contact fatigue crack initiation and propagation under conditions of rolling and sliding contact loading of gear teeth flanks. The equivalent model is subjected to the normal (normal contact pressure) and tangential (frictional forces) contact forces. The material model is assumed as homogeneous, without imperfections such as inclusions, asperities, roughness and residual stresses as often occur in mechanical elements. These drawbacks of the presented model should be improved in the analysis of practical applications (spur gears, rails, etc.), where all enumerating facts actually occur in some manner.

The crack initiation period is based on modified stress–strain analysis in the framework of finite element method, where the main observed point is the inner point of single teeth pair engagement at the gear tooth flank. Rolling–sliding boundary conditions on the contact surface are taken into consideration throughout using the friction coefficient. The presented numerical calculation for the determination of loading cycles considers the transition of sliding–rolling contact loading, the different fatigue analysis approaches, specified critical points of fatigue damage initiation, boundary conditions of geometry and loading (normal, tangential) on and under the surface.

The proposed model enables the determination of the number of loading cycles N_i, required for fatigue damage initiation, by means of certain loading cycles and the introduction of adequate material fatigue parameters. The method of dealing with numerical modelling and the possibility of predicting the fatigue damage initiation in mechanical elements as a consequence of cyclic contact loading represent this contribution to the discussed problems.

The computational model of crack growth is based on appropriate short fatigue crack growth theories. The model attempts to account for the different parameters influencing the crack propagation process leading to pitting (Hertzian contact pressure, friction between contacting surfaces, fluid trapped in the crack, meshing of the gears, etc.), starting from the initial surface fatigue crack to the critical crack length, when the occurrence of a surface pit is expected.

The virtual crack extension (VCE) method in the framework of finite element analysis is then used for two-dimensional simulation of fatigue crack propagation from the initial crack up to the formation of the surface pit.

Subsequent computational determination of the functional relationship $K = f(a)$ enables estimation of the service life of contacting mechanical elements in regard to the surface pitting, if combined with a previously developed model [19,26] and with consideration of some particular material parameters. The small crack lengths imply that the short crack growth theory should be used for the description of fatigue crack propagation as described in [19,26].

Based on the computational results and with consideration of some particular material parameters, the service life N of the gear flanks can then be determined from the number of stress cycles N_i required for crack initiation, and stress cycles N_p required for crack propagation from the initial to the critical crack length, when surface pitting can be expected. Although some influences (non-homogeneous material, travelling of dislocations, residual stresses, etc.) were not taken into account in the computational simulations, the presented model seems to be very suitable for determining the service life of gears. The

model can be further improved with additional theoretical and numerical research, although additional experimental results will be required to provide the required material parameters.

References

[1] Cheng W, Cheng HS, Mura T, Keer LM. Micromechanics modelling of crack initiation under contact fatigue. ASME J Tribol 1994;116:2–8.
[2] Kaneta M, Yatsuzuka H, Murakami Y. Mechanism of crack growth in lubricated rolling/sliding contact. ASLE Trans 1985;28:407–14.
[3] Glodež S, Ren Z, Flašker J. Simulation of surface pitting due to contact loading. Int J Numer Meth Engng 1998;43:33–50.
[4] Fajdiga G, Flašker J, Glodež S, Ren Z. Numerical simulation of the surface fatigue crack growth on gear teeth flanks. J Mech Engng 2000;46(6):359–69.
[5] Bhattacharya BA, Ellingwood B. Continuum damage mechanics analysis of fatigue crack initiation. Int J Fatigue 1998;20(9):631–9.
[6] Zhou RS, Cheng HS, Mura T. Micropitting in rolling and sliding contact under mixed lubrication. ASME J Tribol 1989;111:605–13.
[7] Johnson KL. Contact mechanics. Cambridge University Press; 1985.
[8] Šraml M, Flašker J, Potrč I. Numerical procedure for predicting the rolling contact fatigue crack initiation. Int J Fatigue 2003;25:585–95.
[9] Msc/Corporation, MSC/NASTRAN Quick Start Guide, Version 70.5. Los Angeles: The MacNeal-Schwendler Corporation; 2000.
[10] Smith KN, Watson P, Topper TH. A stress–strain function for the fatigue of metals. J Mater, JMLSA 1970;5:767–78.
[11] Morrow JD. Cyclic plastic strain energy and fatigue of metals. In: Internal friction, damping, and cyclic plasticity. ASTM; 1965. p. 45–86.
[12] Basquin OH. The exponential law of endurance tests. Am Soc Test Mater Proc 1910;10:625–30.
[13] Coffin LF. A study of the effect of cyclic thermal stresses on a ductile metals. Trans ASME 1954;76:931–50.
[14] Manson SS. Fatigue: a complex subject-some simple approximation. Exp Mech 1965;5:193–226.
[15] Neuber H. Kerbspannungslehre. Berlin: Springer; 1958 [in German]: Theory of notch stresses, USOTS, MI; 1961.
[16] Suresh S. Fatigue of materials. 2nd ed. Cambridge: Cambridge University Press; 1998.
[17] NAFEMS. Finite element based fatigue calculations. In: Bishop N, Sherratt F, editors. The international association for the engineering analysis community. Farnham; 2000.
[18] Navarro A, Rios ER. Short and long fatigue crack growth – a unified model. Philos Mag 1988;57:15–36.
[19] Sun Z, Rios ER, Miller KJ. Modelling small fatigue cracks interacting with grain boundaries. Fatigue Fract Engng Meter 1991;14:277–91.
[20] Glodež S. The fracture mechanics model of gear flanks fatigue. PhD thesis, Faculty of Mechanical Engineering, University of Maribor; 1995 [in Slovenian].
[21] Ewalds HL, Wanhill RJH. Fracture mechanics, Delftse U.M., Co-publication of Edward Arnold, Delft; 1991.
[22] Flašker J, Glodež S, Pehan S. Influence of contact area on service life of gears with crack in tooth root. Commun Numer Meth Engng 1995;11:49.
[23] ASTM E 399-83. Standard test method for plane-strain fracture toughness of metallic materials; 1983.
[24] Fajdiga G. Analysis of the surface fatigue crack growth on gear teeth flanks under EHD lubrication conditions. PhD thesis, University of Maribor; 2001 [in Slovenian].
[25] Hellen TK. On the method of virtual crack extensions. Int J Numer Meth Engng 1975;9:187–207.
[26] Bower AF. The influence of crack face friction and trapped fluid on surface initiated rolling contact fatigue cracks. ASME J Tribol 1988;110:704–11.
[27] Winter H, Knauer G. Einfluss von Schmierstoff und Betriebstemperatur auf die Grbchentragfhigkeit einsatzgehrteter Zahnrder. Antriebstechnik 1990;29:65–84.
[28] Msc/Corporation. MSC/FATIGUE Quick Start Guide, Version 8. Los Angeles: The MacNeal-Schwendler Corporation; 1999.
[29] Knauer G. Zur Grbchentragfhigkeit einsatzgehrteter Zahnrder. PhD thesis, TU Munich; 1988.
[30] *BERSAFE*; Users's guides; 1988.
[31] Elstorpff Marc-G. Einflüsse auf die Grübchentragfähigkeit. PhD thesis, TU Munich; 1993.