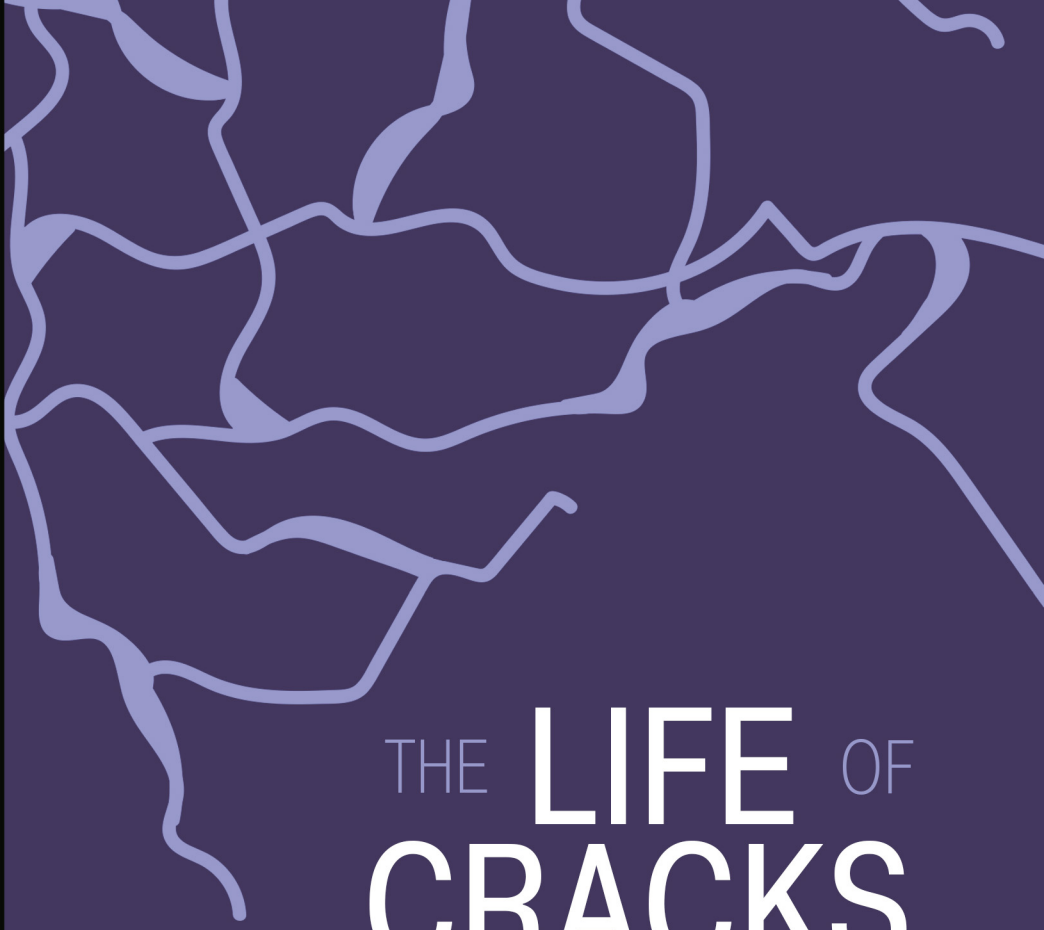

THE LIFE OF CRACKS

Theory and Application

Srečko Glodež and Boris Aberšek

The Life of Cracks

The Life of Cracks:

Theory and Application

By

Srečko Glodež and Boris Aberšek

Cambridge
Scholars
Publishing

The Life of Cracks: Theory and Application

By Srečko Glodež and Boris Aberšek

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-5538-0
ISBN (13): 978-1-5275-5538-9

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The authors gratefully wish to acknowledge:

- All of our colleagues from the Faculty of Natural Sciences and Mathematics, and the Faculty of Mechanical Engineering, University of Maribor for their contribution and helpful comments.
- Many thanks to prof. Jože Flašker, who encouraged us, directed us and allowed us to become who we are now. Thanks for all our conversations, in which we invented questions and jointly examined our intuitive answers.
- We are greatly indebted also to all of our students, for their inspiration and assistance; and many thanks to Helena Fošnjar for translation and proofreading.

The authors wish to express their appreciation to all those who have ensured the quality of this book. Last but not least, thanks to both our **M`s**, who have inspired us, supported us, and given us the opportunity to be who we are.

FOREWORD

The term *fracture and damage mechanics* is somewhat unsettling to many people. This is because, until recently, the major emphasis in mechanics was on the strength and resistance of materials. To speak of fracture is as uncomfortable for some as it is to speak of a fatal illness. However, just as in preventing a fatal disease, one must know its nature, symptoms, and behavior; to ensure the strength of a structure, one must be aware of the causes and nature of its potential failure.

The problem of fracture is vital in the science of strength of materials. However, not only has fracture mechanics, as an independent branch of the mechanics of deformable solids, originated quite recently, but its boundaries are not yet clearly defined. Therefore, it is of paramount interest to combine the efforts of representatives from many different branches of science and engineering for a complete study of the fracture concept. It is also important that differences in terminology (that are usual for different sciences), and the widespread conviction that the solution to everything lies in a particular portion of the general problem, do not lead to a situation in which disputes about the concepts are replaced by arguments about the words.

At present, routine fracture mechanics is the study of conditions under which a crack or a system of cracks undergoes propagation. However, cracks are of different natures, and are considered on different scale levels. The case on one extreme is the fracture of a crystal grain, which initiates with a submicroscopic crack when two atomic layers move apart by such a distance that the forces of interaction between the atoms may be neglected. An example of the other extreme is a crack occurring in a welded turbine rotor in a nuclear reactor, when the crack's length and width may amount to centimeters; this is referred to as a macroscopic fracture.

In the first case, the condition for crack propagation is defined by the configuration of atoms at the crack tip. Considered here is a discrete crystal lattice formed by atoms rather than a continuous medium; therefore, the very concept of the "crack tip" becomes uncertain. The study of this kind of submicroscopic crack and its behavior in interaction with other lattice defects is, essentially, in the province of solid-state physics rather than mechanics; however, the methods of classical theory of elasticity are fully

applicable to problems of this nature. The line between modern physics and mechanics is not well defined; nevertheless, it must be drawn to avoid possible terminological confusion.

A macroscopic fracture has dimensions exceeding by several orders the size of the largest structural constituent of the material (the constituent must contain a sufficient number of crystal grains for its properties not to differ from those of any other element of similar size which may be isolated from the material). It is precisely this condition that makes it possible to solve such a crack problem within the framework of mechanics of a solid body. The formulated condition refers to an ideal situation in order to make the theory applicable; in real conditions one may depart from this stringent requirement, but this in no way makes the theory groundless. Assuming the material to be continuous, homogeneous, and elastic, and using the techniques of the classical theory of elasticity, we inevitably arrive at the paradoxical conclusion that the stresses grow infinitely near the crack tip. This paradox is a sort of penalty paid for the simplicity associated with using the linear theory of elasticity in a region where its application is knowing to be invalid.

So-called linear fracture mechanics assume that a physically impossible singularity is a reality. Such an approach is not new and not so unusual for continuum mechanics; recall, for example, the vortex filaments with zero cross section and finite circulation. It appears that the work of crack propagation, which is done either as a result of increase of external forces or reduction of the elastic energy of the body with the crack size increase, is expressed directly through the coefficient of the singular term in the formula for stress. This coefficient is referred to as the stress intensity factor, and is of fundamental importance for the entire theory. The work of crack propagation may be associated with overcoming the forces of surface tension (Griffith's concept), or the plastic deformation in the small region of the immediate neighborhood of the crack tip, or other physical causes. The factor to be emphasized is that the size of the region, where the laws of the linear theory of elasticity are in some way violated, must be very small. The ability of the crack to further propagate is then determined by the sole characteristic: the work per unit length of the propagation path, or the critical stress intensity factor. If the size of the zone, where the relations of the linear theory of elasticity are violated, is large, one should consider the laws of nonlinear fracture mechanics. It appeared at the beginning that formal indifference of linear fracture mechanics to both the object and the scale, mathematical equivalence of problems associated with entirely

different physical phenomena, would make it possible to establish nonlinear mechanics in a similar uniform manner. It was later found to be quite different.

The principal problem, on which the efforts of scientists have been focused in recent years, concerns the conditions of either equilibrium or the propagation of a large crack in a sufficiently plastic material. Scientists have been involved in the theory and practical applications of fracture mechanics for evaluating the strength of large-scale structural elements. They have shown that the plastic zone ahead of a crack is sufficiently extensive so that the macroscopic theory of plasticity, which assumes that the medium is continuous and homogeneous, holds good. For the plane state of stress, the Leonov-Panasyuk-Dugdale model, which substitutes the plastic zone by a no-thickness segment extending the crack, appears to be satisfactory. In particular, this book presents an analysis of the corresponding elastic-plastic problem that is solved numerically by using the finite element method (FEM). The presented FEM-solution confirms the validity of the model used.

This book by S. Glodež and B. Aberšek is one of the first Slovenian monographs in international space on the above-discussed subject. It is based mainly on the results obtained by the authors during their original research and concerns the problems of fatigue and fracture mechanics. The greater part of the book is devoted to fatigue problems related to gears and other mechanical elements.

In spite of certain limits imposed by linear fracture mechanics, a wide variety of problems may be reliably solved using its methods. Development of this theory is focused on accumulating data from already solved elasticity problems concerning cracks of various shape in various bodies. The amount of such information continually grows both abroad and in Slovenia. Many results obtained by foreign authors became available by means of numerous books and published articles.

The present book may be considered as a significant contribution to the database of fracture mechanics, especially for gears. Some features of the book deserve special mention. First, it is the new variational principle that makes it possible to approximately solve numerous problems, in particular, to find the trajectory of crack propagation in a nonuniform stress field. Secondly, a straightforward approach for an approximate determination of the stress intensity factor is included; it enables one to obtain a reasonable evaluation for those cases where an exact solution of the elasticity problem

is impossible, and the numerical computation is extremely laborious. In addition, a series of newly solved dynamic problems for bodies subjected to cyclic (periodic) loading is provided.

CHAPTER 1

INTRODUCTION

Fatigue of engineering components and structures is a localised damage process produced by cyclic loading. In general, it has been observed that the fatigue process involves the following stages [1.1, 1.2]: (1) crack nucleation; (2) short crack growth; (3) long crack growth; and (4) final fracture. Fatigue cracks usually start on the localised shear plane or near high stress concentrations (persistent slip bands, inclusions, pores, etc.). Crack nucleation is the first phase in the complete fatigue process. Once the initial crack is nucleated and cyclic loading continues, the fatigue crack tends to grow along the plane of maximum shear stress.

Figure 1.1 shows a schematic representation of the fatigue process under cyclic loading where the crack nucleation (stage I) starts along the persistent slip bands. The next stage in the fatigue process is crack growth, which can be divided into short (stage II) and long (stage III) crack growth. Crack nucleation and short crack growth (stages I and II) are generally considered as a crack propagation across a few crystal grains in the plane of maximum shear stress. Here, the crack tip plasticity is greatly affected by slip band characteristics, grain size and their orientation, because the crack length is comparable to the material microstructure. Stage III corresponds to the long crack growth in the direction normal to the principal tensile stress. The long crack growth is less affected by the material microstructure because the crack tip plastic zone is much larger if compared to the microstructural properties (size of crystal grains). The fatigue process is finished when the crack reaches a critical length and final fractures occur (stage IV in Figure 1.1).

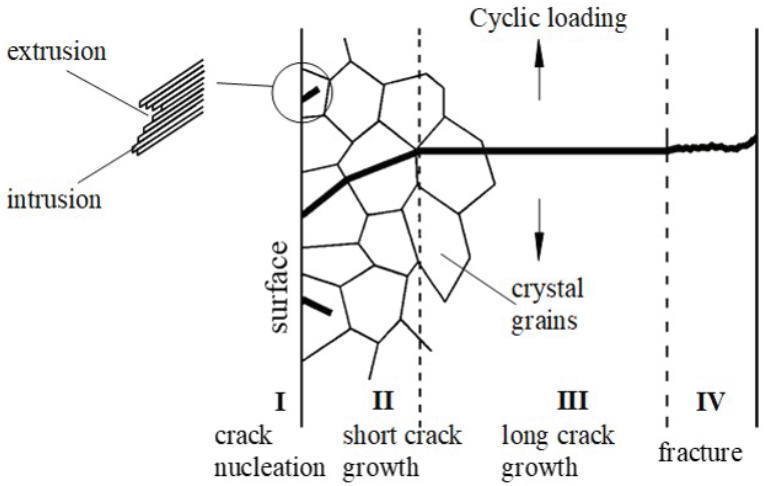

Figure 1.1: The fatigue process under cyclic loading [1.2, 1.3]

In engineering applications, the first two stages (crack nucleation and short crack growth) are usually termed as "*crack initiation period* N_i", while the last two stages (long crack growth and final fracture) are characterised as "*crack propagation period* N_p". The complete fatigue life of an analysed engineering component can then be determined from the number of stress cycles N_i required for fatigue crack initiation 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:

$$N = N_i + N_p \tag{1.1}$$

An exact definition of the transition period from the initiation of an "engineering" crack to its propagation is usually not possible. However, for engineering components made of steels the size of initial crack a_i after stages I and II (see Figure 1.1) is of the order of a few crystal grains of the material. This crack size usually ranges from about 0.1 to 1.0 mm. According to Dowling [1.4], the crack initiation size can also be estimated by the following equations:

$$a_i \approx \frac{1}{\pi}\left(\frac{\Delta K_{th}}{\Delta S_e}\right)^2 \qquad \text{smooth specimen} \tag{1.2}$$

$$a_i \approx (0.1 \ldots 0.2) \cdot R \qquad \text{notched specimen} \tag{1.3}$$

where ΔS_e is the stress range at the fatigue limit, ΔK_{th} is the range of the threshold intensity factor, and R is the notch-tip radius.

Once a crack has formed and propagated until the final fracture, the fatigue fracture surface can be inspected. A bending fatigue failure, as presented in Figure 1.2, generally leaves behind clamshell or beach markings. It is evident that the crack is nucleated at one side of the shaft and then propagates away from the nucleation site, usually in a radial manner. A semi-elliptical pattern is left behind. When the crack reaches a critical length, the final fracture occurs. In some cases, inspection of the size and location of the beach marks left behind may indicate where a different period of crack growth began or ended.

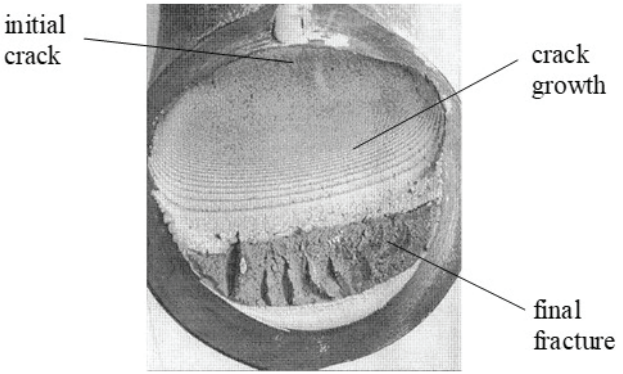

Figure 1.2: Fatigue fracture surface from a shaft in bending [1.5]

Figure 1.3 schematically shows fatigue fracture surfaces of cylinder specimens subjected to the axial or bending loading as a function of load magnitude (nominal stress) and geometry (notch effect) of the specimen. The white coloured regions identify the crack nucleation and the subsequent crack propagation while the grey coloured region represents the surface of final fracture. The direction of crack growth is marked with arrows where the beginnings of the arrows show the location of the crack nucleation. In each case shown in Figure 1.3, the fatigue cracks nucleate at the surface and then propagate in the plane of maximum tensile stress. In the case of uniaxial tension, tension/compression and unidirectional bending initial cracks are nucleated only at one side of the specimen, which corresponds to the locations of tensile stress. For reversed bending, cracks usually nucleate at opposite sides, since both sides are subjected to repeated tensile stress. In

rotating bending, the crack may be initiated around the whole circumference of the specimen. The surface of the final fracture is significantly influenced by the magnitude of nominal stress and notch effect of the specimen.

Similar presentations for fracture surfaces could also be drawn for round specimens subjected to torsional loading and also for flat specimens. Detailed information about these problems can be found in [1.2, 1.4].

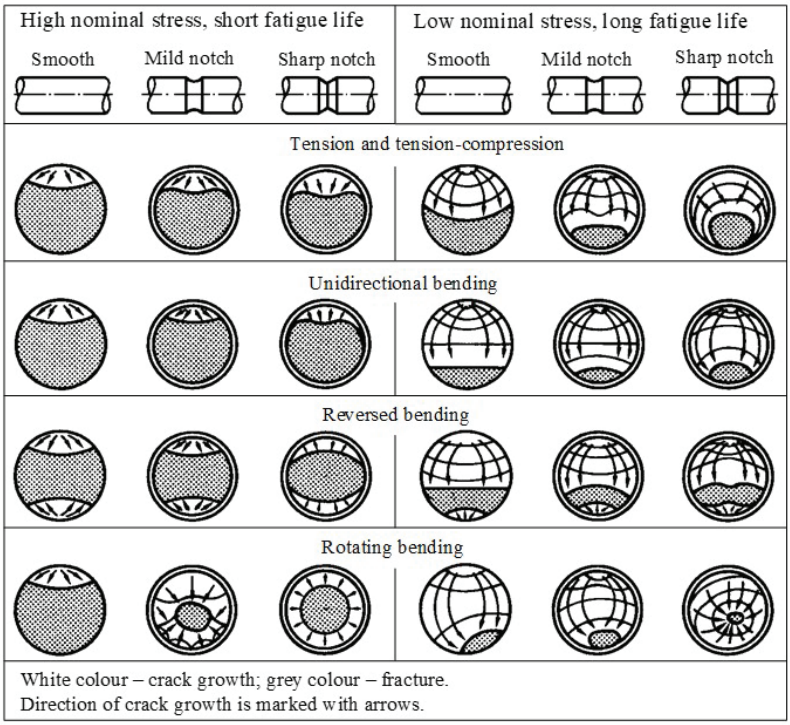

Figure 1.3: Schematic fatigue fractures of round specimens [1.2, 1.6]

1.1 Fatigue design approaches

The available fatigue design approaches have many similarities but also differences: an engineering component may be safety critical or nonsafety critical, the structure may be very simple or very complex, and failures may be a nuisance or catastrophic. Choosing the appropriate fatigue design

approach is a crucial decision for engineers when dimensioning dynamically loaded structural components or machine parts. Three main approaches currently exist in fatigue design [1.2]:

- Stress-life approach (*S*–*N*),
- Strain-life approach (ε–*N*),
- Fatigue crack growth approach (d*a*/d*N*–Δ*K*).

The stress-life approach has been in use for more than 150 years, while the other two approaches have been available only since the 1960s. Some authors [1.2, 1.4] have proposed a fourth model, the so-called "two-stage approach", which consists of combining the strain-life approach and the fatigue crack growth approach to incorporate both, macroscopic fatigue crack initiation and the subsequent fatigue crack growth.

1.1.1 Stress-life approach

The stress–life approach is the oldest method for dimensioning dynamically loaded machine parts and structural components. This approach is based on the *S*–*N* curves that are commonly plotted in terms of stress amplitude σ_a versus number of loading cycles to failure *N*. The most basic *S*–*N* curve is considered to be the one for zero mean stress $\sigma_m = 0$, which corresponds to the stress ratio $R = \sigma_{min}/\sigma_{max} = -1$. The relationship between the stress amplitude and the number of cycles to failure may be mathematically expressed as follows:

$$\sigma_a = \sigma_f{}'(2 \cdot N)^b \tag{1.4}$$

where σ_f' is the fatigue strength coefficient and *b* is the fatigue strength exponent. Material parameters σ_f' and *b* can be determined experimentally, usually by means of the rotating bending test [1.7, 1.8].

S–*N* curves vary with the material and its prior processing (thermal treatment). They are also affected by mean stress and geometry, especially the presence of notches, as well as by surface finish, residual stresses, loading frequency, environment conditions, etc. If an unnotched structural component is loaded with the mean stress $\sigma_m \neq 0$, the equivalent completely reversed stress amplitude σ_{ac} should be obtained when determining the fatigue life using eq. (1.4):

$$\sigma_{ac} = \frac{\sigma_a}{1 - \frac{\sigma_m}{\sigma_f'}} \quad \text{Morrow correction} \tag{1.5}$$

$$\sigma_{ac} = \sqrt{\sigma_{max} \cdot \sigma_a} \quad \text{SWT correction} \tag{1.6}$$

The equivalent completely reversed stress amplitude σ_{ac} actually represents the applied combination of stress amplitude σ_a and mean stress σ_m, which results in the same fatigue life as the stress amplitude σ_{ac} applied at zero mean stress. Beside the equations (1.5) and (1.6), other criteria to determine the value of σ_{ac} may be found in specialist literature [1.9, 1.10]. The choice of the appropriate criterion is mainly dependent on the material of the treated structural component.

Equation (1.4) basically represents the relationship between the stress amplitude and the number of cycles when uniaxial fatigue loading is applied. However, this equation may also be generalized and used to determine fatigue life under multiaxial loading. In such cases, the values σ_a and σ_m in eq. (1.4) are replaced by equivalent stresses σ_{aE} and σ_{mE}. Here, the equivalent stress amplitude σ_{aE} is proportional to the amplitude of the octahedral shear stress while the equivalent mean stress σ_{mE} is proportional to the hydrostatic stress due to mean stresses in three directions.

If more than one amplitude or mean stress level occurs (variable amplitude loading), the complete loading spectrum may be divided into the *M*-loading intervals inside which the values σ_{ai} and σ_{mi} are constant. The fatigue life of treated structural component can then be obtained using the Palmgren-Miner rule:

$$\sum_{i=1}^{M} \frac{n_i}{N_i} \leq 1 \tag{1.7}$$

where n_i is the real number of loading cycles inside the i^{th} interval and N_i is the number of loading cycles to failure from the *S–N* curve for σ_{ai} and σ_{mi} in this interval.

Machine parts and structural components often comprise different geometric discontinuities (holes, fillets, grooves, keyways, etc.), which are unavoidable in their design. These stress raisers (notches) reduce fatigue strength and require careful attention in the design process. In the stress-life approach, the notch-effect is taken into account by modifying the unnotched *S–N* curve considering the fatigue notch factor K_f:

$$K_f = 1 + q \cdot (K_t - 1) \tag{1.8}$$

where K_t is the stress concentration factor based on the linear elastic theory and q is the notch sensitivity factor dependent on the material and notch radius. A value $q = 0$ indicates that the notch effect can be neglected ($K_f = 1$), whereas a value $q = 1$ indicates full notch sensitivity ($K_f = K_t$). Two approaches which are often used for determining the notch sensitivity factor q are Neuber's method [1.11], and Peterson's [1.12] method.

In engineering design, the stress-life approach is usually applied by dimensioning machine parts and structures according to the *Infinite-Life Design* criterion, or the *Safe-Life Design* criterion (see section 1.2). Both criteria assume that the stresses in the critical cross-sections are in the linear elastic area. The calculation procedure is focused on the appearance of a final failure (fracture) of the treated component in the critical cross-section, and is not conditioned with crack nucleation and its growth.

1.1.2 Strain-life approach

The strain-life approach is based on the knowledge of stresses and strains that occur at locations where fatigue crack nucleation is likely to start, such as holes, fillets, grooves, etc. In this approach, the behavior of the material is characterized by the use of the stable cyclic stress-strain curve (eq. 1.9) and the strain-life curve (eq. 1.10) from uniaxial loading:

$$\sigma_a = \frac{\sigma_a}{E} + \left(\frac{\sigma_a}{K'}\right)^{\frac{1}{n'}} \tag{1.9}$$

$$\varepsilon_a = \frac{\sigma_f{}'}{E}(2N_i)^b + \varepsilon_f{}' \cdot (2N_i)^c \tag{1.10}$$

where σ_a is the stress amplitude, E is Young's modulus, K' is the cyclic strength coefficient, n' is the cyclic strain hardening exponent, σ_f' is the fatigue strength coefficient, b is the fatigue strength exponent, ε_f' is the fatigue ductility coefficient, c is the fatigue ductility exponent, and N_i is the number of loading cycles required for fatigue crack initiation (according to the eq. 1.1). Material parameters K', n', σ_f', b, ε_f' and c can be determined experimentally, usually by means of uniaxial fatigue tests according to the ASTM E606 standard [1.11]. These parameters vary over a range of different engineering materials and their processing histories (such as thermal treatment).

Equation (1.10) generally applies for zero mean stress $\sigma_m = 0$, which corresponds to completely reversed straining, $R = \varepsilon_{min}/\varepsilon_{max} = -1$. However, a mean strain or stress can be present in many engineering applications. In this case, the effect of mean stress σ_m can be considered using the following equations [1.2, 1.4]:

$$\varepsilon_a = \frac{\sigma_f{}' - \sigma_m}{E}(2N_i)^b + \varepsilon_f'(2N_i)^c \qquad \text{Morr} \tag{1.11}$$

$$\varepsilon_a = \frac{\sigma_f{}' - \sigma_m}{E}(2N_i)^b + \varepsilon_f'\frac{\sigma_f{}' - \sigma_m}{\sigma_f{}'}(2N_i)^c \quad \text{Manson} \tag{1.12}$$

$$\sigma_{max}\varepsilon_a E = \left(\sigma_f{}'\right)^2(2N_i)^{2b} + \sigma_f{}'\varepsilon_f' E(2N_i)^c \quad \text{SWT} \tag{1.13}$$

If multiaxial loading occurs, then the stress-strain and the strain-life relations need to be used in a more general form. Analogous to the equivalent stress approach (see section 1.1.1), the equivalent strain amplitude ε_{aE} can be used when determining fatigue life N_i. The most commonly used theories for determining the equivalent strain amplitude ε_{aE} under proportional loading conditions are the maximum principal strain theory (appropriate for brittle materials), and the maximum shear strain theory (appropriate for ductile materials):

$$\varepsilon_{aE} = \varepsilon_{a1} \qquad \text{maximum principal strain theory} \tag{1.14}$$

$$\varepsilon_{aE} = \frac{\varepsilon_{a1} - \varepsilon_{a3}}{1+\nu} \qquad \text{maximum shear strain theory} \tag{1.15}$$

In equations (1.14) and (1.15), ν is Poisson's ratio, and ε_{a1} and ε_{a3} are principal alternating strains, where $\varepsilon_{a1} > \varepsilon_{a3}$. Once an equivalent strain amplitude, ε_{aE}, has been calculated using the appropriate strain theory, the value ε_a is replaced with the value ε_{aE} to determine fatigue life using equations (1.10) to (1.13).

In the case of variable amplitude loading, the strain-life approach accounts for load sequence effects and is generally advantageous for cumulative damage analyses of notch members where significant plasticity usually occurs due to notch effect. The strain-life approach for fatigue of notched members consists of two stages. First, the stress and strains around the notch are determined (usually numerically by FEM). Afterwards, a life prediction

is made using the appropriate strain-life equation with consideration of the notch root stresses and strains.

The strain-life approach is a comprehensive approach, which can be applied for determining fatigue life of structural components in both, low-cycle fatigue (LCF) and high-cycle fatigue (HCF) regimes. In the low-cycle region, the component of plastic strain is dominant, while in the high-cycle region the elastic strain component is dominant. From that perspective, ductile materials have better fatigue resistance at large strain, whereas the material strength is the crucial parameter against fatigue failure at smaller strains.

The strain-life approach can also be used in combination with the fatigue crack growth approach (see section 1.1.3) to obtain total fatigue lives for crack initiation and the subsequent crack growth. From this point of view, the *Safe-Life Design* criterion and the *Fail-Safe Design* criterion (see section 1.2) are usually used when determining the fatigue life of cyclic loaded components.

1.1.3 Fatigue crack growth approach

The fatigue crack growth approach requires the use of fracture mechanics and its integration into the fatigue crack growth theory to obtain the number of loading cycles required for crack propagation form initial to the critical crack length, when final fracture of the treated component can be expected to occur. This approach can be applied to determine total fatigue life when it is used in conjunction with information on the existing initial crack, which has been, for example, detected by previous examination. The "two-stage approach" (see section 1.1) means that the fatigue crack growth approach is combined with the strain-life approach. In this case, total fatigue life can be determined from the number of stress cycles N_i required for fatigue crack initiation (previously determined using the strain-life approach) and the number of stress cycles N_p required for a crack to propagate from the initial to the critical crack length (subsequently determined using the fatigue crack growth approach); see also equation (1.1).

In engineering design, the fatigue crack growth approach is usually used by dimensioning the machine parts and structures considering the *Fail-Safe Design* criterion (see section 1.2). Moreover, this approach is the most often used approach in different case studies presented in this book. In that respect, the fatigue crack growth approach is described in more detail in section 2.

1.2 Fatigue design criteria

As presented in the previous section, three basic approaches can be used for dimensioning structural elements subjected to cyclic loading. However, from a philosophical point of view, four fatigue design criteria can be combined with these three approaches [1.2]:

- Infinite-Life Design,
- Safe-Life Design,
- Fail-Safe Design,
- Damage-Tolerant Design.

Infinite-Life Design

Infinite-Life Design is the oldest fatigue design criterion. This criterion is based on the *S–N* curve with the assumption that the engineering component is going to reach "infinite" life (usually several millions of cycles). According to this criterion, the stresses are in the elastic area and should not exceed the fatigue limit of the material.

This criterion is suitable for dimensioning dynamically loaded machine parts or structures which are in the framework of their fatigue life actually exposed to millions of cycles (engine valve springs, axes and bearings of railway wagons, shafts and gears of high stages in change-speed gear drives, etc.). However, most engineering components undergo significant variable amplitude loading, and the pertinent fatigue limit is difficult to obtain. In addition, this criterion may not be economical in many design situations where the expected fatigue life is shorter (shafts and gears of low stages in change-speed gear drives, certain parts in the aircraft industry, etc.).

The infinite-life design criterion is exclusively combined with the stress-life approach (see section 1.1.1). The main advantage of this method is the fact that fatigue time is known for most engineering materials, and that this information is available in specialist literature. Furthermore, certain design parameters (surface roughness, notch effect, residual stresses, temperature, corrosion, etc.), which may significantly influence the fatigue strength, are also well understood.

Safe-Life Design

Safe-Life is a fatigue design criterion where the engineering component is designed for a finite life, which is often known in advance. The safe-life

criterion should include a margin for the scatter of fatigue results and for other unknown factors. The dimensioning process may be based on the stress-life approach, if stresses are in the elastic area, or on the strain-life approach, if plastic deformation occurs in the critical cross section of the treated component.

This criterion is suitable for dimensioning dynamically loaded machine parts or structures with an expected specific finite life (i.e., reverse gears in car drives, pressure vessels design, jet engine design, etc.). Similar to the infinite-life design criterion, some design parameters (such as surface roughness, notch effect, residual stresses, etc.) should also be considered when determining the finite fatigue life.

Fail-Safe Design

Fail-Safe Design is a fatigue design criterion, which assumes that some initial failures (cracks) may appear in individual parts of the engineering structure but these failures are not critical and do not lead to a catastrophic failure of the structure. This fatigue design criterion was developed in the aircraft industry. Namely, aircraft engineers could not tolerate the added weight required by large safety factors, or the danger to life created by small safety factors, or the high cost of the safe-life design. In that respect, the fail-safe design is based on the requirement that the system does not fail if one part fails. This principle recognizes that fatigue cracks may occur, and structures are arranged so that cracks will not lead to failure of the structure before they are detected and repaired. Following this idea, multiple load paths, load transfer between members, crack stoppers built at intervals into the structure, and inspection are some of the means used to achieve a fail-safe design. Although this approach was originally applied mainly to aircraft structures (fuselages, wings), it is now used in many other applications.

Damage-Tolerant Design

The *Damage-Tolerant Design* criterion is actually a refinement of the Fail-Safe Design criterion. It is based on the assumption that cracks or initial defects exist in engineering structures which were caused either by mechanical and thermal treatment of components during the manufacturing process, or by fatigue. A fracture mechanics analysis can then be performed, in order to determine whether such cracks will grow large enough to produce failures before they are detected by periodic inspection. When

dimensioning machine parts and structures using the damage-tolerant design criterion, at least three issues should be taken into account:

- crack detection involving nondestructive inspection,
- residual strength of treated component, and
- fatigue crack growth behavior.

In recent decades, several nondestructive inspection methods have been developed to detect possible defects (cracks) in a treated engineering component. If a crack is detected, the residual strength of the treated component should be obtained using fracture mechanics theory. As a crack propagates under cyclic loading, the residual strength decreases up to the critical crack length, when final failure (fracture) occurs. If there is no crack, the residual strength is equal to the ultimate tensile strength or yield stress of the material. Apart from those described above, some other influencing parameters, such as environmental conditions, load history, statistical evaluation, etc., should also be incorporated into this methodology.

The damage-tolerant design criterion is often used when evaluating the residual strength of complex and expensive engineering components, which should be retired from service because they have reached their designed safe-life service life, based upon analytical and experimental results. However, it has often been established that such components could have significant additional service life. To allow for possible extended service life, damage tolerant methodology based upon both, analytical (or numerical) analyses and additional experimental testing, is required.

References

[1.1] Lee YL, Pan J, Hathaway RB, Barkey ME. *Fatigue testing and analysis*, Elsevier, Amsterdam, 2005
[1.2] Stephens RI, Fatemi A, Stephens RR, Fuchs HO. *Metal Fatigue in Engineering*: John Wiley & Sons Inc, New York, 2001.
[1.3] Glodež S, Flašker J, *Dimensioning on the service life*, Scientific monograph, University of Maribor, Maribor, 2005.
[1.4] Dowling NE, *Mechanical Behaviour of Materials*, Prentice Hall, New Jersey, 1999.
[1.5] Draper J, *Modern metal Fatigue Analysis*, Emas Publishing, Warrington, 2007.
[1.6] Radaj D, *Ermüdungsfestigkeit*, Springer Verlag, Heidelberg, 2003. (in German)

[1.7] ISO 1143:2010(en), *Metallic materials–Rotating bar bending fatigue testing*.
[1.8] ASTM E2948, *Standard Test Method for Conducting Rotating Bending Fatigue Tests of Solid Round Fine Wire.*
[1.9] Zahavi E., *Fatigue Design*, CRC Press, Boca Raton, FL, 1996.
[1.10] Rice RC., *SAE Fatigue Design Handbook*, 3rd ed., AE-22, SAE, Warrendale, PA, 1997.
[1.11] ASTM E606, *Standard Test Method for Strain-Controlled Fatigue Testing.*

CHAPTER 2

FATIGUE AND FRACTURE MECHANICS – THEORETICAL BACKGROUND

The presence of a crack in an engineering component can significantly reduce its fatigue life, as already discussed in Chapter 1. This chapter introduces the concept and use of *fracture mechanics* in the fatigue crack growth approach (see section 1.1.3) when dimensioning machine parts and structural components according to the fail-safe design criterion or the damage-tolerant design criterion (see section 1.2). When a crack is detected in a treated engineering component, the following questions are usually topical for designers:

- What is the residual strength of the treated engineering component?
- What is the critical crack length that still assures the safety operation of the component?
- What is the operation time (number of loading cycles) needed for crack extension from the initial to the critical length?
- What is the fatigue life of a component with a detected micro-crack (i.e., as a consequence of mechanical or thermal treatment)?
- How frequent are periodic inspections of components with detected cracks?

As implied above, the effective use of fracture mechanics requires periodic inspections of cracked components in order to determine what sizes and geometries of cracks are present or might be present. Such periodic inspections are often performed on aircraft, bridges, nuclear constructions, pressure vessels and many other engineering applications. According to this philosophy, a crack cannot grow to a dangerous length and the critical part of the structure will be repaired before a critical situation could happen. Methods of inspection to detect the presence of possible cracks can be simple visual examinations, or more comprehensive examinations such as X-ray photography or ultrasonics. Repairs necessitated by cracks may consider replacing a part or repairing it, for example, by machining away a small crack to leave a smooth surface [2.1–2.3].

This chapter provides the theoretical background for fracture mechanics and its use when solving engineering problems related to the fatigue of cyclically loaded machine parts and structural components. The topic is focused only on that part of the fracture mechanics theory, which is then actually used in the subsequent case studies presented in Chapters 3 to 6 of this book. In specialist literature, different concepts may be found that explane the evaluation of a component with a crack [2.4–2.6]:

- *K*-concept (the stress intensity factor concept)
- *G*-concept (the energy release rate concept)
- *COD*-concept (the crack opening displacement concept)
- *J*-concept (the J-integral concept)

This chapter focuses predominantly on the *K*-concept and its application to fatigue problems.

2.1 Linear elastic fracture mechanics

Linear Elastic Fracture Mechanics (*LEFM*) is used to determine crack growth in components under the basic assumption that material conditions are predominantly linear elastic during the fatigue process. This is usually the case in the High Cycle Fatigue (HCF) regime, or when using hard and brittle materials (i.e., high strength steels), where the actual stresses are lower if compared to the yield stress of the material.

Basic investigations related to LEFM have been made by Griffith [2.7], who defined the energy release rate, *G*, which represents the elastic energy per unit crack surface area required for crack extension. However, Griffith's researches were limited on very brittle materials (glass). Irwin [2.8] and Orowan [2.9] made significant advances by applying Griffith's theory to metals while taking into account small plastic deformations at the crack tip. The next stage was the energy approach to crack extension, which was developed by Irwin [2.10] and further improved through the implementation of the stress intensity factor *K* [2.11, 2.12.].

2.1.1 Loading modes

The theoretical background of fracture mechanics is based on three loading modes of cracked components as shown in Figure 2.1:

- *Mode I* – opening mode

- *Mode II* – shearing mode
- *Mode III* – tearing mode

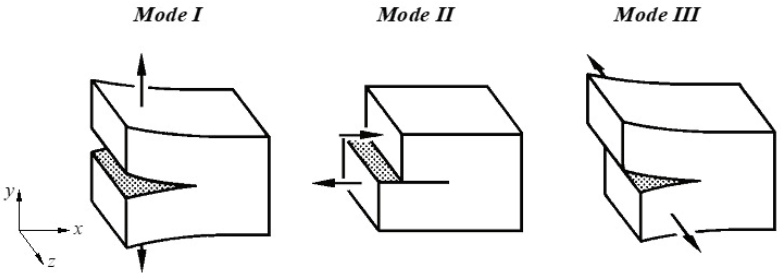

Figure 2.1: The basic loading modes of cracked components

Mode I (*opening mode*) is the most common when solving uniaxial fatigue problems and has received the greatest amount of investigation in the past. In this mode, tensile loading acts perpendicular to the crack front and causes a crack opening in the plane of the maximum tensile stress. In Mode II (*shearing mode*), shear loading acts on the plane x-y along the crack front and causes the sliding of crack faces relative to one another along the plane of the maximum shear stress. Mode III (*tearing mode*) also involves relative sliding of the crack faces, but now in plane y-z (perpendicular to the crack front).

Machine parts and structural elements are often subjected to external loading where more than one mode is present. In such cases, the crack propagates according to the *mixed-mode crack extension*. An example of mixed-mode crack extension I and II is shown in Figure 2.2, where appropriate stress intensity factors K_I and K_{II} are also designated. This involves the axial loading of a crack inclined for an angle β in respect to the x-axis. When $\beta = 0°$, the pure mode I is present ($K_{II}=0$), when $\beta = 90°$, the pure mode II is present ($K_I=0$). For all other values of β, a mixed-mode of crack extension appears.

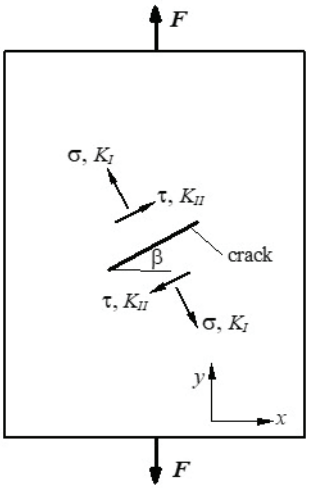

Figure 2.2: Mixed-mode I and II

2.1.2 Stress intensity factor

In general, the *stress intensity factor K* characterizes the intensity of the stresses in the vicinity of an ideally sharp crack tip in a linear-elastic and isotropic material [2.1]. Figure 2.3 shows the elastic stress σ_y near the crack tip ($r/a << 1$) in an elastic isotropic body subjected to the pure mode I loading. It is evident that the magnitude of this stress at a given point is dependent entirely on K_I.

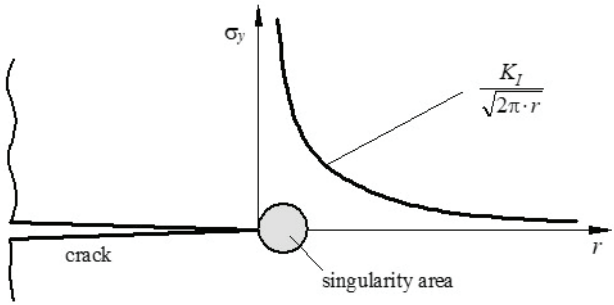

Figure 2.3: Elastic stresses near the crack tip for pure mode I loading [2.5]

It can be seen from Figure 2.3, that the elastic stress distribution σ_y approaches infinity ($\sigma_y \to \infty$) as r approaches zero ($r \to 0$). Following this finding, the stress intensity factor for mode I loading can be defined mathematically as:

$$K_I = \lim_{r \to 0} \sigma_y \cdot \sqrt{2\pi r} \tag{2.1}$$

Similar equations can also be expressed for mode II and mode III loading. Values of K are generally dependent on the external loading, crack length and geometry of the cracked member. Explicit equations for the determination of K can be found in specialist literature [2.13]. On the other hand, K can also be determined numerically using the available numerical approach (i.e., FEM).

When the crack is small compared to other dimensions of the component, the crack is viewed as being contained within an infinite body (Figure 2.4). If nominal stress σ_{nom} acts at a large distance from the crack, the stress intensity factor results in:

$$K_I = \sigma_{nom} \sqrt{\pi \cdot a} \tag{2.2}$$

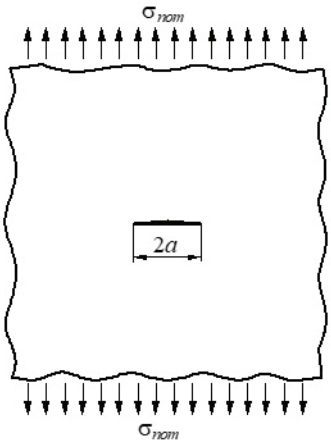

Figure 2.4: Middle crack of length 2a in an infinity body

For other crack configurations and external loadings, the stress intensity factor can be determined with the modification of Eq. (2.2) as follows:

$$K_I = \sigma_{nom} \sqrt{\pi \cdot a} \cdot f\left(\frac{a}{W}\right) \tag{2.3}$$

where $f(a/W)$ is the dimensionless function dependent on the crack length, geometry of the specimen and loading. For many standardized crack configurations and specimen geometries, the appropriate equations for the function $f(a/W)$ may be found in different standard procedures [2.14, 2.15], or in specialist literature related to fracture mechanics [2.5, 2.13].

2.1.3 Critical stress intensity factor

It can be seen from eqs. (2.2) and (2.3) that the stress intensity factor increases with the increase of crack length and loading. The critical stress intensity factor K_c refers to the condition when a crack propagates in an unstable manner. According to the eq. (2.3), K_c can be expressed mathematically as follows:

$$K_c = \sigma_c \sqrt{\pi \cdot a_c} \cdot f\left(\frac{a_c}{W}\right) \tag{2.4}$$

where σ_c is the applied nominal stress at crack instability and a_c is the crack length at instability. The critical stress intensity factor K_c is the main designing parameter when dimensioning cyclically loaded components using the damage-tolerant design criterion. Namely, it represents the critical value of the stress intensity factor K for a given load, as well as the crack length and geometry required to cause the fracture.

In general, K_c mainly depends on the material and geometry (thickness) of the cracked member (see Figure 2.5). It is evident that the highest K_c corresponds to the thinnest specimen where the plane stress appears in the critical cross section. As the thickness of the specimen increases the value of K_c, it decreases up to the thickness B_1 when plane strain conditions appear. The minimum value of K_c is called *fracture toughness* K_{Ic} and strictly refers to the plane strain conditions. The type of fracture (brittle or ductile) is also dependent on the loading conditions (plane stress or plane strain) as it can be seen in Figure 2.5.

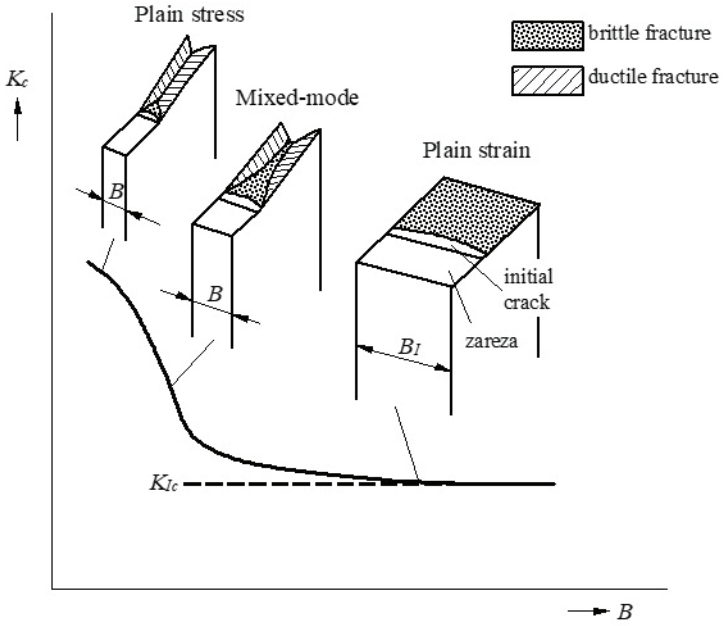

Figure 2.5: Critical stress intensity factor K_c for given material [2.2]

Fracture toughness K_{Ic} is a material parameter and is independent of the geometry of the cracked member. The values of K_{Ic} for different engineering materials can be found in specialist literature [2.16, 2.17]; the fields of K_{Ic}-values for typical engineering materials are also shown in Figure 2.6. From this diagram, it can be seen that the values of K_{Ic} decrease with the increase of the yield stress of the material. Therefore, high strength materials (i.e., high strength steels) are more sensitive to the occurrence of cracks if compared to low strength materials. Furthermore, fracture toughness is also dependent on temperature (see Figure 2.7). As the temperature decreases, the value of K_{Ic} decreases.

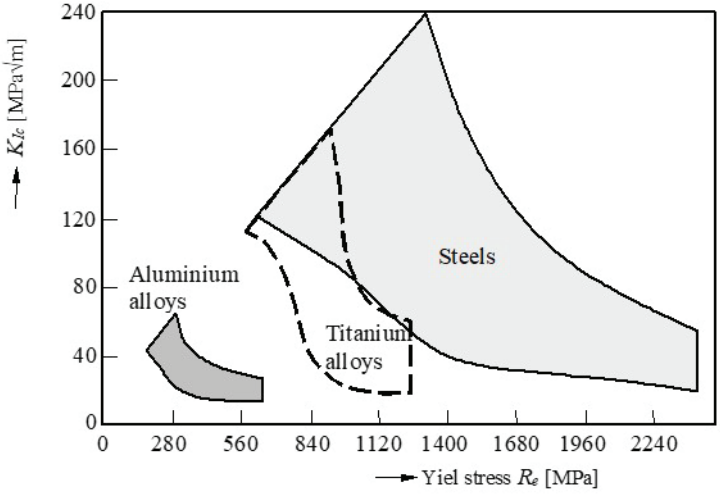

Figure 2.6: The fields of K_{Ic}-values for typical engineering materials [2.2]

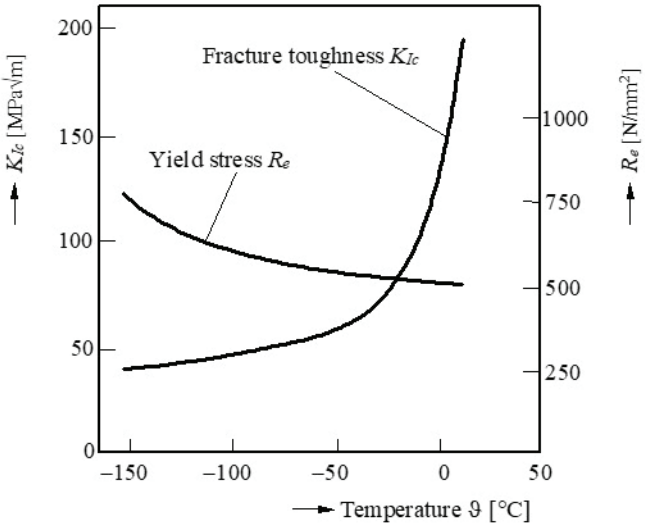

Figure 2.7: Variation of K_{Ic} with temperature for low alloy steel A533B [2.2]

2.1.4 Plastic zone and its limitation on LEFM

At the beginning of section 2.1 it is pointed out that the LEFM-theory is suitable when the material behavior is mainly elastic. However, this theory can also be used if the plastic zone size ahead of the crack is small in comparison to the other dimensions of the component. Figure 2.8 schematically shows the plastic zone in a tensile loaded cracked plate, where Irwin's formulation [2.18] is used to obtain the plastic zone size ahead of the crack. Assuming a circular shape of the plastic zone, the plastic zone size r_d results in:

$$r_d = \frac{1}{\pi} \cdot \left(\frac{K_I}{R_e}\right)^2 \quad \text{plane stress} \tag{2.5}$$

$$r_d = \frac{1}{3\pi} \cdot \left(\frac{K_I}{R_e}\right)^2 \quad \text{plane strain} \tag{2.6}$$

where K_I is the stress intensity factor for Mode-I loading and R_e is the yield stress of the material. As might be expected, the plastic zone size increases with the increase of the stress intensity factor, and is smaller for the same K_I in materials with a higher yield stress R_e. It is also evident that the plastic zone size r_d is three-times larger for plane stress conditions (thin plate in Figure 2.8*a*) in comparison to plane strain conditions (thick plate in Figure 2.8*b*).

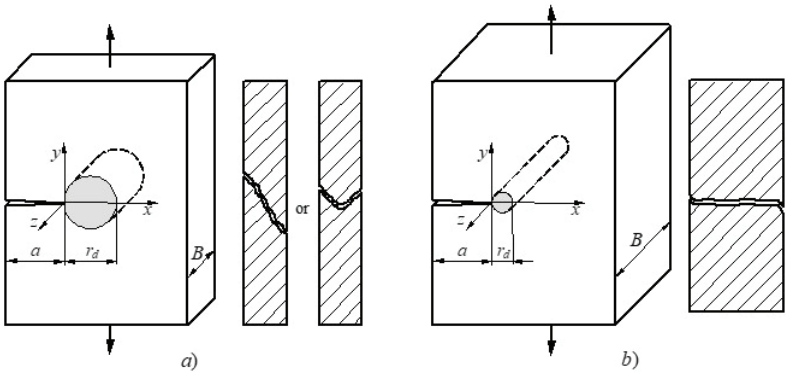

Figure 2.8: Plastic zone and fracture mode in tensile loaded cracked plate [2.2]
a) thin plate (plane stress), *b*) thick plate (plane strain)

Irwin's plastic zone formulation as described above provides only a rough estimation of the plastic zone ahead of the crack and actually represents the plastic zone in the plane of the crack ($\theta = 0$; see Figure 2.9). For an arbitrary value of θ, according to von Mises' plastification criterion, the plastic cone can be described as:

$$r(\theta) = \frac{1}{4\pi}\left(\frac{K_I}{R_e}\right)^2 \cdot \left[1 + \cos\theta + \frac{3}{2}\sin^2\theta\right] \qquad \text{plane stress} \tag{2.7}$$

$$r(\theta) = \frac{1}{4\pi}\left(\frac{K_I}{R_e}\right)^2 \cdot \left[(1-2\nu)^2(1+\cos\theta) + \frac{3}{2}\sin^2\theta\right] \quad \text{plane strain} \tag{2.8}$$

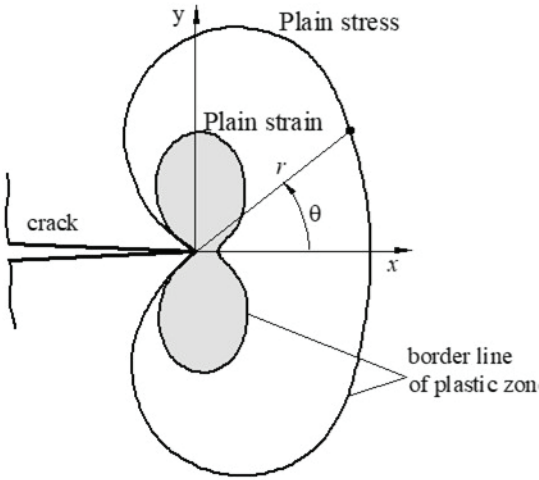

Figure 2.9: Plastic zone ahead of the crack considering von Mises' yield criterion

Plastic zone configuration in real engineering components is usually a combination of plane stress and plane strain conditions. In the case of a through-thickness crack (see Figure 2.10), the plane stress condition is dominant at the surface, while in the middle of component the plane strain occurs.

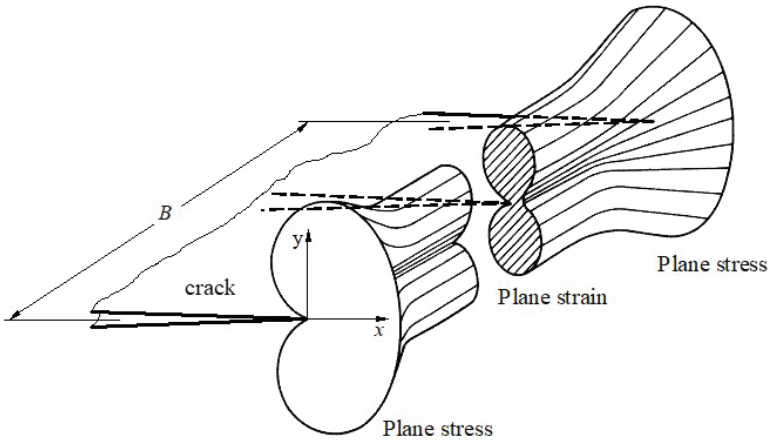

Figure 2.10: Plastic zone size at the tip of a through-thickness crack

Plasticity limitations on LEFM

If the plastic zone is sufficiently small, there will be an elastic region outside it where an elastic stress field occurs. The existence of such a region is necessary for the application of the LEFM theory. In that respect, it is necessary that the plastic zone is smaller in comparison to the distance from the crack tip to any boundary of the cracked member, such as (see Figure 2.11):

- crack length a,
- uncracked ligament $(W - a)$,
- height of cracked member H.

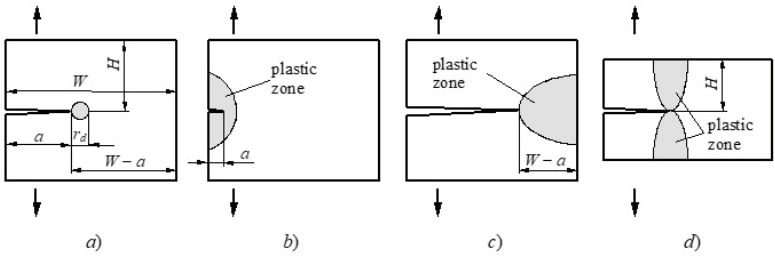

Figure 2.11: Different configurations of plastic zone in cracked members [2.1]

a) small plastic zone (the use of LEFM is valid), *b*) crack length is too small, *c*) ligament is too small, *d*) height of member is too small

Considering the explanation above, an overall limitation of LEFM may be expressed as follows:

$$a, (W - a), H \geq \frac{4}{\pi}\left(\frac{K_I}{R_e}\right)^2 \tag{2.9}$$

where K_I is the Mode-I stress intensity factor and R_e is the yield stress of the material.

2.2 Fatigue crack growth

In section 2.1, the theoretical background of LEFM related to monotonic loading is presented. The following questions can be answered with the use of this knowledge: (i) What is the highest permissible loading of a cracked component and, (ii) What is the critical crack length that still assures the safety operation of the component? However, machine elements and structural components are often exposed to cyclic loading which varies between a maximum (F_{max}) and a minimum (F_{min}) value. In such cases, the designer should first provide the answer to the following question: What is the operation time (number of loading cycles) needed for crack extension from the initial to the critical length, when final fracture can be expected to occur?

Figure 2.12 schematically shows the curve *a*–*N* (fatigue crack length versus number of loading cycles) for three different nominal stress ranges $\Delta\sigma_{nom}$ which are acting far away from the crack. All specimens have an equal initial crack length, a_0, and in each test, the minimum stress is almost zero. In the case of the highest nominal stress range $\Delta\sigma_{nom1}$, fatigue life until fracture is the shortest and the fracture appears at the shortest crack length. The opposite is valid for the lowest nominal stress range $\Delta\sigma_{nom3}$.

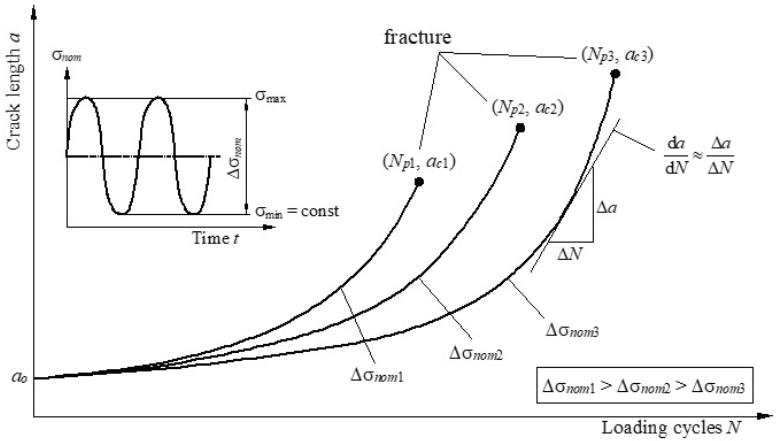

Figure 2.12: Fatigue crack length a versus the number of loading cycles N

From Figure 2.12, the fatigue crack growth rate, da/dN, can also be obtained for each nominal stress range $\Delta\sigma_{nom}$. Namely, the fatigue crack growth rate, da/dN, is simply the slope of the a versus N curve at a given crack length or given number of cycles. Guidelines for the experimental determination of the crack growth rate, da/dN, based on the data pairs (a, N), are described in the standard procedure ASTM E647 [2.19].

When analyzing cracked components using the damage-tolerant design criterion (see Chapter 1) under LEFM conditions, it is reasonable to combine the crack growth rate, da/dN, with the *stress intensity factor range* ΔK. For Mode-I loading, ΔK can be obtained as:

$$\Delta K = \Delta K_I = K_{\max} - K_{\min} \tag{2.10}$$

where K_{max} and K_{min} are the maximum and minimum stress intensity factor, respectively. Considering eq. (2.3), K_{max} and K_{min} can be determined as:

$$K_{\max} = \sigma_{\max}\sqrt{\pi \cdot a} \cdot f\left(\frac{a}{W}\right) \tag{2.11}$$

$$K_{\min} = \sigma_{\min}\sqrt{\pi \cdot a} \cdot f\left(\frac{a}{W}\right) \tag{2.12}$$

where σ_{max} and σ_{min} are the maximum and minimum nominal stress (see Figure 2.12), respectively, and $f(a/W)$ is the dimensionless function dependent on the crack length, geometry of the specimen and loading (see section 2.1.2). Based on K_{max} and K_{min}, the stress ratio R is defined as:

$$R = \frac{K_{min}}{K_{max}} \tag{2.13}$$

2.2.1 Basic formulation of fatigue crack growth

For a given material and set of test conditions, crack growth behavior can be described as the relationship between cyclic crack growth rate da/dN and the stress intensity factor range ΔK. The needed material parameters are usually determined by pulsating loading ($R = 0$ or approximately zero) due to the fact that during compression loading the crack is closed, and hence no stress intensity factor, K, can exist.

Figure 2.13 schematically shows the typical $\log(da/dN) - \log(\Delta K)$ plot, which may be divided into three major regions:

- region 1 (crack growth threshold),
- region 2 (stable crack growth),
- region 3 (instable crack growth).

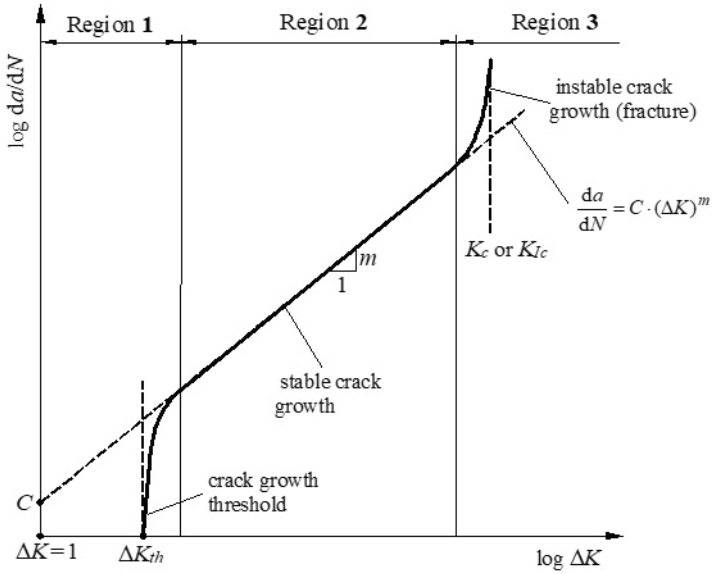

Figure 2.13: Schematic presentation of diagram log (da/dN) – log (ΔK)

Region 1

Region 1 indicates a threshold value, ΔK_{th}, below which there is no observable crack growth. This threshold occurs at crack growth rates on the order of 10^{-10} m/cycle or less, as defined in standard ASTM E647 [2.19]. Below ΔK_{th}, fatigue cracks are characterized as non-propagating cracks. The main parameters influencing the value of ΔK_{th} are the microstructure of the material, mean stress (or R-value), loading frequency, environment, etc. For many engineering materials, the appropriate values for ΔK_{th} can be found in specialist literature [2.2, 2.17, 2.20]. For steels with an ultimate tensile strength $R_m \leq 1400$ MPa and stress ratio $R = 0$, the ΔK_{th} value can be approximately determined as [2.21]:

$$\Delta K_{th} \approx \frac{3.48 \cdot 10^3}{\sqrt{\sigma_D}} \tag{2.14}$$

where σ_D is the fatigue limit of the material obtained by $R = 0$.

Region 2

Region 2 shows a linear relationship between log (da/dN) and log (ΔK), which can be expressed mathematically with the *Paris equation* [2.22]:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C \cdot (\Delta K)^m \tag{2.15}$$

where C and m are material parameters which can be determined experimentally. Graphically (see Figure 2.13), m is the slope of the line and C is the coefficient found by extending the straight line to the value $\Delta K = 1$ MPa $\sqrt{\text{m}}$. Microstructure and mean stress have less influence on fatigue crack growth behaviour in Region 2 in comparison to Region 1.

Once C and m are known, and ΔK is obtained using an appropriate analytical procedure or numerical approach (FEM), the number of loading cycles N required for a crack to propagate from initial (a_0) to the critical (a_c) length can be determined with the integration of eq. (2.15):

$$N = \frac{1}{C} \int_{a_0}^{a_c} \frac{\mathrm{d}a}{(\Delta K)^m} \tag{2.16}$$

Region 3

In Region 3, the fatigue crack growth rates are very high as they approach instability (final fracture of component). This region is controlled primarily by the critical stress intensity factor K_c or fracture toughness K_{Ic}, which in turns depends on the microstructure, mean stress, and environment.

For a given material and environment, the fatigue crack growth behavior as presented in Figure 2.13 is essentially very similar for different engineering components because the stress intensity factor range ΔK is the principal controlling factor in the fatigue crack growth process. This allows the fatigue crack growth rate da/dN versus ΔK data, obtained under constant amplitude conditions with simple specimen configurations, to be used in real design problems.

2.2.2 Effect of the stress ratio R on fatigue crack growth

Material parameters to describe fatigue crack growth (ΔK_{th}, C, m) taken from specialist literature are generally related to the stress ratio $R = 0$.

However, if different values of R appear, this fact should be considered when dimensioning structural components using the damage-tolerant design criterion. The general influence of the stress ratio R on crack growth is shown in Figure 2.14 where the R-value is presented as the principal parameter. The presented curves indicate that increasing the R-ratio, which means increasing the mean stress, has a tendency to increase the crack growth rates. It is evident that this influence is more significant in Regions 1 and 3, while in Region 2, the influence is smaller.

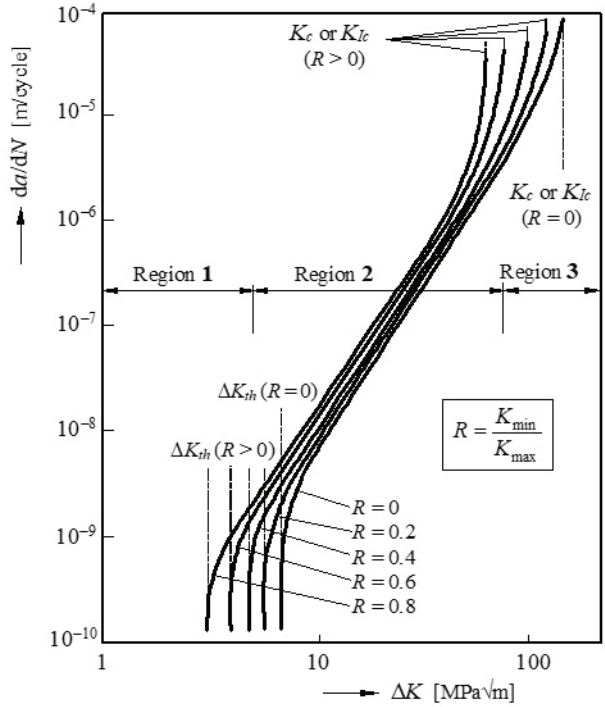

Figure 2.14: Diagram log (da/dN) – log (ΔK) for different R-values [2.2]

Region 1

Figure 2.14 indicates that the threshold stress intensity ranges ΔK_{th} decrease with the increase of the stress ratio R. The effect of the stress ratio R on in the threshold stress intensity ranges ΔK_{th} for steels can be expressed empirically as [2.23]:

$$\Delta K_{th(R>0)} \approx \Delta K_{th(R=0)}(1-R)^{\gamma} \tag{2.17}$$

where γ is the empirical constant (γ = 0.7 ... 1.0 for steels). The relationship between ΔK_{th} and R for other engineering materials is presented in Figure 2.15.

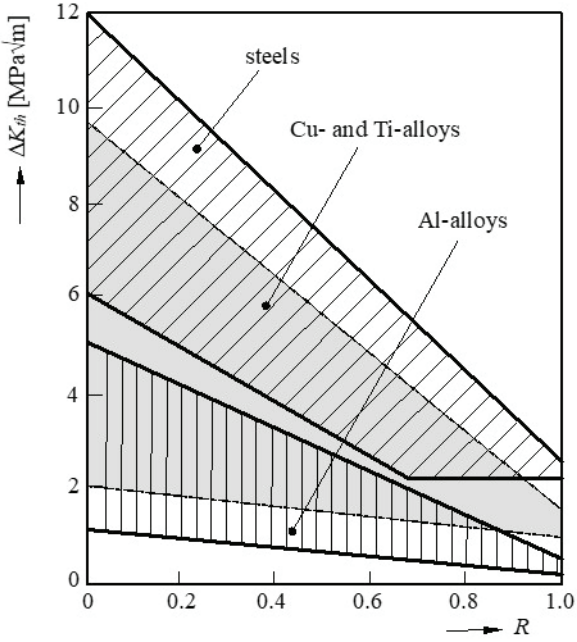

Figure 2.15: Dependence between ΔK_{th} and R for different engineering materials [2.24]

Regions 2 and 3

A commonly used equation depicting the mean stress effect (or R-ratio) on fatigue crack growth in Regions 2 and 3 is the Forman equation [2.25]:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = \frac{C_1 \cdot (\Delta K)^{m_1}}{(1-R) \cdot K_c - \Delta K} \tag{2.18}$$

where C_1 and m_1 are empirical material fatigue crack growth rate constants and K_c is the applicable critical stress intensity factor for a given material and thickness of component. The Forman equation is actually a

modification of the Paris equation that incorporates mean stress and Region 3 fatigue crack growth behavior.

2.3 Fatigue crack growth by multiaxial fatigue

The fatigue crack growth theory presented in the previous sections is related to uniaxial fatigue considering the Mode-I crack propagation. However, engineering components are often subjected to multiaxial fatigue, in which case combinations of loading modes I, II and III may appear at the same time. Such multiaxial loading results in *mixed-mode crack extension* where both, crack growth direction and crack growth rate are important. Many theories describing the mixed-mode crack growth exist in specialist literature. Some of them are discussed in the following sections.

2.3.1 Maximum tangential stress criterion

The *maximum tangential stress* (MTS) criterion was first proposed by Erdogan and Sih [2.26] and is generally suitable for mixed-mode I and II crack growth. This criterion is based on the elastic stress field around the crack tip as shown in Figure 2.16. Assuming that the material is isotropic and homogeneous, the elastic stress field in polar coordinates can be described as follows:

$$\sigma_r = \frac{K_I}{\sqrt{2\pi r}}\left(\frac{5}{4}\cos\frac{\theta}{2} - \frac{1}{4}\cos\frac{3\theta}{2}\right) + \frac{K_{II}}{\sqrt{2\pi r}}\left(-\frac{5}{4}\sin\frac{\theta}{2} + \frac{1}{4}\sin\frac{3\theta}{2}\right) \tag{2.19a}$$

$$\sigma_\theta = \frac{K_I}{\sqrt{2\pi r}}\left(\frac{3}{4}\cos\frac{\theta}{2} + \frac{1}{4}\cos\frac{3\theta}{2}\right) + \frac{K_{II}}{\sqrt{2\pi r}}\left(-\frac{3}{4}\sin\frac{\theta}{2} - \frac{3}{4}\sin\frac{3\theta}{2}\right) \tag{2.19b}$$

$$\tau_{r\theta} = \frac{K_I}{\sqrt{2\pi r}}\left(\frac{1}{4}\sin\frac{\theta}{2} + \frac{1}{4}\sin\frac{3\theta}{2}\right) + \frac{K_{II}}{\sqrt{2\pi r}}\left(\frac{1}{4}\cos\frac{\theta}{2} + \frac{3}{4}\cos\frac{3\theta}{2}\right) \tag{2.19c}$$

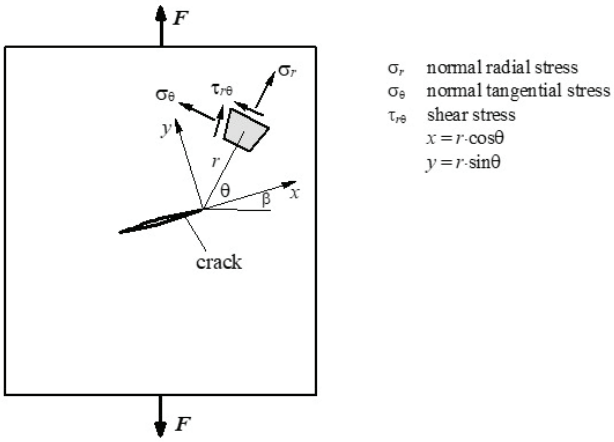

Figure 2.16: Elastic stress field around crack tip in polar coordinates for mixed-mode I and II crack growth

The MTS criterion considers the following assumptions:

- The crack propagates under the angle $\theta = \theta_0$ where the normal tangential stress is maximum ($\sigma_{\theta} = \sigma_{\theta max}$); it means that the direction of the crack extension is perpendicular to the direction of $\sigma_{\theta max}$.
- Instable crack growth (fracture) occurs at the moment when the maximum normal tangential stress $\sigma_{\theta max}$ reaches its critical value $\sigma_{\theta c}$, or, when the equivalent stress intensity factor K_E reaches its critical value K_{Ec}.

The crack extension angle θ_0 can be determined from eq. (2.19b) with a consideration of the following conditions:

$$\left.\frac{\partial \sigma_\theta}{\partial \theta}\right|_{\theta=\theta_0} = 0 \quad \text{and} \quad \left.\frac{\partial^2 \sigma_\theta}{\partial \theta^2}\right|_{\theta=\theta_0} < 0 \tag{2.20}$$

which leads to the following value of crack extension angle θ_0:

$$\theta_0 = \pm \arccos\left[\frac{3{K_{II}}^2 + \sqrt{{K_I}^4 + 8{K_I}^2{K_{II}}^2}}{{K_I}^2 + 9{K_{II}}^2}\right] \tag{2.21}$$

It can be seen from eq. (2.21) that the angle θ_0 can be positive or negative which is dependent on the stress intensity factor K_{II}. Here, the following rule can be applied: $\theta_0 < 0$ if $K_{II} > 0$ and $\theta_0 > 0$ if $K_{II} < 0$, respectively. It is also evident from eq. (2.21) that $\theta_0 = 0°$ for pure normal loading according to Mode I ($K_I \neq 0$, $K_{II} = 0$) and $\theta_0 \pm 70.5°$ for pure shear loading according to Mode II ($K_I = 0$, $K_{II} \neq 0$).

Once the crack extension angle θ_0 is known, the combined (equivalent) stress intensity factor K_E can be determined as a function of K_I and K_{II}:

$$K_E = \left[K_I \cos^2 \frac{\theta_0}{2} - \frac{3}{2} K_{II} \sin\theta_0\right] \cos\frac{\theta_0}{2} \tag{2.22}$$

The above equations are actually related to the stress intensity factors K_I and K_{II} in monotonic loading. When studying multiaxial fatigue crack growth where cyclic loading is applied, the stress intensity factor ranges $\Delta K_I = K_{I\max} - K_{I\min}$ and $\Delta K_{II} = K_{II\max} - K_{II\min}$ should be determined first. Based on their values, the equivalent stress intensity factor range ΔK_E can be obtained using eq. (2.22).

2.3.2 Minimum strain energy density criterion

The *minimum strain energy density* (SED) criterion was first proposed by Sih [2.27]. This criterion is often used for mixed-mode I and II crack growth, although it is also suitable if an additional Mode III appears [2.28]. Figure 2.17 shows the inclined crack inside the tensile loaded plate. The elastic stress field around the crack in Cartesian coordinates is:

$$\sigma_x = \frac{K_I}{\sqrt{2\pi r}} \cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) - \frac{K_{II}}{\sqrt{2\pi r}} \sin\frac{\theta}{2}\left(2 + \cos\frac{\theta}{2}\cos\frac{3\theta}{2}\right) \tag{2.23a}$$

$$\sigma_y = \frac{K_I}{\sqrt{2\pi r}} \cos\frac{\theta}{2}\left(1 + \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) + \frac{K_{II}}{\sqrt{2\pi r}} \sin\frac{\theta}{2}\cos\frac{\theta}{2}\cos\frac{3\theta}{2} \tag{2.23b}$$

$$\tau_{xy} = \frac{K_I}{\sqrt{2\pi r}} \sin\frac{\theta}{2}\cos\frac{\theta}{2}\cos\frac{3\theta}{2} + \frac{K_{II}}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) \tag{2.23c}$$

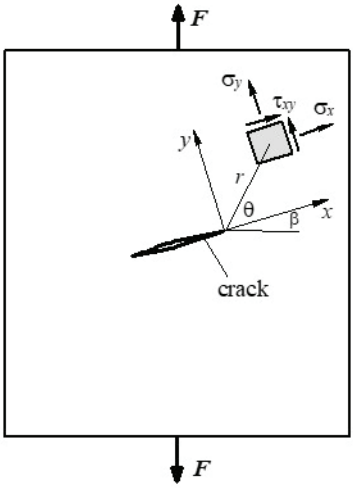

Figure 2.17: Elastic stress field around crack tip in Cartesian coordinates for mixed-mode I and II crack growth

Based on the stresses σ_x, σ_y nad τ_{xy}, the strain energy density w is defined as follows:

$$w = \frac{1+\nu}{2E}\left[\frac{\kappa+1}{4}\left(\sigma_x + \sigma_y\right)^2 - 2\left(\sigma_x\sigma_y - {\tau_{xy}}^2\right)\right] \tag{2.24}$$

where ν is Poisson's ratio, E is the modulus of elasticity and κ is the auxiliary parameter defined as:

$$\kappa = \frac{3-\nu}{1+\nu} \quad \text{plane stress} \tag{2.25a}$$

$$\kappa = 3 - 4\nu \quad \text{plane strain} \tag{2.25a}$$

With combining eqs. (2.23) and (2.24) it follows:

$$w = \frac{1}{r}\left(a_{11}{K_I}^2 + 2a_{12}K_I K_{II} + a_{22}{K_{II}}^2\right) \tag{2.26}$$

where a_{ij} are the auxiliary coefficients:

$$a_{11} = \frac{1+\nu}{8E}[(1+\cos\theta)(\kappa-\cos\theta)] \tag{2.27a}$$

$$a_{12} = \frac{1+\nu}{8E}\sin\theta\,[(2\cos\theta-\kappa+1)] \tag{2.27b}$$

$$a_{22} = \frac{1+\nu}{8E}[(\kappa+1)(1-\cos\theta)+(1+\cos\theta)(3\cos\theta-1)] \tag{2.27c}$$

The strain energy density w can also be written in the form:

$$w = \frac{S}{r} \tag{2.28}$$

where S is the coefficient of strain energy density:

$$S = a_{11}{K_I}^2 + 2a_{12}K_I K_{II} + a_{22}{K_{II}}^2 \tag{2.29}$$

When analyzing fatigue crack growth, the SED criterion considers the following assumptions:

- The crack propagates under the angle $\theta = \theta_0$ which corresponds to the minimum strain energy density ($w = w_{min}$; $S = S_{min}$).
- Instable crack growth (fracture) occurs at the moment when the minimum coefficient of strain energy density S_{min} reaches its critical value S_c, or, when the equivalent stress intensity factor K_E reaches its critical value K_{Ec}.

The crack extension angle θ_0 can be determined from eq. (2.29) with a consideration of the following conditions:

$$\left.\frac{dS}{d\theta}\right|_{\theta=\theta_0} = 0 \quad \text{and} \quad \left.\frac{d^2S}{d\theta}\right|_{\theta=\theta_0} > 0 \tag{2.30}$$

Once the crack extension angle θ_0 is obtained, the appropriate minimum coefficient of strain energy density $S_{\min}$ can be determined from eq. (2.29), considering that $\theta = \theta_0$ when determining the coefficients a_{11}, a_{12} and a_{22} according to eq. (2.27).

The combined (equivalent) stress intensity factor K_E can be obtained on the assumption that an equivalent coefficient of strain energy density $S_{E\min}$ for Mode I loading has a similar effect on crack growth as $S_{\min}$, which is generally valid for a combined Mode I and Mode II loading:

$$S_{E\min} = a_{11}(\theta = 0){K_E}^2 = \frac{(\kappa - 1)(1 + \nu)}{4E}{K_E}^2 \tag{2.31}$$

and with consideration of eq. (2.29):

$$K_E = \sqrt{\frac{4E\left(a_{11}{K_I}^2 + 2a_{12}K_I K_{II} + a_{22}{K_{II}}^2\right)}{(\kappa - 1)(1 + \nu)}} \tag{2.32}$$

In the given form, the above equations are valid for monotonic loading. When fatigue loading appears, the stress intensity factors K_I and K_{II} should by replaced with their ranges $\Delta K_I = K_{I\max} - K_{I\min}$ and $\Delta K_{II} = K_{II\max} - K_{II\min}$. Once ΔK_I and ΔK_{II} are known, there is no difficulty in obtaining the equivalent stress intensity factor range ΔK_E using eq. (2.32), which is then considered in the subsequent fatigue analysis.

2.3.3 Energy release rate criterion

This criterion is based on the energy release rate G during crack extension. For pure Mode I loading, the following relation between G and K_I is known:

$$G = \frac{1}{E}{K_I}^2 \qquad \text{plane stress} \tag{2.33a}$$

$$G = \frac{1 - \nu^2}{E}{K_I}^2 \quad \text{plane strain} \tag{2.33b}$$

where E is the modulus of elasticity and ν is Poisson' ratio. For mixed-mode I and II loading (see Figure 2.17), the relationship between the energy release rate G and the stress intensity factors K_I and K_{II} can be expressed with the following equations:

$$G = \frac{1}{E}\left({K_I}^2 + {K_{II}}^2\right) \qquad \text{plane stress} \tag{2.34a}$$

$$G = \frac{1-\nu^2}{E}\left({K_I}^2 + {K_{II}}^2\right) \quad \text{plane strain} \tag{2.34b}$$

The equivalent stress intensity factor K_E is obtained on the assumption that the energy release rate for Mode I loading, $G=f(K_E)$, has a similar effect on crack growth as the energy release rate $G=f(K_I, K_{II})$ for combined Mode I and Mode II loading. Thus, with combining eqs. (2.33) and (2.34), it follows:

$$K_E = \frac{1}{E}\sqrt{{K_I}^2 + {K_{II}}^2} \qquad \text{plane stress} \tag{2.35a}$$

$$K_E = \frac{1}{E}\sqrt{\left({K_I}^2 + {K_{II}}^2\right)(1-\nu^2)} \quad \text{plane strain} \tag{2.35b}$$

In this criterion, the crack extension direction is obtained according to the *virtual crack extension (VCE)* method, which was developed by Hellen [2.29]. As shown in Figure 2.18, different virtual crack extensions of length δr are assumed around the crack tip. For each virtual crack extension with the belonging angle θ, the equivalent stress intensity factor $K_E = f(K_I, K_{II})$ is calculated using eq. (2.35). In the next step, the crack is actually extended under the angle θ_0 which corresponds to the maximum value of K_E. Similar to the MTS and SED criterions, the ranges ΔK_I, ΔK_{II} and ΔK_E should be considered when analysing fatigue crack growth.

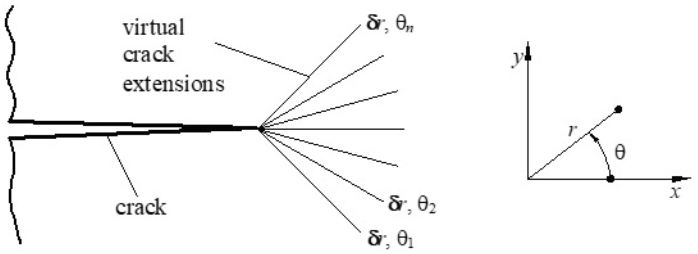

Figure 2.18: Virtual crack extensions in Polar coordinates

2.4 Short crack growth

Fatigue crack growth, as considered thus far, is based on the stress analysis of an isotropic and homogenous material. In effect, the microstructural features of the materials (i.e., crystal grains and their orientations) are assumed to occur on such a small scale that only the average behavior needs to be considered. However, if a crack is sufficiently small, it can interact with the microstructural features in ways that cause crack growth behavior to differ from what would otherwise be expected.

One of the most convenient presentations of short crack growth is the Kitagawa-Takahashi plot [2.30] as shown in Figure 2.19. In this diagram, the log-log relationship between the range of the cyclic stress $\Delta\sigma$ and crack length a is presented. The bolded curve represents the border between propagating and non-propagating cracks, while the dashed line represents the constant value of the threshold stress intensity factor range ΔK_{th} as a boundary condition to use the LEFM approach (the condition $\Delta K > \Delta K_{th}$ is needed for the subsequent crack growth). On the other hand, crack growth according to the LEFM can also be assumed when the crack length is large enough ($a > a_2$), or, when the stress range is in the elastic area ($\Delta\sigma < \Delta\sigma_2$). When higher stress range is applied, significant plastic deformation occurs around the crack tip and the crack growth behavior should be studied with a consideration of elastic-plastic fracture mechanics (EPFM). The second boundary condition in Figure 2.19 is the range of the fatigue limit $\Delta\sigma_D$. A crack will be classified as a propagating crack, if $\Delta\sigma > \Delta\sigma_D$. According to the ESTM E 647 [2.19], fatigue cracks are generally classified into three categories:

- microstructural short cracks ($a < a_1$),
- physical short cracks ($a_1 < a < a_2$),
- long cracks ($a > a_2$).

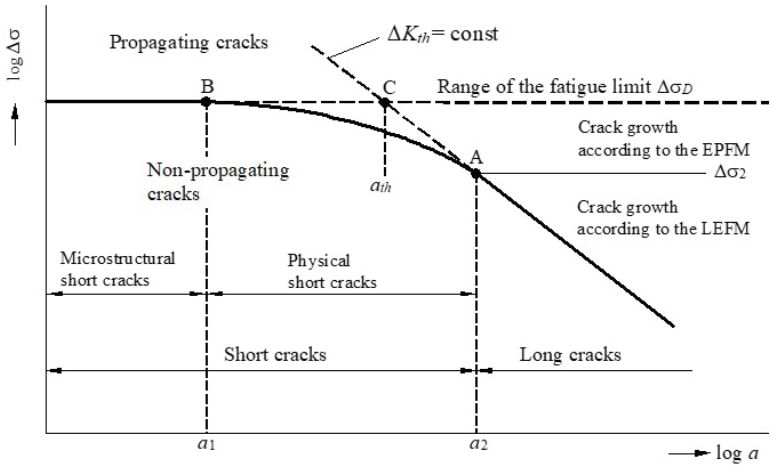

Figure 2.19: The Kitagawa-Takahashi plot of short crack growth [2.30]

For most engineering materials it is known that short cracks tend to grow faster than long cracks (see Figure 2.20), according to the diagram log (da/dN) – log (ΔK). As regards microstructural short cracks, all of their dimensions are similar to or smaller than the dimensions of greatest microstructural significance, such as the average crystal grain size. However, a physical short crack has one dimension that is large in comparison to the microstructure. Several theoretical models to describe short fatigue crack growth can be found in specialist literature. Some of them are presented in the continuation of this section.

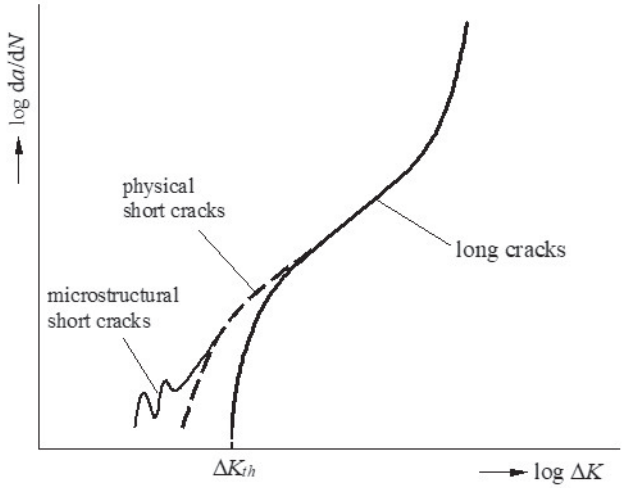

Figure 2.20: Diagram log (da/dN) – log (ΔK) for short and long cracks [2.1, 2.2]

A modified Paris model

The modified Paris model is based on the assumption that the linear part of Paris' curve log (da/dN) – log (ΔK) is simply extrapolated into Region 1 (see Figure 2.21). However, one must be careful when performing this operation, as extrapolation could result in either conservative or non-conservative behavior, depending on the short fatigue crack behavior observed.

Using this modification, the Paris equation with a consideration of material parameters C and b can also be used to describe fatigue crack growth in Region 1 [2.1]. In engineering applications, this approach is often used to describe physical short cracks and is not recommended for microstructural short cracks.

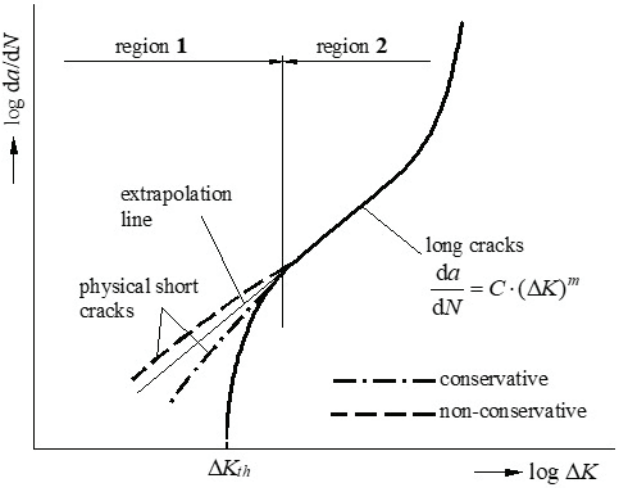

Figure 2.21: The principle of short crack growth using a modified Paris model

The Sähn model

The Sähn model [2.31] is suitable for describing fatigue behavior in physical short cracks without a consideration of the crack closure effect. The main influencing factor in this model is the coefficient of material microstructure d^*, which is a basic magnitude when determining the threshold stress intensity factor range for short cracks $\Delta K_{th}{}^*$ according the following equation:

$$\Delta K_{th}{}^* = \frac{\Delta K_{th}}{\sqrt{1+\frac{d^*}{a}}} \tag{2.36}$$

where ΔK_{th} is the threshold stress intensity factor range for long cracks and a is the crack length. If $a >> d^*$ (long cracks), the eq. (2.36) results in $\Delta K_{th}{}^* = \Delta K_{th}$.

In this model, the crack growth rate, da/dN, is defined as:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C\left(\Delta K\sqrt{1+\frac{d^*}{a}} - \Delta K_{th}\right)^m \tag{2.37}$$

The coefficient of material microstructure d^* is a material parameter and can be obtained experimentally. For steels with a yield stress R_e = (300 … 700) MPa, the material parameter d^* can be obtained empirically using the equation:

$$d^* = 105 - 0.0424 \cdot R_e \text{ [µm]} \tag{2.38}$$

Figure 2.22 shows the crack growth rate, da/dN, determined with eq. (2.37) and with a comparison of experimental results for non-alloyed steel C10.

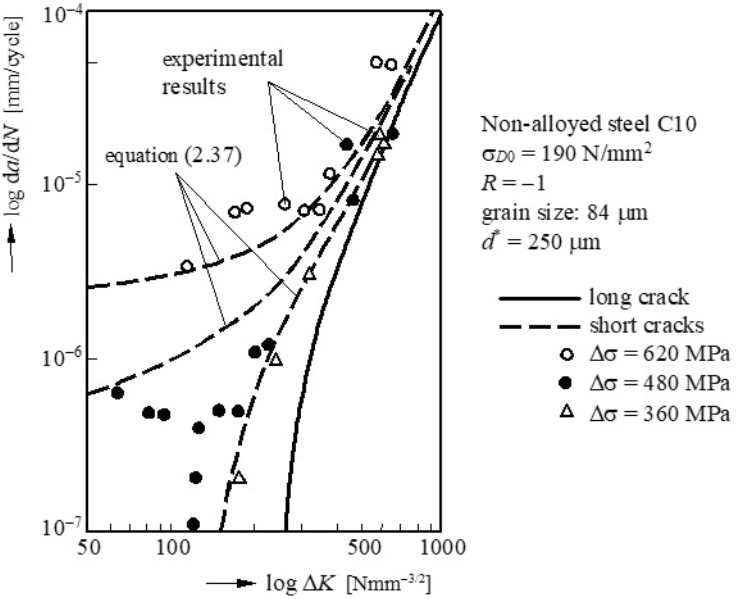

Figure 2.22: Diagram log (da/dN) – log (ΔK) for non-alloyed steel C10 [2.31]

The Haddad-Smith-Topper model

The Haddad-Smith-Topper model [2.32] is, similarly to the Sähn model, suitable for describing fatigue behavior in physical short cracks without a consideration of the crack closure effect. The model is based on the threshold crack length a_{th} between short and long cracks (see Figure 2.19). For steels, the threshold crack length a_{th} can be obtained using the following empirical equations:

$$a_{th} \approx \frac{13.6}{(R_e - 77.6)^2}\ [\text{mm}]\ ;\ \ R = -1 \tag{2.39a}$$

$$a_{th} \approx \frac{4.11}{(R_e - 72.6)^2}\ [\text{mm}]\ ;\ \ R = 0 \tag{2.39b}$$

where R_e is the yield stress of the material and R is the loading ratio. Once the threshold crack length a_{th} is known, the virtual crack length a^* is defined as:

$$a^* = a + a_{th} \tag{2.40}$$

where a is the actual crack length. Furthermore, the virtual stress intensity factor range, ΔK_{th}^*, is then:

$$\Delta K^* = \Delta\sigma\sqrt{\pi \cdot a^* \cdot} f\left(\frac{a^*}{W}\right) \tag{2.41}$$

where $f(a^*/W)$ is the dimensionless function (see section 2.1.2).

Once a^* and ΔK^* are known, further short crack growth behavior can be analyzed using the same equations as for long cracks, where a is replaced with a^* and ΔK is replaced with ΔK^*. A practical example of short crack growth using the Haddad-Smith-Topper model is shown in Figure 2.23.

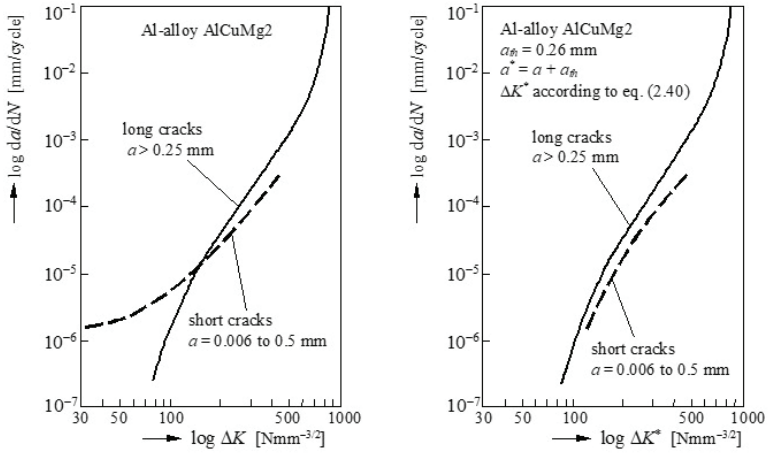

Figure 2.23: Diagram log (da/dN) – log (ΔK) for Al-alloy AlCuMg2 [2.32] *a*) real stress intensity factor range ΔK, *b*) virtual stress intensity factor range ΔK^*,

References

[2.1] Dowling NE, Mechanical Behavior of Materials, Prentice Hall, New Jersey, 1999.
[2.2] Stephens RI, Fatemi A, Stephens RR, Fuchs HO. Metal Fatigue in Engineering: John Wiley & Sons Inc, New York, 2001.
[2.3] Glodež S, Flašker J, Dimensioning on the service life, Scientific monograph, University of Maribor, Maribor, 2005.
[2.4] Ewalds HL, Wanhill RJH, Fracture Mechanics, Edward Arnold, 1991.
[2.5] Anderson TL, Fracture Mechanics: Fundamentals and Applications, CRC Press, Boca Raton, FL, 1995.
[2.6] Broek D, Elementary engineering fracture mechanics, Springer, 1982.
[2.7] Griffith AA, The Phenomena of Rupture and Flow in Solids, Philosophical Transactions, Series A, Vol. 221, 1920, pp. 163-198.
[2.8] Irwin GR, Fracture Dynamics, Fracturing of Metals, American Society for Metals, Cleveland, 1948, pp. 147-166.
[2.9] Orowan E, Fracture and Strength of Solids, Reports on Progress in Physics, Vol. XII, 1948, p. 185.
[2.10] Irwin GR, Onset of Fast Crack Propagation in High Strength Steel and Aluminium Alloys, Sagamore Research Conference Proceedings, Vol. 2, 1956, pp. 289-305.
[2.11] Irwin GR, Analysis of Stresses and Strains near the End of a Crack Traversing a Plate, Journal of Applied Mechanics, Vol. 24, 1957, pp. 361-364.
[2.12] Williams M.L., On the Stress Distribution at the Base of a Stationary Crack, Journal of Applied Mechanics, Vol. 24, 1957, pp. 109-114.
[2.13] Murakami Y, Stress Intensity Factors Handbook, Vol. 2, Perg. Press, Oxford, 1987.
[2.14] ASTM E399, Standard Test Method for Plane-Strain Fracture Toughness of Metallic Materials, Vol. 03.01, West Conshohocken, PA, 2000, p. 431.
[2.15] BS 7448, Fracture Mechanics Toughness Tests; Part 1: Method for Determination of K_{Ic}, critical CTOD and critical J values of Metallic Materials, British Standard, 1991.
[2.16] Damage Tolerant Design Handbook, A Compilation of Fracture and Crack Growth Data for High Strength Alloys, CINDAS/Purdue University, Lafeyette, 1994.
[2.17] Schwalbe KH, Bruchmechanik metallischer Werkstoffe, Carl Hanser Verlag, 1980.

[2.18] Irwin GR, Plastic Zone near a Crack and Fracture Toughness, Proceedings of the Seventh Sagamore Ordnance Materials Conference, Vol. 4, 1960, pp. 63-78.
[2.19] ASTM E647, Standard Test Method for Measurement of Fatigue Crack Growth Rates, Vol. 03.01, West Conshohocken, PA, 2000, pp. 591.
[2.20] Zahavi E, *Fatigue Design*, CRC Press, Boca Raton, FL, 1996.
[2.21] Radaj D, *Ermüdungsfestigkeit*, Springer Verlag, Heidelberg, 2003. (in German).
[2.22] Paris PC, Gomez MP, Anderson WE, A Rational Analytical Theory of Fatigue, *Trend Eng.*, Vol. 13, 1961, pp. 9-14.
[2.23] Klesnil M, Lukaš P, Influence of Strength and Stress History on Growth and Stabilization of Fatigue Cracks, *Engng. Fract. Mech.*, Vol. 4, 1972, pp. 77-92.
[2.24] Liaw PK, Leax TR, Logsdon WA, Near-Threshold Fatiogue Crack Growth Behaviour in Metals, *Acta Metallurgica* , Vol. 31, 1983, pp. 1581-1587.
[2.25] Forman RG, Kearney VE, Engle RM, Numerical Analysis of Crack Propagation in Cyclic-Loaded Structures, *Trans. ASME, J. Basic Eng.*, Vol. 89, 1967, pp. 459.
[2.26] Erdogan F, Sih GC, On the Crack Extension in Plates under Plane Loading and Transverse Shear, *Int. Journal of Basic Engineering*, 1963, pp. 519-525.
[2.27] Sih GC, Strain Energy density factor applied to mixed mode crack problems, *Int. Journal of Fracture*, Vol. 10, 1974, pp. 305-321.
[2.28] Sih GC, *Mechanics of fracture initiation and propagation*, Kluwer Academic Publisher, Dodrecht, 1990.
[2.29] Hellen TK, On the Method of Virtual Crack Extensions, *International Journal for Numerical Methods in Engineering*, Vol. 9, 1975, pp. 187-297.
[2.30] Kitagawa H., Takahashi S., Applicability of Fracture Mechanics to Very Small Cracks or the Cracks in Early Stage, *Proceedings of the Second International Conference on Mechanical Behaviour of Materials*, Metals Park, OH, 1976, p. 627.
[2.31] Sähn S, Festigkeitsverhalten von Bauteilen mit kleinen Rissen und Kerben bei zyklischer Belastung, *Konstruktion*, Vol. 43, 1991, pp. 9-16 (in German).
[2.32] El Haddad MH, Smith KH, Topper TH, Fatigue Crack Propagation of Short Cracks, *J. Engng. Mater. Technol.*, Vol. 101, 1979, pp. 42-46.

CHAPTER 3

FRACTURE BEHAVIOUR OF COUNTERWEIGHT BARS IN MOBILE CRANES

3.1 Introduction

Concerning the design of cyclically loaded engineering structures and components, the prediction of their service life is of great importance. There is a growing interest in many structures (such as mobile cranes) to reduce the weight of the structure. One way to achieve this goal is to reduce required cross sections using high strength steel. In some cases, the fatigue strength of steel can be increased using appropriate heat treatment of the treated machine part or structure. In [3.1] it is shown that the fatigue strength of AISI 4340 steel can be increased by up to 91 % by the ion nitriding process. A similar effect can be reached with case carburizing and surface hardening of such elements [3.2, 3.3]. However, such heat treatment cannot be used for large elements because of the limit magnitude of heating devices. In such cases, the required properties of steel are achieved through metallurgical purification and cooling directly after rolling. During a controlled rolling process, high toughness due to fine-grained structure is also achieved [3.4].

However, the complete process of fatigue failure may be divided into the following stages: (1) microcrack nucleation; (2) short crack growth; (3) long crack growth; and (4) occurrence of final failure. In engineering applications, the first two stages are usually referred to as the "*crack initiation period*", while long crack growth is referred to as the "*crack propagation period*" [3.5]. The complete service life can then be determined from the number of stress cycles N_i required for fatigue crack initiation, 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:

$$N = N_i + N_p \tag{3.1}$$

The service life calculation of a cyclically loaded component such as the counterweight bar of a mobile crane is based on the knowledge of stresses or deformations in critical cross sections, which is usually calculated by means of the finite element analysis (FEA) or measured using appropriate measuring instruments. The main parameters influencing fatigue life are the external loads and the strength behaviour of the material. Therefore, in order to perform such an analysis, the appropriate fatigue properties of the material should be known.

The strain-based approach to fatigue problems is widely used at present [3.6, 3.7]. The most common application of the strain-based approach is in fatigue of notched members. In a notched component subjected to cyclic external loads, the behaviour of material at the root of the notch is best considered in terms of strain. A reasonable expected fatigue life (number of stress cycles N_i), based on the nucleation or formation of small macrocracks, can then be determined iteratively using the Coffin-Manson equation:

$$\varepsilon_a = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\sigma_f'}{E} \cdot (2N_i)^b + \varepsilon_f' \cdot (2N_i)^c \tag{3.2}$$

where ε_a is the total strain amplitude, E is the modulus of elasticity, σ_f' is the fatigue strength coefficient, b is the fatigue strength exponent, ε_f' is the fatigue ductility coefficient and c is the fatigue ductility exponent. Strain-life fatigue properties σ_f', b, ε_f' and c, which are often referred to as "*low-cycle fatigue properties*", are obtained experimentally according to the ASTM E 606 standard [3.8]. When the total strain amplitude ε_a of a specific machine part or structure is known (ε_a can be measured experimentally or determined numerically), the number of stress cycles N_i can be calculated iteratively using eq. (3.2).

The initiation phase of fatigue life in a virgin material is often assumed to constitute the growth of short cracks up to the size a_{th}, which is the transition length of short cracks into long cracks and may be estimated as [3.9]

$$a_{th} \approx \frac{1}{\pi} + \left(\frac{\Delta K_{th}}{\Delta\sigma_{FL}}\right)^2 \tag{3.3}$$

where ΔK_{th} is the threshold stress intensity range and $\Delta\sigma_{FL}$ is the fatigue limit of the material. However, a wider range of values has been selected for a_{th}, usually between 0.1 and 1 mm for steels where the high strength steels take the smallest values [3.10].

In the presented analysis, the simple LEFM theory [3.11] is used to describe fatigue crack growth from the initial (a_{th}) to the critical (a_{cr}) crack length. The appropriate number of stress cycles N_p with regard to eq. (3.1) is therefore:

$$\int_0^{N_p} dN = \frac{1}{C} \cdot \int_{a_{th}}^{a_{cr}} \frac{da}{\Delta K^m} \tag{3.4}$$

where ΔK is the stress intensity range ($\Delta K = K_{max} - K_{min}$), which is usually determined numerically using appropriate numerical code. In eq. (3.4), C and m are the material parameters, which can be determined experimentally, usually by means of three-point bending specimens according to the ASTM E 647 standard [3.12].

To describe the fatigue and fracture behaviour of structural elements such as counterweight bars of mobile cranes, the low cycle fatigue parameters σ_f', ε_f', b and c (for fatigue crack initiation) and the material parameters C and m (for fatigue crack growth) of the treated material should be determined first. Furthermore, a fatigue assessment of the treated component can be made.

3.2 Material properties of high strength steel S1100Q

In the present study, it is assumed that the treated structural element (counterweight bar of a mobile crane) is made of high strength steel S1100Q. Table 1 shows the chemical composition of this material, which was supplied in the form of hot-rolled plates. The test specimens were cut out of the plate in the rolling direction and machined to the final shape, as described in sections 3.2.1 and 3.2.2. The microstructure of the material is lower needle bainite with extremely fine needles (see Figure 3.1). Because of the different orientations of crystal growth, it can be assumed that the average mechanical properties can be used in all loading directions.

Table 3.1: Chemical composition of high strength steel S1100Q

C	Si	Mn	P	S	Cr	Ni
0.18 %	0.20 %	0.83 %	0.007 %	0.003 %	0.56 %	1.88 %
Mo	V	Cu	Al	Nb	N	B
0.56 %	0.057 %	0.01 %	0.61 %	0.017 %	0.006 %	0.002 %

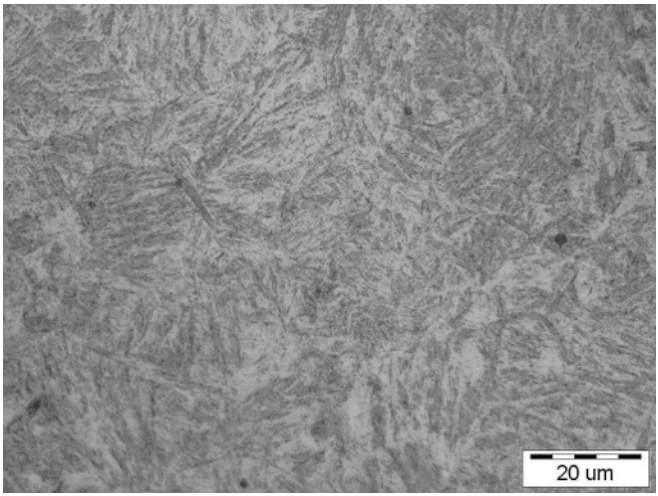

Figure 3.1: The microstructure of high strength steel S1100Q

3.2.1 Low-cycle fatigue parameters of high strength steel S1100Q

For determining low-cycle fatigue parameters, test specimens according to the ASTM E 606 standard have been used (see Figure 3.2). Before fatigue tests, monotonic tensile testing has been performed using the same specimen. Based on the monotonic test, the ultimate tensile strength R_m = 1450 MPa, the yield stress R_e = 1148 MPa, and the modulus of elasticity E = 194889 MPa, have been recorded.

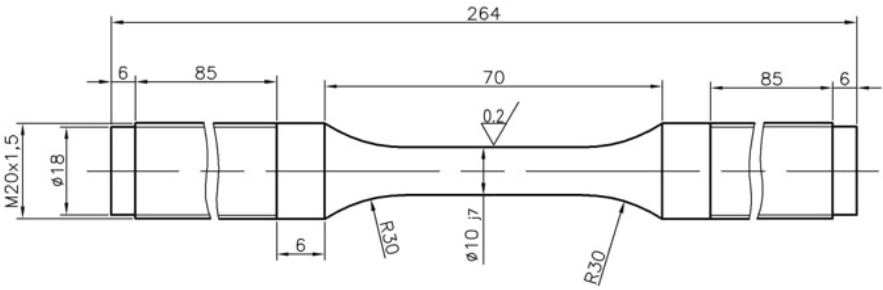

Figure 3.2: Test specimen according to the ASTM E 606 standard

The low-cycle fatigue tests were carried out in the strain-controlled regime on a servo-hydraulic fatigue machine, the Instron 1255. The loading waveform was triangular with loading ratio $R = -1$. The specimen temperature was 20°C and was manually checked during the test procedure. The loading frequency was higher for specimens with a lower deformation amplitude, as the energy generated in each cycle is lower. The low-cycle fatigue parameters have been determined using the results of 8 specimens, where specimen separation was chosen as the failure criterion.

Figure 3.3 shows the cyclic stress-strain curve, which is constructed on the basis of 8 stable hysteresis loops describing the fatigue behaviour of specimens loaded on a different strain level. From Figure 3.3, the cyclic modulus of elasticity E' = 183443 MPa and the cyclic yield stress R_e' = 875 MPa, are recorded.

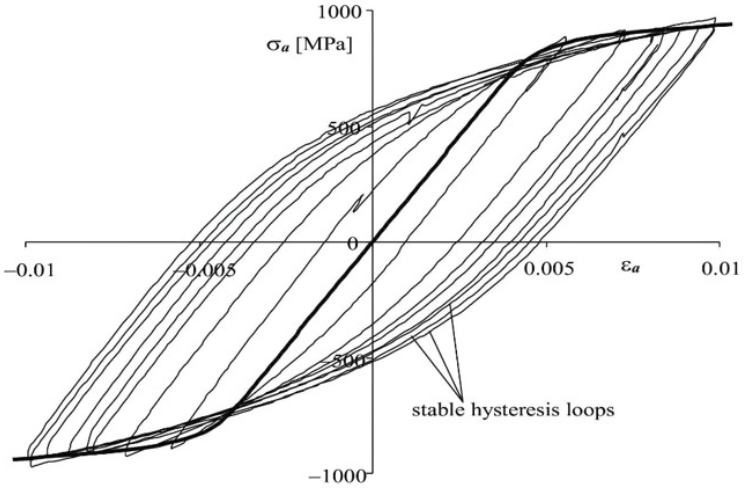

Figure 3.3: The cyclic stress-strain curve

The plot of the true stress amplitude, σ_a, versus the true plastic-strain amplitude, $\Delta\varepsilon_p/2$, in log-log coordinates results in a linear curve (see Fig. 3.4) represented by the power function

$$\sigma = K' \left(\frac{\Delta\varepsilon_p}{2}\right)^{n'} \tag{3.5}$$

where K' is the cyclic strength coefficient and n' is the cyclic strain-hardening exponent. It is evident from Figure 3.2 that K' = 1280 MPa (stress intercept at $\Delta\varepsilon_p/2 = 1$) and n' = 0,0595 (slope of the line).

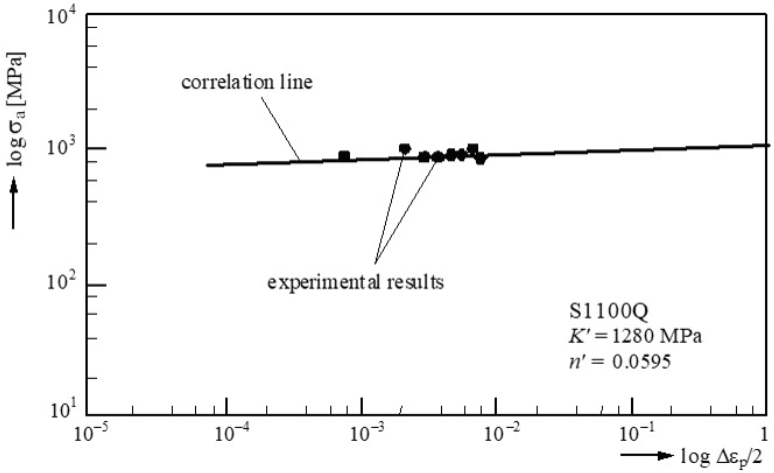

Figure 3.4: True stress amplitude versus true plastic-strain amplitude behaviour

Figure 3.5 shows the strain-life fatigue curves plotted on log-log scales, where N is the number of cycles to failure for each tested specimen. If the magnitudes in Figure 3.5 are compared with the theoretical ones in the ASTM E 606 standard [3.8], the low-cycle fatigue parameters for the high strength steel S1100Q result in:

- fatigue strength coefficient: $\sigma_f' = 2076$ MPa
- fatigue strength exponent: $b = -0.0997$
- fatigue ductility coefficient: $\varepsilon_f' = 9.93$
- fatigue ductility exponent: $c = -0.978$

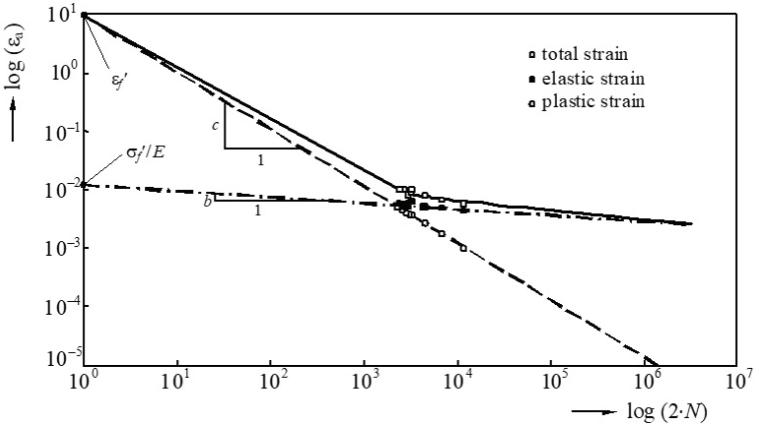

Figure 3.5: Strain-life curves of high strength steel S1100Q

3.2.2 Fracture mechanics parameters of high strength steel S1100Q

The fracture mechanics parameters C and m have been determined according to the standardised procedure ASTM E 647 [3.12] using a three-point bending specimen shown in Figure 3.6.

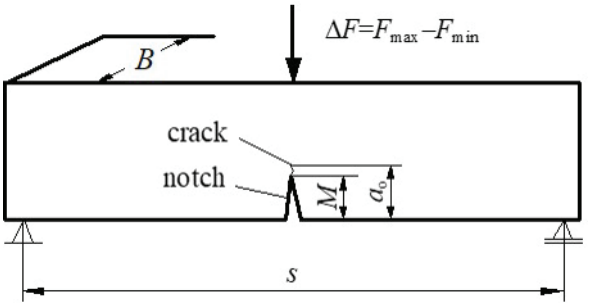

Figure 3.6: Three-point bending specimen

For the three-point bending specimen, the stress intensity range can be expressed as a function of its geometry and crack length [3.13]

$$\Delta K = \frac{\Delta F \cdot s}{B \cdot W^{3/2}} \cdot Y \tag{3.6}$$

where Y=f(a/W) is a shape function as defined in [3.13]. Using the standard procedure ASTM E 647, the test specimens are subjected to cyclic loading and crack propagation is measured in relation to the number of loading cycles N. In this way, the experimental relationship between a and N is determined and approximated with the appropriate function a=f(N). The crack growth rate da/dN then follows from:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = \frac{\mathrm{d}}{\mathrm{d}N}[f(N)] \tag{3.7}$$

Assuming that the relationship ΔK=f(a) is known, the diagram log(da/dN) −log(ΔK) can then be easily constructed. Taking into account the linear dependence between log(da/dN) and log(ΔK), the crack growth rate can be defined in the following linearized form:

$$\log\left(\frac{\mathrm{d}a}{\mathrm{d}N}\right) = C_1 + C_2 \cdot \log\left(\Delta K\right) \tag{3.8}$$

where C_1 and C_2 are the constants of the linear equation. Using substitution log(C_3)=C_1 it follows:

$$\log\left(\frac{\mathrm{d}a}{\mathrm{d}N}\right) = \log\left[C_3 \cdot (\Delta K)^{C_2}\right] \tag{3.9}$$

which finally leads to:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C_3 \cdot (\Delta K)^{C_2} \tag{3.10}$$

Comparing equations (3.4) and (3.10), the fracture mechanics parameters C and m result in:

$$C = C_3$$
$$m = C_2$$

The crack growth test was carried out on a servo-hydraulic fatigue machine, the Instron 1255, at a loading ratio R = 0.1. During the test, crack lengths have been recorded at certain load cycle intervals, which resulted in an appropriate relationship between the crack length a and the number of stress cycles N, and consequently between the stress intensity range ΔK and crack growth rate da/dN (see the diagram logΔK–log da/dN in Figure 3.7).

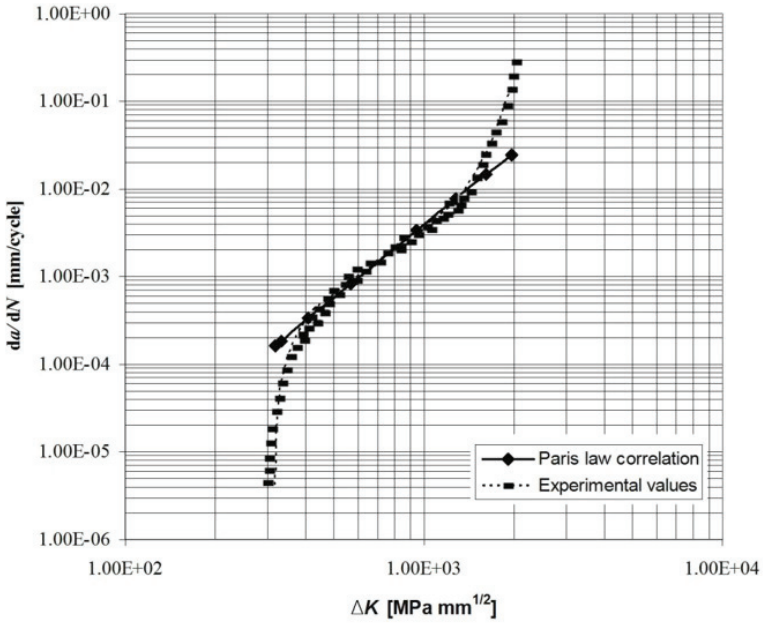

Figure 3.7: The diagram logΔK – log(da/dN)

Following the mathematical procedure as described above and the theoretical background as described in the ASTM E 647, the parameters C and m and the threshold stress intensity range ΔK_{th} result in:

$C = 2.02 \cdot 10^{-11}$ mm/(cycl·MPa$\sqrt{}$mm)
$m = 2{,}761$
$\Delta K_{th} = 315$ MPa·$\sqrt{}$mm

3.3 Fatigue and fracture analysis of counterweight bars in mobile cranes made of high strength steel S1100Q

Crawler cranes (Figure 3.8) are devices in which a counter weight is usually supported by a steel chain, which consists of a series of highly loaded bars (usually made of high-strength steels like S1100Q [3.14, 3.15]) with lug joints at their ends. Based on the material parameters determined in Chapter 3.2, the fatigue assessment of such steel bars is analysed using experimental testing and computational analysis as described in the following sections.

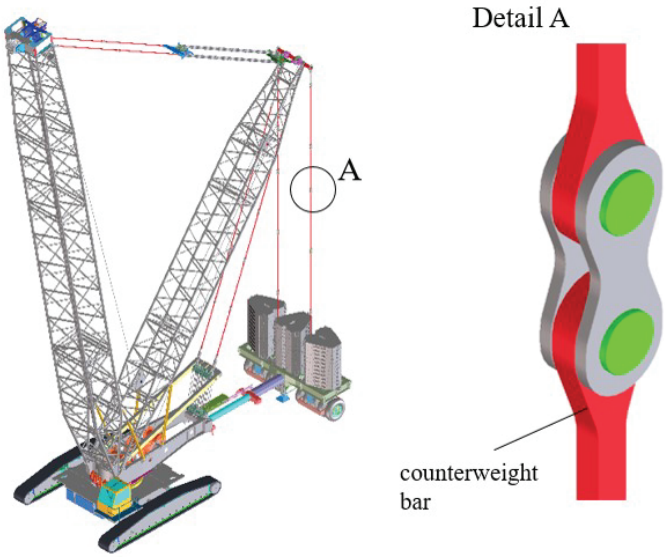

Figure 3.8: Schematic presentation of a crawler crane

3.3.1 Experimental testing

The fatigue tests were carried out in a specially designed testing machine made of two basic rigid plates, which are connected with a central lattice (Figure 3.9*a*). The load is applied using the oil pressure pumped into hydraulic cylinders, which are mounted on the base plate. The oil pressure provides a simple means of measuring the force applied. Changing the direction of the load is done by reversing the oil flow by electrical command. The machine enables testing four bars simultaneously, with a 1000 kN maximum tensile force in each bar. Actual stresses were controlled by means of the oil pressure and checked using strain gauges. The loading machine has certain disadvantages, like the need for large forces to achieve the necessary stresses and difficulty in providing high testing frequency. The testing bar (Figure 3.9*b*) has a rectangular cross-section (30 × 50 mm). Each side of the bar consists of a head with a hole for a pin. The pin/hole used in the experiments complied with a tolerance of H7/g6, with a measured fit clearance of approximately 25 μm. The fabrication of approximately 6-meter long bars was carried out under normal production procedure (gas cutting). Grinding was done on gas-cut surfaces. The bars

were loaded in tension so that the nominal applied stress was controlled in the critical cross-section.

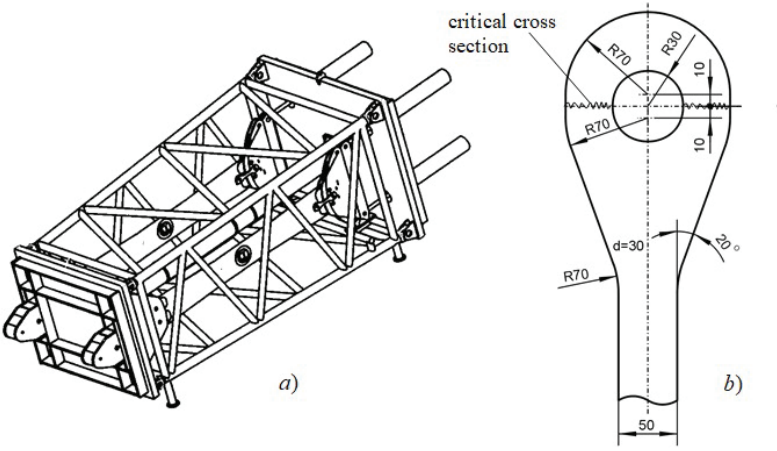

Figure 3.9: Fatigue testing machine (*a*) and testing bar (*b*)

3.3.2 Computational analysis

The computational analysis was performed in three steps. First, a stress and deformation field in the critical cross-section of the bar was determined numerically using the FEM program code Abaqus [3.16]. The FE model shown in Figure 3.10*a* and the loading pattern shown in Figure 3.10*b* were used in the computational analysis. Here, the loading pattern in Figure 3.10*b* is the same as was used in experimental testing, described previously. However, a zero clearance fit was assumed in the numerical model, which means that no initial stress was induced because of the pin itself. The stress field in the lug joint was determined numerically, using contact boundary conditions between the pin and the hole surface.

In the next step, fatigue analysis was performed using the FE-safe program code [3.17]. The fatigue analysis was based on the strain-life method (ε-*N*), where the Coffin-Manson relationship with a Morrow mean stress correction was used to determine the number of stress cycles N_i required for fatigue crack initiation in a critical cross-section of the treated bar connection [3.18]:

$$\frac{\Delta\varepsilon}{2} = \frac{(\sigma_f' - \sigma_m)}{E} \cdot (2N_i)^b + \varepsilon_f' \cdot (2N_i)^c \tag{3.11}$$

where $\Delta\varepsilon$ is the true strain range, σ_m is the mean stress and E, σ_f', ε_f', b and c are material parameters described in sections 3.1 and 3.2.

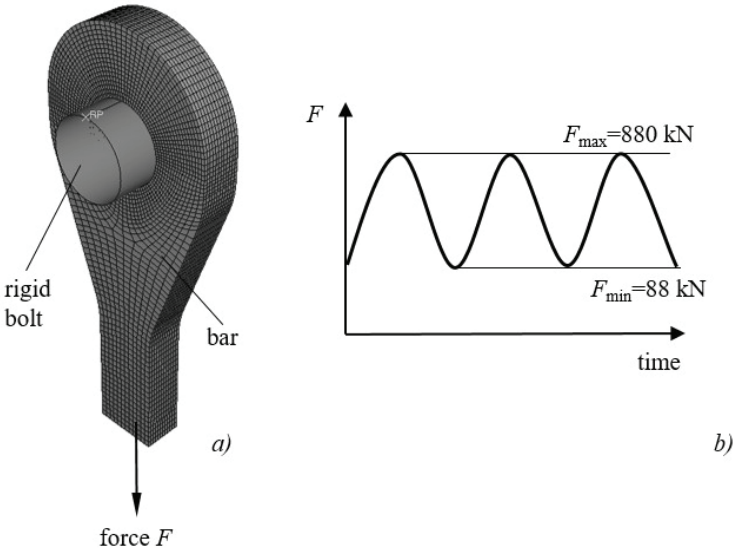

Figure 3.10: Numerical model (*a*) and loading pattern (*b*)

The third step represents the numerical analysis of fatigue crack growth. The numerical model consists of the geometry with a modeled crack in the region where a crack was being observed in experimental test bars. The model includes symmetry about central longitudinal and traverse vertical planes, to lower the computation time and reduce the size of the model (see Figure 3.11). The initial crack is modeled as a quarter elliptical edge crack with main ellipse axis c along the hole surface and a along the face of the lug. The initial crack length $a = c = 0.2$ mm has been determined using eq. (3.3) with consideration of previously determined threshold stress intensity range $\Delta K_{th} = 315$ MPa·√mm (see section 3.2.2) and fatigue limit $\sigma_{FL} = 390$ MPa as presented in literature [3.19, 3.20]. The region around the crack uses a highly refined mesh in comparison to the rest of the model. This region consists of a special block, which is tied in the base model, making it possible to have a coarser mesh in the remaining part of the model. The mesh consists of hexahedral brick elements using a structured mesh

technique in the crack region block. The material is defined as isotropic linear elastic with elastic modulus of E = 183442 MPa and Poisson's ratio ν = 0.3. Load is applied by means of the applied pressure p at the bottom end surface according to Figure 11, to obtain a load of 880 kN at the lug with a load ratio R=0.1. The whole model is supported by the fixed rigid pin in the hole with defined contact between the main part and its outer surface.

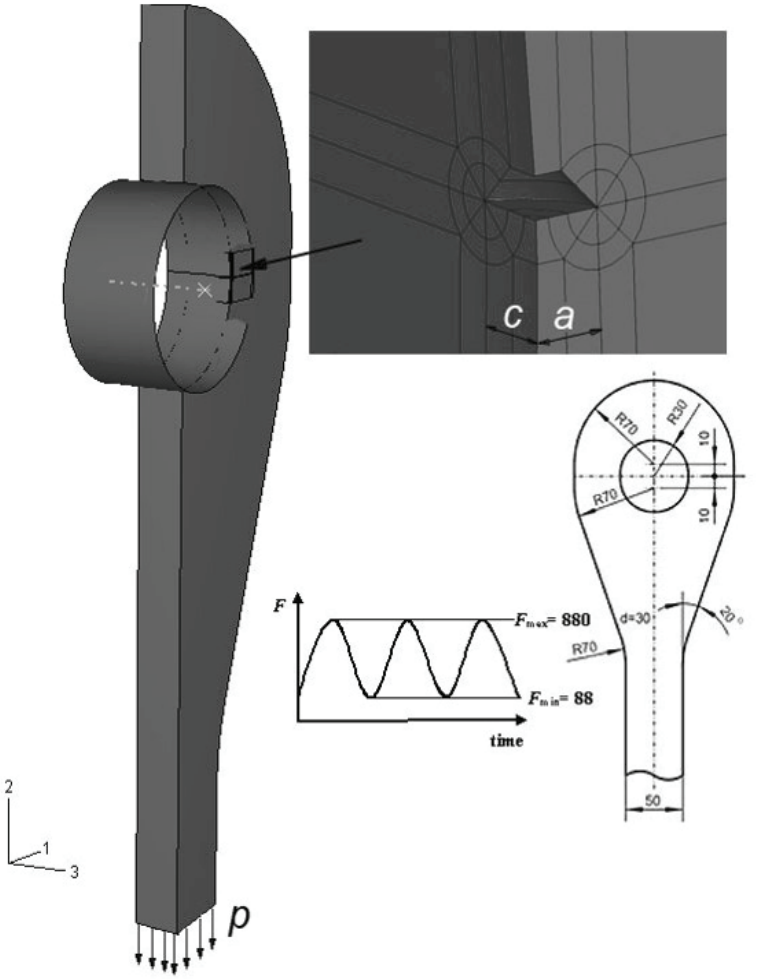

Figure 3.11: Numerical model for fatigue crack growth

In a numerical simulation of crack growth, stress intensity factors were determined for different sizes of the crack. The obtained results were then used for derivation of the correction function to be used with the standard model, which assumes the plate with quarter elliptical edge crack, loaded in tension and bending [3.21]. As the stress field in the main part of the lug end can be approximated with tension and bending stress, stress intensity factors have been estimated over the cracks front region using an analytical model given in [3.21]. For each numerical solution, the crack length was scaled to obtain a reasonable agreement of stress intensity factors over the crack front. Using obtained scale factors, the correction functions over the whole crack growth domain were obtained. The scale function f_a is proposed for the main axis a:

$$f_a = 1.458 \cdot 10^{-6} \cdot a^3 - 2.050 \cdot 10^{-4} \cdot a^2 - 1.428 \cdot 10^{-2} \cdot a + 1.125 \tag{3.12}$$

and the scale function f_c for the main axis c:

$$f_c = -4.080 \cdot 10^{-4} \cdot c^2 - 3.0350 \cdot 10^{-3} \cdot c + 9.732 \cdot 10^{-1} \tag{3.13}$$

The scale functions f_a in f_c have then been used to determine the stress intensity range ΔK using analytical procedure as described in [3.21]. The crack growth has then been analyzed with eq. (3.4) using material parameters C and m as described in section 3.2.2.

On the basis of stress intensity factors for appropriate crack lengths along axis a and c, the crack extension in the next step is given for 5 load cycles. Numerical analysis has shown that for the initial crack length $a = c = 0.2$ mm, the stress intensity factor exceeds its threshold value $\Delta K_{th} = 315$ MPa·√mm. The numerical analysis was stopped when the stress intensity factor reached its critical value $K_{Ic} = 2100$ MPa·√mm. The numerical results are shown in Figure 3.12. It is evident that the critical crack lengths are:

$a_c = 2.03$ mm
$c_c = 2.3$ mm

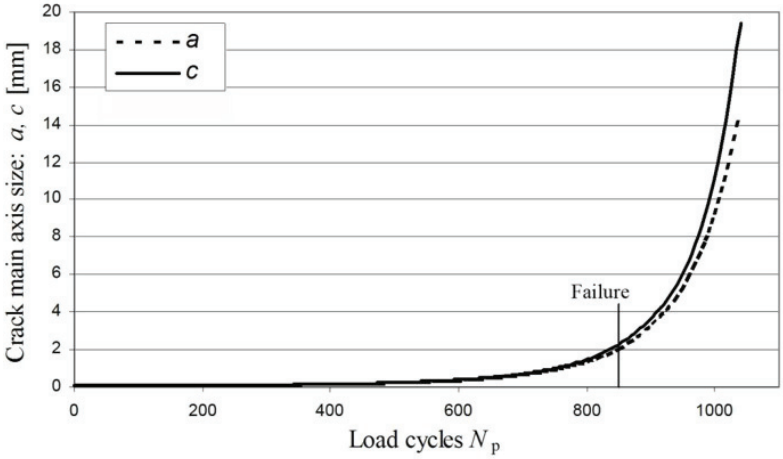

Figure 3.12: Numerical analysis of fatigue crack growth

3.3.3 Comparison between experimental and computational results

Figure 3.13*a* shows the fatigue breakage of the tested bar. The fatigue crack was initiated at the edge of the hole, which can be seen from Fig. 3.13*b*. The initial crack then propagates until the final fracture in the critical cross-section. Figure 14 shows the numerically determined crack path in the critical cross-section of the bar.

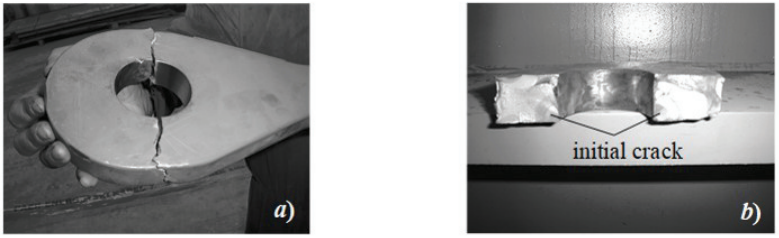

Figure 3.13: Fatigue breakage of testing bar (*a*) and example of fracture surface (*b*)

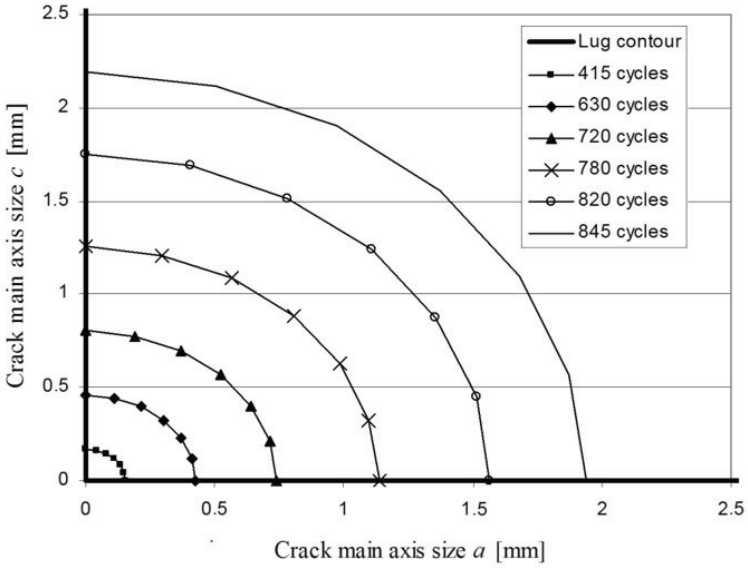

Figure 3.14: Numerically determined crack path in the critical cross-section of the bar

The number of stress cycles required for fatigue crack initiation N_i and fatigue crack propagation N_p determined using the presented computational model is shown in Table 2. It is evident that the comparison of experimental and computational results shows a reasonable agreement.

Table 3.2: The number of stress cycles N required for the final failure of the bar

Experimental results				Computational results		
Test 1	Test 2	Test 3	Test 4	Crack initiation N_i	Crack propagation N_p	Total N
38029	26727	24795	29036	28705	845	29550

3.4 Concluding remarks

The experimental determination of fatigue and fracture mechanics parameters of high strength steel S1100Q is presented in this study. The low-cycle fatigue parameters σ_f' = 2076 MPa, b = –0,0997, ε_f' = 9,93 and c = –0,978 are determined following the standard procedure ASTM E 606. On the basis of this parameter the fatigue initiation period (the number of stress cycles N_i required for fatigue crack initiation) can be determined using the Coffin-Manson relationship as described in section 3.1. Furthermore, the complete procedure for determining the Paris coefficients $C = 2.02 \cdot 10^{-11}$ mm/(cycl·MPa$\sqrt{}$mm) and m = 2,761 of the treated material is presented in section 3.2. Based on these parameters, the crack propagation period (the number of stress cycles N_p required for a crack to propagate from the initial to the critical crack length) can be determined using the Paris equation.

The proposed computational model is used to determine the service life of a counterweight bar with a lug joint at its end, and has been experimentally tested. Both procedures have shown that the connection hole in the bar end is the most critical location for crack initiation and final failure. The results presented in Table 2 represent useful information for the designer about the fatigue assessment of a counterweight bar as an important supporting part for cranes.

References

[3.1] Sule YS, Kahraman S, Erdinc K. Effect of the ion nitriding surface hardening process on fatigue behaviour of AISI 4340 steel, *Materials Characterization* 2008; 59: 351-358.

[3.2] Bensely A, Senthilkumar D, Mohan Lal D, Nagarajan G, Rajadurai A, Effect of cryogenic treatment on tensile behaviour of case carburized steel-815M17, *Materials Characterization* 2007; 58: 485-491.

[3.3] Glodez S, Sraml M, Kramberger J, A computational model for determination of service life of gears, International Journal of Fatigue 2002; 24: 1013-1020.

[3.4] *Design with Weldox and Hardox*, SSAB Oxelösund, 1991.

[3.5] Podrug P, Jelaska J, Glodež S. Influence of different load models on gear crack path shapes and fatigue lives, *Fatigue Fract. Eng. Mater. Struct*. 2008; 31: 327-339.

[3.6] Stephens RI, Fatemi A, Stephens RR, Fuchs HO. *Metal Fatigue in Engineering*: John Wiley & Sons Inc, New York, 2001.

[3.7] Draper J. *Modern Metal Fatigue Analysis*: EMAS Publishing, 2007.
[3.8] ASTM E 606. *Standard Practice for Strain-Controlled Fatigue Testing*, ASTM standard, 1998.
[3.9] Bhattacharya B and Ellingwood B. Continuum damage mechanics analysis of fatigue crack initiation, *Int. J. Fatigue* 1998, 20, 631-639.
[3.10] Bishop NW and Sherratt F. *NAFEMS-Finite Element Based Fatigue Calculations*, The International Association for the Engineering Analysis Community, Farnham, 2000.
[3.11] Ewalds HL, Wanhill RJ. *Fracture Mechanics*, Edward Arnold Publication, London, 1989.
[3.12] ASTM E 647, *Standard Test Method for Measurement of Fatigue Crack Growth Rates*, ASTM standard, 2000.
[3.13] ASTM E 399, *Standard Test Method for Plane-Strain Fracture Toughness Of Metallic Materials*, ASTM standard, 2000.
[3.14] Kramberger J, Potrč I, Bombek G and Flašker J. Fatigue assessment of high strength steel beam for crawler track cranes, *Gep* 2003, 10-11, 87-90.
[3.15] Vogwell J., Minguez J.M. Failure in lug joints and plates with holes, *Engineering Failure Analysis*, 1995, 2, pp. 129-135.
[3.16] *Abaqus*, Version 6.4, Online Documentation, 2003.
[3.17] *FE-Safe*, Version 5, User's Manual, 2003.
[3.18] Suresh S. *Fatigue of materials*, Cambridge University Press, Cambridge, 1998.
[3.19] Glodež S, Knez M and Kramberger J, Fatigue assessment of high loaded bolted bar connection using strain-life approach, *Key eng. mater.* 2006, 324/325, 711-714.
[3.20] Glodež S, Knez M, Jezernik N and Kramberger J, Fatigue and fracture behaviour of high strength steel S1100Q, *Engineering Failure Analysis.* 2009, 16, 2348-2356.
[3.21] Anderson TL. *Fracture Mechanics-Fundamentals an Applications*, CRC Press, 1995.

CHAPTER 4

FATIGUE CRACK INITIATION AND GROWTH IN MARTENSITIC STEELS

4.1 Introduction

Current industrial design is heavily concerned with the high-cycle fatigue resistance of mechanical components and structures which are often manufactured by the thermal cutting process [4.1]. This process may cause a significant roughness on the cut edge and also changes the microstructural properties of the material in the heat affected zone (HAZ). Surface topography, which is usually represented by roughness parameters, such as average roughness height *Ra*, decreases fatigue life as it causes more stress concentrations on the surface of the structural element. Heat input of the cutting process causes hardening and softening in HAZ and creation of internal residual stresses. This also has an effect on fatigue life, which may be detrimental or beneficial. On the other hand, surface roughness represents a very small notch in the surface, from which microcracks may be initiated.

In general, the complete process of fatigue failure in structural elements may be divided into the following stages [4.2, 4.3]: (1) microcrack nucleation; (2) short crack growth; (3) long crack growth; and (4) occurrence of final failure. In engineering applications, the first two stages are usually referred to as the *"crack initiation period",* while long crack growth is referred to as the *"crack propagation period".* An exact definition of the transition from the initiation to the propagation period is usually not possible. However, the crack initiation period generally accounts for most of the service life, especially in high-cycle fatigue (see Figure 4.1). The total number of stress cycles N can then be determined from the number of stress cycles N_i required for fatigue crack initiation 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 ($N = N_i + N_p$).

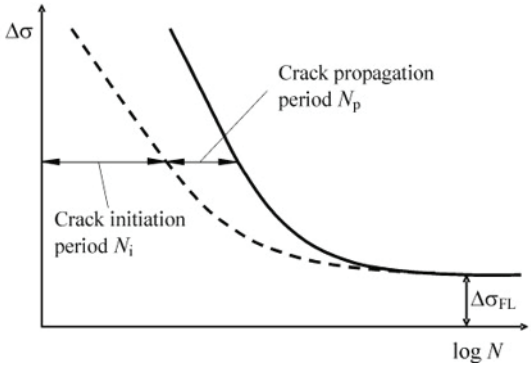

Figure 4.1: Schematic representation of the service life of structural elements

4.2 Fatigue crack initiation

Fatigue crack initiation, as presented in this section, is mainly restricted to microcrack nucleation and short crack growth. As presented in [4.4, 4.5], the behaviour of short cracks under fatigue loading differs substantially from the behaviour of long cracks. Existing research on microcrack nucleation has shown that with many materials microcracks are initiated on slip bands of grains and stretch across the whole grain. One of the frequently proposed methods for solving this problem is using the Tanaka-Mura model for microcrack initiation [4.6]. Using this model, the number of stress cycles N_c required for microcrack nucleation along the slip band of crystal grains can be determined as follows

$$N_C = \frac{8 \cdot G \cdot W_S}{n \cdot (1-v) \cdot d \cdot (\Delta\overline{\tau} - 2k)^2} \tag{4.1}$$

where d is the slip band length, $\Delta\overline{\tau}$ is the average shear stress range on the slip band, G is the shear modulus of the material, W_s is the specific fracture energy per unit area, ν is Poisson's ratio, and k is the frictional stress of dislocation on the slip plane. For martensitic steels, the material parameters in eq. (4.1) may be obtained from specialist literature [4.7].

However, the Tanaka-Mura model has two main deficiencies that limit its use. The model deals only with microcrack nucleation in separate grains and does not deal with macrocrack coalescence from existing microcracks. Another problem is that this model uses average shear stress to determinate

microcrack nucleation [4.8]. For that reason, it often happens that an existing microcrack significantly raises stresses of neighbouring grain, yet the average shear stress in this grain is still below the threshold needed for microcrack nucleation, therefore no nucleation occurred. This problem becomes increasingly pronounced when using lower loads, that is in high-cycle fatigue regime. Therefore, the basic Tanaka-Mura model needs to be significantly improved to describe the crack nucleation process more exactly. In that respect and through a consideration of the research work of other researchers (especially in Ref. [4.7]), some improvements have been suggested, and are presented in this study. Multiple slip bands in each crystal grain are used, as opposed to one slip band going through the centre of the crystal grain. The Tanaka-Mura model is also modified so that microcracks are nucleated in multiple steps (microcrack segmentation). Therefore, the deformation energy is accumulated during algorithm iterations. This modification substantially increases the usability and accuracy of the model in high-cycle fatigue regime. Finally, a special plug-in in the framework of the program package ABAQUS has been written in the Python programming language, for handling microcrack nucleation and coalescence.

4.1.1 Physical model

The problem of microcrack nucleation and short crack growth can be studied on a multi-scale model as shown in Fig. 2. The global model represents a full-scale specimen subjected to a chosen load level. The proposed microcrack nucleation model simulates fatigue crack initiation in particular grains, therefore a complete microstructure needs to be modelled. Since it would be impossible to model a detailed microstructure on a global model (due to memory and computational constraints), a sub-model was used in the place where the highest stresses are present (at the inner hole) and formation of cracks is expected [4.8, 4.9, 4.10]. The size of the sub-model was chosen to be 0.5×0.5 mm, as this would be enough to asses the crack initiation period. The loading of the sub-model was implemented by applying stress distribution of the global model to the boundary edges of the sub-model (top, right, and bottom edge on the right-hand side of the image in Figure 4.2).

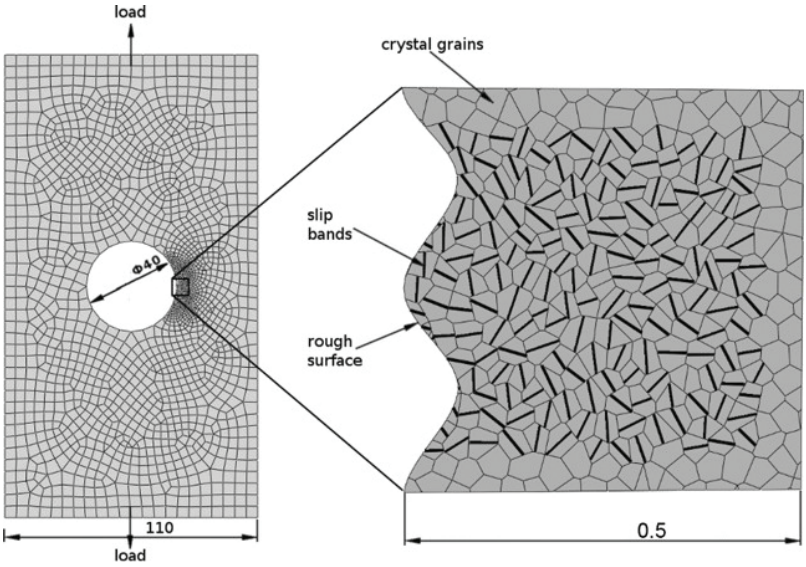

Figure 4.2: Global model and sub-model of the specimen

A numerical simulation of crack initiation was performed with the ABAQUS [4.11], using a plug-in that was written specially for handling microcrack nucleation and coalescence. Individual grains are simulated with a randomly generated Voronoi tessellation and were given randomly oriented [4.12, 4.13] orthotropic material properties of martensite: $C11 = C22 = C33 = 233$ GPa, $C12 = C13 = C23 = 135$ GPa, $C44 = C55 = C66 = 118$ GPa. Considering the average grain size of 20 μm, the whole model consists of approximately 400 individual grains. The right-hand side image in Figure 2 shows such a model, with slip bands drawn through centres of crystal grains. The surrounding grains (with no slip bands indicated), serve as a buffer zone to minimize the numerical mismatch at boundaries (top, right and bottom edge on the right-hand side of the image in Figure 2) and do not partake in microcrack evaluation.

Multiple slip bands

When simulating crack nucleation in a particular grain, multiple slip bands were evaluated. For each grain there was one slip band created through the centre of the grain and additional slip bands were created with a 2 μm offset from the first (see Figure 4.3). Figure 4.4 shows an example where the first microcrack was nucleated in the left grain and influenced its neighbouring

grain on the right, so that a microcrack in it nucleated far from the centre (black dot), close to the first microcrack. The proximity of nucleated microcracks also enables easier microcrack coalescence than in a model where microcracks are allowed to go through grain centres only.

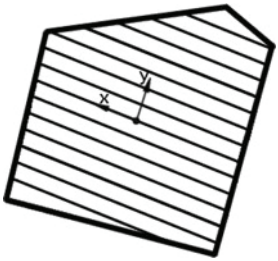

Figure 4.3: A crystal grain with multiple slip bands

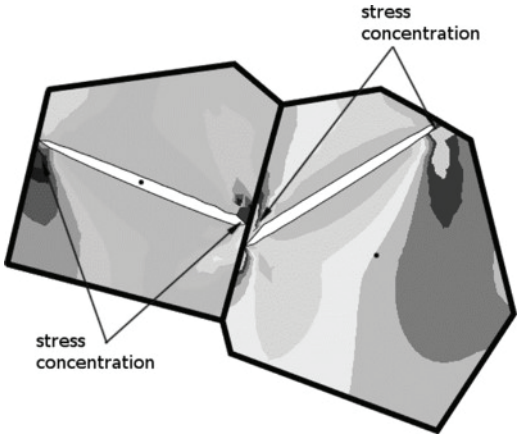

Figure 4.4: Example of nucleated cracks along slip lines

Crack coalescence

In the presented study, a conservative approach was taken to enable crack coalescence. Whenever a new microcrack occurred in the model, every possible combination of microcracks was analysed to see if the average stress on a straight line between two microcracks surpasses the yield stress of the material. If this was the case, a crack was extended through that line, effectively transforming two microcracks into a single larger crack (Figure 4.5). Note the significant stress relaxation in the area between the cracks

with respect to Figure 4.4. When calculating cycles required for fatigue crack initiation, no cycles were attributed to crack coalescences (it is simulated as being instantaneous), so the total cycles of crack initiation equal the sum of cycles needed for each microcrack to nucleate.

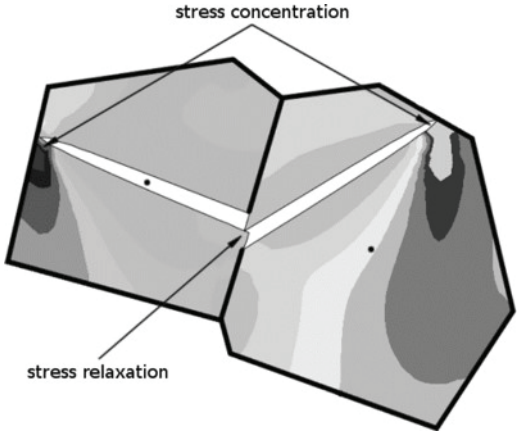

Figure 4.5: Example of two microcracks coalescing

Segmented microcracks

Figure 4.6a shows the shear stress range $\Delta\tau$ (black solid line) along a slip band of a grain, whose neighbouring grain already has a microcrack. This causes a significant stress concentration on the side of the slip band that is adjacent to the existing microcrack. Using the Tanaka-Mura model would not cause a microcrack to nucleate in this case, as average shear stress range $\Delta\overline{\tau}$ (dotted grey line) is lower than the required threshold $2k$ (solid grey line). This problem was solved using segmented microcracks. Figures 4.6 b, c, and d show the evolution of microcrack nucleation using four segments along the slip band. Figure 4.6a shows that the stress level on the leftmost segment $\Delta\overline{\tau}_s$ (dotted black line) surpasses the threshold stress range and a microcrack can occur there. Figure 4.6b shows the next iteration where a microcrack was created on the first segment, causing stress relaxation there and a stress increase on the second segment, so that the stress range surpasses the threshold. Figures 4.6c and 4.6d show the next two iterations where a microcrack progresses through the third and fourth segment and finally forms along the whole slip band.

Since the proposed algorithm for crack nucleation goes through multiple iterations, it is necessary to account for dislocation pile-up by keeping track of the damage that was accumulated over previous stages. In the beginning, the total deformation energy needed for microcrack nucleation is calculated for all segments of slip bands of every grain, using the following equation:

$$U_s^{th} = \frac{8 \cdot W_S}{d_s} \tag{4.2}$$

where U_s^{th} represents a threshold amount of total deformation energy that is required for microcrack nucleation on a particular segment (index s), d_s is the length of that segment and W_s is the specific fracture energy per unit area. Next, in each stage the amount of deformation energy acquired per cycle for each segment is calculated:

$$U_s = \frac{\pi(1-v)(\Delta\bar{\tau}_s - 2k)^2}{G} \tag{4.3}$$

where $\Delta\overline{\tau_s}$ and ΔU_s represent average shear stress range and the amount of deformation energy that is accumulated in one loading cycle on a segment s. The number of cycles N_s^i, required for microcrack nucleation on each segment, is calculated with the following equation:

$$N_s^i = \frac{U_s^{th} - U_s^i}{\Delta U_s} \tag{4.4}$$

where U_s^i represents the already accumulated damage in each segment of stage i.

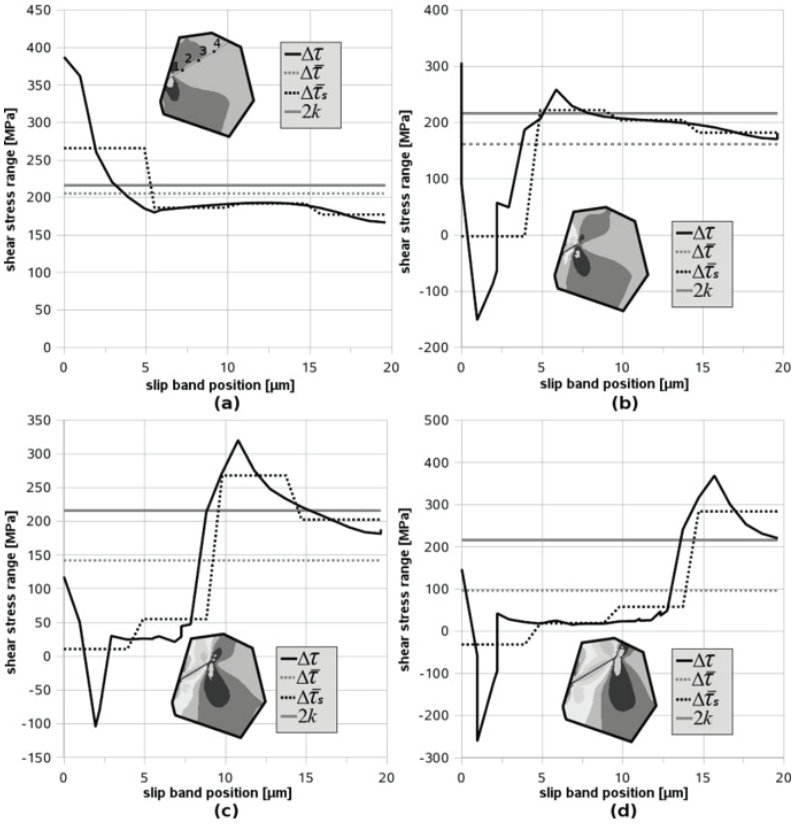

Figure 4.6: Shear stress distribution along the slip band

Considering the procedure as described above, a microcrack will be nucleated on the segment with the smallest N_s^i. For other segments, the accumulated damage in the next stage U_s^{i+1} is increased as expressed in eq. (4.5). The model with the new microcrack is then recalculated and used for the next iteration where the whole process is repeated.

$$U_s^{i+1} = U_s^i + \Delta U_{s\cdot} \cdot N_{s\,min}^i \tag{4.5}$$

In the beginning, microcracks tend to occur scattered around the model and form preferably in larger grains that are favourably oriented and have higher shear stresses. But after a while existing microcracks start coalescing, causing local stress concentrations and amplifying the likelihood of new

microcracks forming near the already coalesced crack. When this crack grows to sufficient length, crack initiation is considered to be finished. The resulting macrocrack is then used as an initial crack for crack propagation. The sum of $N_{smin}{}^{i}$ for all stages represents the total crack initiation life N_i:

$$N_i = \sum_{i=1}^{n} N_{min}^{i} \tag{4.6}$$

4.3 Fatigue crack growth

In order to determine the total lifetime, the crack propagation period should also be calculated. The length and direction of the initial crack were selected so as to resemble the coalesced microcracks obtained in the crack initiation stage. This crack was then propagated through the macro-model according to linear-elastic fracture mechanics. Crack propagation cycles were calculated using Paris' law:

$$\frac{d_a}{dN_p} = C[\Delta K(a)]^m \tag{4.7}$$

with crack growth parameters $C = 6 \cdot 10^{-9}$ (units, MPa, m) and $m = 3$, taken from [4.14] for the treated material. The stress intensity factor range ΔK and the angle of crack propagation were both calculated using the ABAQUS Software [4.11]. As a measure of crack propagation direction, the maximum tangential stress criterion was used.

4.4 Numerical modelling

The commercially available software ABAQUS [4.11] was used, as it enabled an easy enough extension in order to meet the specific demands: additional routines were written, which are needed for the generation, adaptation and evaluation of the model. The developed code that can be invoked through GUI was written in the Python programming language, using the ABAQUS Scripting Interface (ASI) [4.15]. ASI is an object-oriented extension library based on Python, for advanced pre- and post-processing tasks in ABAQUS.

Since in terms of fatigue only a small portion of a structural element needs to be analysed at the place where crack initiation is expected, the decision was made to use a process of sub-modelling in order to transfer the load of

the global-model to the boundaries of the sub-model. Creating a sub-model is a two-step process and is often used for a detailed study of an area of interest, usually a region of high stress. First, a global model needs to be created and analysed. Next, a sub-model is created in the place of interest. The boundary nodes of the sub-model are driven by the macro-model solution. The benefits of sub-modelling are twofold. Firstly, it allows for shorter calculation times, as only a part of whole model needs to be analysed. Secondly, a sub-model can be structured in a different way from a global model. In the presented example, the sub-model is modelled with a disordered microstructure of randomly oriented orthotropic crystal grains, where the global model is continuous and isotropic. Due to the disparity of model properties at the boundary between the global model and the sub-model, a sufficient buffer zone is needed, and many numerical errors of FEM analysis are present close to the boundaries.

In order to assess the microcrack nucleation within and microcrack coalescence between grains, an alternative model description compared to what is usually used in the ABAQUS, was used. This approach utilises a concept of two-dimensional multi-material systems [4.16] with two sets of components: a set of continua and a set of continuum interfaces (Figure 4.7).

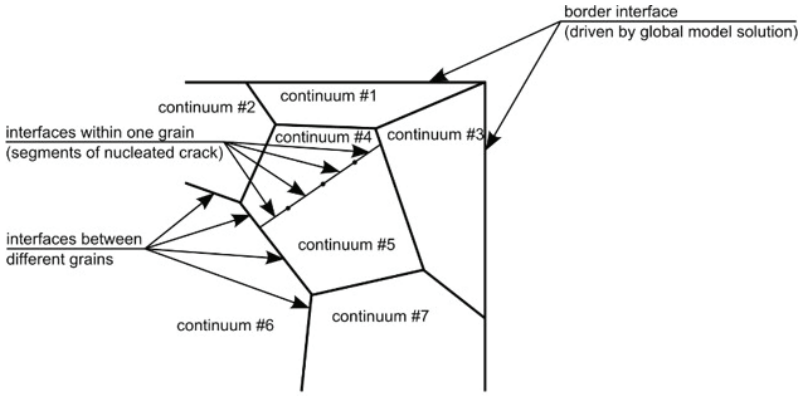

Figure 4.7: Sub-model description by continuum decomposition

In our example, a *continuum* represents a 2D sub-domain of the sub-model, which has its own material orientation. A *continuum interface* defines the interaction between two continua and can exist between two grains of within one grain, if it has a nucleated microcrack. In order to enable segmented

microcrack nucleation, the continuum interface within a grain consists of multiple interface segments with different interaction properties. There are two kinds of interaction properties of continuum interfaces. Initially, the nodes of two continua are aligned at the interface (similar to TIE command in ABAQUS) and have identical deformation and stress values. Upon crack nucleation within a grain or crack coalescence between grains the appropriate continuum interfaces are considered to fail, causing no interaction between continua through that interface. In the example presented in this study, where the micro-model is loaded with tension only in the stress field, this is sufficient. However, if compression stresses occurred in the sub-model, failed interfaces would require some form of contact interaction that would prevent protrusion of one continua into another.

The whole process of crack initiation was developed in Python using the ABAQUS Scripting Language. In this respect, the implementation has two sides. A Python side that controls the entire process and implements novelties described in chapter 4.2, and an ABAQUS side that is used for pre-processing, post-processing and as a solver (see Figure 4.8).

At the beginning, a sub-model is generated according to input parameters (grain size, perturbation). In this sense, a multitude of crystal grains is created using a Voronoi tessellation. Each grain represents one continuum with a unique material orientation, while the properties of the boundary interface are drawn from the macro-model displacement field. In addition, a separate file with properties of all potential slip bands and slip segments for all grains is created, which holds the information about the existing microcrack geometry and the accumulated deformation energy for a particular slip band segment. This model serves as an initial state for the following crack initiation procedure.

Using the ABAQUS/CAE side, the model is first automatically meshed and then calculated using ABAQUS Solver, which produces an output database of stress and displacement field. Shear stress field is used by a custom developed routine (Python side) to determinate where in the sub-model the next microcrack will occur and how many loading cycles it takes. If a microcrack occurs on an already existing slip band segment, the appropriate continuum interface is marked as failed. Otherwise, in case a microcrack occurs in a grain that has no slip bands yet, that grain is split in two parts that have identical material orientation. The interface between these two parts represents a newly created slip band, with a predetermined number of segments where the appropriate segment is marked as failed. In addition, all

adjacent grains with existing microcracks are inspected to determinate if microcrack coalescence can occur. If it does, the appropriate interface between those grains is also marked as failed. All relevant data regarding the new model geometry and deformation energy accumulation for all slip band segments is recorded. In this manner, the created model is then used as an input for the next iteration. The iteration loop is exited when one of two conditions is met. Either a predetermined number of loading cycles has passed or a crack of sufficient size has formed so that the crack initiation process can be considered as finished. Steels are believed to have a fatigue limit at 2×10^6 cycles. To be on the safe side, 5×10^6 was chosen as the maximum amount of cycles that a simulation can run through. In cases where simulation exited after 5×10^6 cycles and no significant crack formation occurred, it was considered that the load was below the fatigue limit.

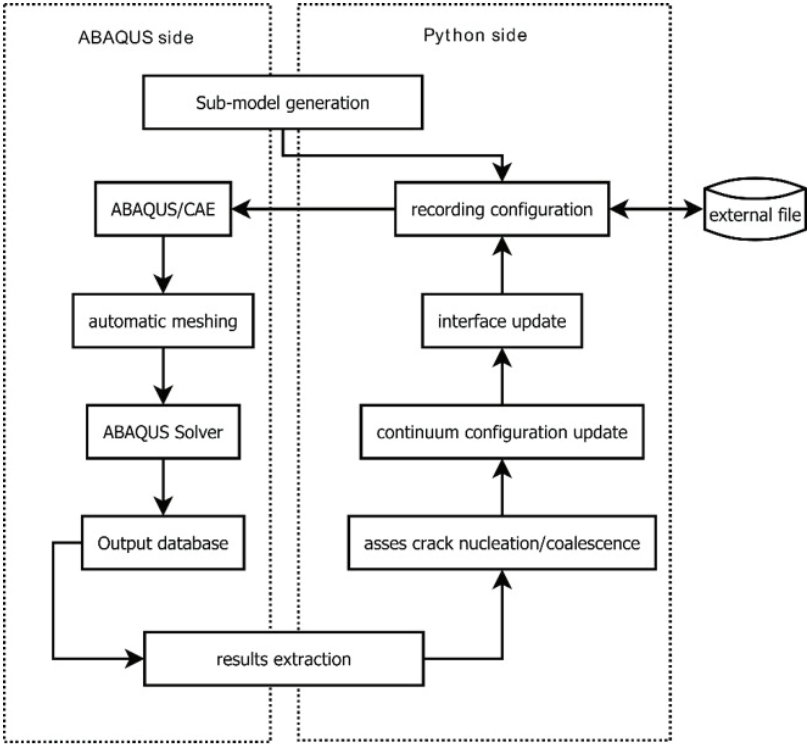

Figure 4.8: Flow diagram of the modelling of crack initiation

4.5 Practical example

4.5.1 Test specimen

The presented model has been used for determining crack initiation and growth in the specimen made of high strength steel S960QL (commercial WELDOX 960E). Components made of this material are often used in heavy-lifting industries where they are subjected to dynamically loaded environments with a high number of required loading cycles [4.14]. Since components used in these cases are usually thermally cut (with laser for thinner and plasma for thicker components) and built-in without any additional treatment, microstructural properties [4.17, 4.18] and surface roughness [4.19, 4.20] have to be considered when determining their fatigue strength.

Fatigue testing was performed on nine laser cut specimens. Specimen dimensions were 200 mm × 110 mm × 5 mm (Figure 4.9), with a circular hole of 40 mm, which creates a stress concentration with a factor $K_t = 2.28$. According to the static tensile test, the yield stress of the material is $R_{p0,2} = 1026$ MPa and the ultimate tensile stress is $R_m = 1064$ MPa, with a 15% elongation at the break. The material can also be used at low temperatures. The standard Charpy-V test (D×10) performed at –40 °C, shows that the material has an impact toughness of 40 J. Table 4.1 shows the chemical composition of the investigated material. Material parameters of the martensitic layer, required for the Tanaka-Mura model, were taken from literature [4.7]: shear modulus $G = 118$ GPa, specific fracture energy per unit area $W_s = 2.0$ kJ/m^2, Poisson's ratio $\nu = 0.3$, and frictional stress of dislocations on the slip plane $k = 108$ MPa.

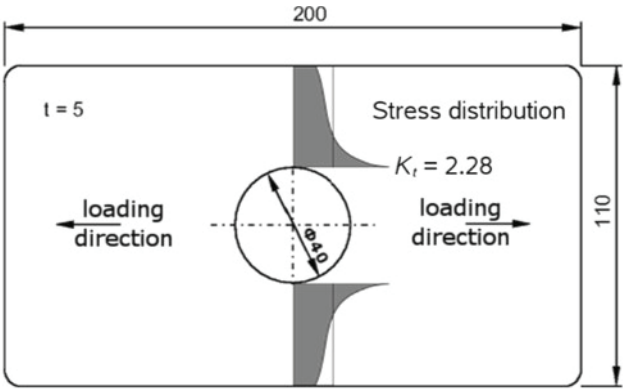

Figure 4.9: Test specimen

Table 4.1: Chemical composition of steel S960QL

C	Si	Mn	P	Cr	Ni	Mo	V	Cu	Al	Nb	N
0.17	0.21	1.24	0.01	0.19	0.05	0.57	0.04	0.01	0.06	0.01	0.01

Cut edge properties

Since the numerical model is directed at simulating fatigue properties of thermally cut steel, edge properties of the test specimen are additionally inspected in terms of surface roughness and residual stresses. Surface roughness of the cut edge was determined using a perthometer. Since the expected roughness is in the range $5 < R_a < 10$ μm, the reference measurement length was chosen to be 12.5 mm. The average arithmetic roughness height was found to be $R_a = 5.0$ μm and the maximum roughness height was $R_z = 37.6$ μm. In simulation, the maximum roughness was modelled, as it can be expected that microcracks will likely form where the highest stress concentration exists.

A microstructural analysis was performed to determine the type of material in the heat-affected zone (HAZ) and its thickness. With electron microscope imaging it was determined that the average grain size in HAZ is 20 μm and the typical distance between martensitic laths is 2 μm. Figure 4.10 shows an electron microscope image of the investigated material, with some distinct grains shaded for better visualization. These properties were then incorporated in the numerical simulation.

No actual measurements of residual stresses have been performed, so the required data was obtained from other works [4.21]. By way of analogy, it was assumed that the maximum compressive stress amounts to 300 MPa at the cut edge, and almost linearly drops to zero at 0.3 mm from the edge. It achieves maximum tensile value at 0.6 mm from the cut edge and fades out asymptotically so it can no longer be detected at 2 mm from the cut edge. Residual stresses have been applied to the appropriate region in the global model.

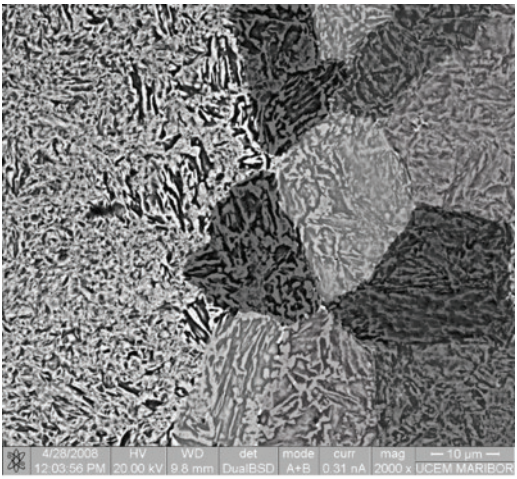

Figure 4.10: Electron microscope image of investigated material

4.5.2 Experimental testing

Experimental fatigue tests were carried out on a pulsating testing machine with uni-axial loading and load control. The stress ratio was $R = 0.1$ with a constant stress amplitude. The used frequency of loading was approximately 150 Hz. Specimens were tested in the same state as they were produced and no additional treatment was performed. Nine specimens were loaded at different stress levels with the maximum stress ranging from 460 to 590 MPa. Stress loads were chosen so as to evaluate high-cycle fatigue range from $2 \cdot 10^5$ to $2 \cdot 10^6$ loading cycles.

4.5.3 Numerical modelling

Fatigue crack initiation

A numerical simulation of fatigue crack initiation was performed using a computational model described in Section 4.2. Four stress levels have been considered for the subsequent numerical analysis: 600, 550, 500, and 485 MPa. Figure 4.11 shows stages of microcrack evolution for a load level of 600 MPa. It is evident that in the beginning microcracks form near the stress concentration due to surface roughness, and where the grain orientation is 45° with respect to the load direction, so that maximum shear stresses occur. After a while, new microcracks nucleate near the existing microcracks, and start coalescing into a larger macrocrack. When the macrocrack reaches

sufficient length (0.3 mm), the simulation stops. If no significant crack coalescence occurs, crack initiation cannot be completed (the load level is below the fatigue limit).

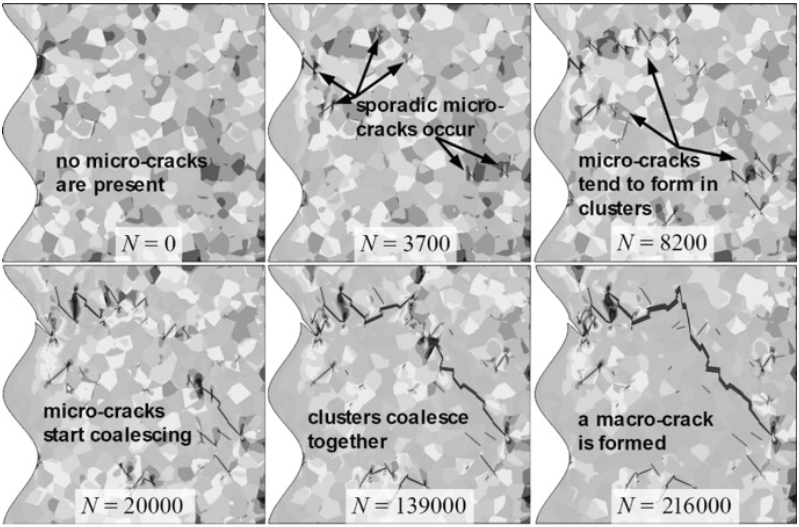

Figure 4.11: Microcrack evaluation for the load level of 600 MPa

Figure 4.12 shows the shear stress distribution and nucleated microcracks for a load level of 485 MPa after $5 \cdot 10^6$ cycles. Since no significant crack coalescence occurred, it was concluded that crack initiation cannot be completed and that the load level is below the fatigue limit.

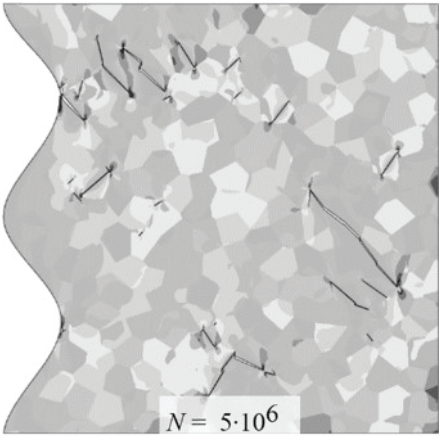

Figure 4.12: Microcrack evaluation for the load level of 485 MPa

Fatigue crack growth

A numerical simulation of fatigue crack growth was performed using the commercial program package ABAQUS. The length and direction of the initial crack were selected in order to make them resemble those of the coalesced microcracks obtained in the crack initiation stage. The crack then propagated through the macro-model according to the simple Paris law, as described in section 4.3. Figure 4.13 shows the relationship between crack length and the stress intensity factor range.

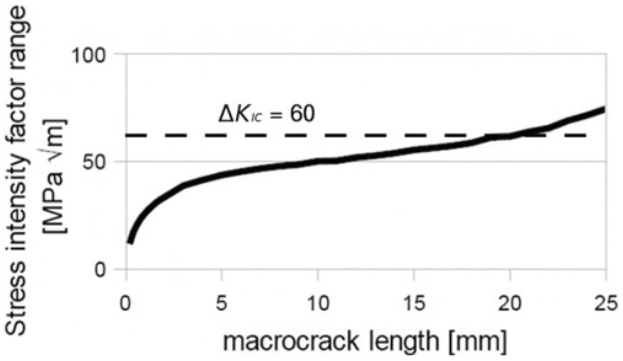

Figure 4.13: The relationship between crack length and the stress intensity factor range

The ABAQUS software package offers basic support in calculating the stress intensity factor K and in assessing the crack propagation direction according to the selected criteria (e.g. maximum tangential stress), but it does not offer support for actual crack propagation. In this respect, an add-on package was developed for the ABAQUS, which automates the whole crack propagation procedure. Figure 4.15 shows how this add-on works. It requires two data input elements: propagation speed (in mm per iteration), and the maximum number of iterations allowed. The output of the add-on is the value of the stress intensity factor K with respect to crack length. The number of cycles required for crack propagation can then be evaluated using the Paris law and appropriate material parameters. Crack propagation after a few calculations is shown in Figure 4.14.

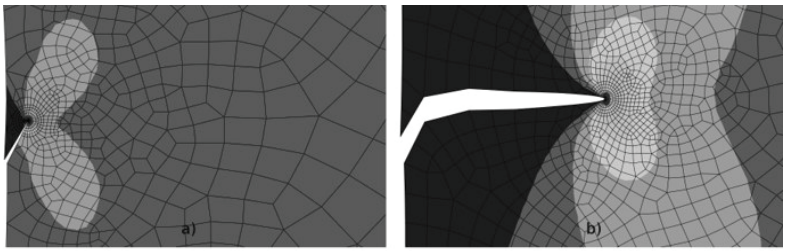

Figure 4.14: Crack path of an initial crack (a) and a long crack (b)

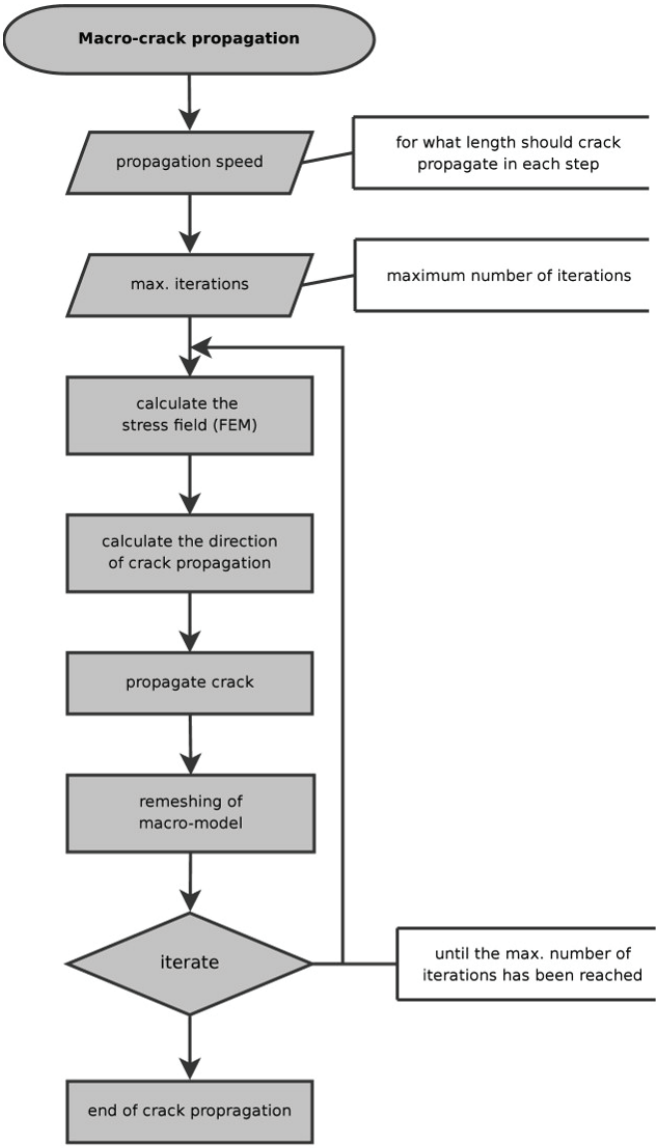

Figure 4.15: Algorithm for calculating crack growth

4.5.4 Comparison between numerical and experimental results

Figure 4.16 shows a comparison between numerical and experimental results. The best fit of the *S–N* curve is also included in the diagram. Since only one test was carried out at each load level, a statistical evaluation of the data or an assessment of its probability were not possible. Nevertheless, the exponent of fatigue strength is equal to b=−0.087, assuming that the coefficient of fatigue strength equals σ_f' = 1167 MPa.

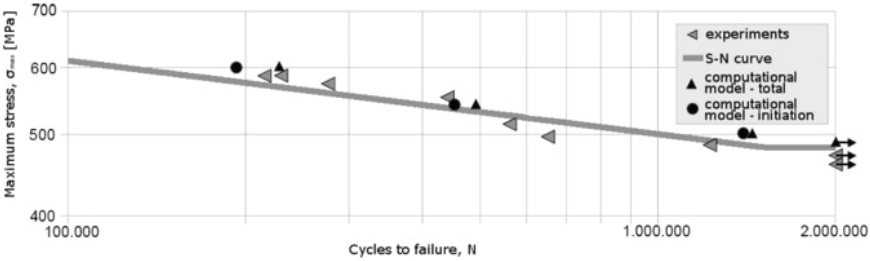

Figure 4.16: Comparison between numerical and experimental results

The computational results presented in Figure 4.16 consist of combined crack initiation and crack propagation. It is important to point out that no fine-tuning of parameters has taken place. All required material parameters have been taken from specialist literature [4.7, 4.14], whereas geometrical properties (roughness and grain size) have been measured. Thought has been given to the fact that two inspected features have an opposite trend. Note that compressive residual stresses retard crack initiation, while edge roughness accelerates it (see Table 4.2).

Table 4.2: Comparing the effects of roughness and residual stresses on a number of initiation cycles for 600 MPa load

	smooth	rough
no residual stresses	238 000	194 000
with residual stresses	296 000	216 000

4.6 Conclusions

The fatigue strength of martensitic steel after thermal cutting is determined using both, a numerical and an experimental approach. Crack initiation becomes increasingly important in high-cycle fatigue (HCF), as it can amount to more than 90% of component life cycles. Crack initiation assessment is also very problematic as it is highly dependent on minute features, such as microstructure and surface roughness. Using the Tanaka-Mura approach for solving crack initiation still leaves open the problem of microcrack coalescence. Furthermore, it does not handle the problem of significant stress gradients caused by existing microcracks as it uses average stress along the slip band. The presented study presents a possible solution to this problem, by introducing segmented slip bands where microcracks nucleate in multiple stages. Crack coalescence is solved by connecting two microcracks, if stresses between them reach the yield stress of material. A plug-in for the ABAQUS package has been created to handle these features.

The proposed method shows a reasonably good correlation with experimental testing, but it still has some deficiencies. Crack coalescence is solved very conservatively and some method should be applied to evaluate the number of loading cycles needed for a crack to extend along the grain boundary. Some preliminary tests have shown that the number of segments on a slip band influences the rate at which microcracks nucleate. Perhaps, not all segments should be treated as equally susceptible to microcrack nucleation, as the proximity of grain boundaries may inhibit the rate at which the microcrack occurs.

References

[4.1] Kramberger J, Jezernik N, Göncz P, Glodež S, Extension of the Tanaka-Mura model for fatigue crack initiation in thermally cut martensitic steels, *Engineering Fracture Mechanics*, 59 (2008), 351–358.

[4.2] Stephens RI, Fatemi A, Stephens RR, Fuchs HO, *Metal Fatigue in Engineering*, John Wiley & Sons Inc, New York, 2001.

[4.3] Podrug S, Jelaska D, Glodež S, Influence of different load models on gear crack path shapes and fatigue lives, Fatigue Fract. Eng. Mater. Struct., 31 (2008), 327–339.

[4.4] Miller KJ. The behaviour of short fatigue cracks and their initiation, parts I+II, Fatigue Fract. Engng. Mater. Struct., 10 (1987), 75–113.

[4.5] Tryon RG, Cruse TA. A reliability-based model to predict scatter in fatigue crack nucleation life, Fatigue Fract. Engng. Mater. Struct., 21 (1998), 257–267.
[4.6] Tanaka K, Mura T. A dislocation model for fatigue crack initiation, J. Appl. Mech., 48 (1981), 97–103.
[4.7] Brückner-Foit A, Huang X, Numerical simulation of microcrack initiation of martensitic steel under fatigue loading, Int. J. Fatigue, 28 (2006), 963–971.
[4.8] Jezernik N, Glodež S, Kramberger J, Evaluating microcrack initiation with a finite element model, CADAM 2008, Proceedings (2008), 25–26.
[4.9] Glodež S, Jezernik N, Kramberger J, Lassen T. Numerical modelling of fatigue crack initiation of martensitic steel, Advances in Engineering Software, 41 (2010), 823–829.
[4.10] Jezernik N, Kramberger J, Lassen T. Glodež S, N umerical modelling of fatigue crack initiation and growth in martensitic steels, Fatigue Fract Engng Mater Struct, 33 (2010) 714–723.
[4.11] Dasault Systems, SIMULIA, ABAQUS, Online Documentation: Version 6.12.
[4.12] Andersson J, The influence of grain size variation on metal fatigue, Int. J. of Fatigue, 27 (2005) 847–852.
[4.13] Simonovski I, Cizelj L, The influence of grains' crystallographic orientations on advancing short crack, Int. J. Fatigue, 29 (2007) 2005–2014.
[4.14] Pusch G, Hübner P, Bruchverhalten des Stahles StE 960 und seiner Schweisswerbindung bei statischer und zyklischer Belastung, European Commission, (1998) Luxembourg (in German).
[4.15] Abaqus Scripting Reference Guide, Version 6.13.
[4.16] Cojocaru D, Karlson AM, An object-oriented approach for modelling and simulation of crack growth in cyclically loaded structures, Adv. in Eng. Soft., 39 (2008) 995–1009.
[4.17] Rodopoulos CA, Rios ER, Theoretical analysis on the behaviour of short fatigue cracks, Int. J. Fatigue, 24 (2002) 719–724.
[4.18] Lankford J, The influence of microstructure on the growth of small fatigue cracks, Fatigue Eng. Mater. Struct., 8 (1985) 161–175.
[4.19] Yue ZF, Surface roughness evolution under constant amplitude fatigue loading using crystal plasticity, Engng. Fract. Mechanics, 72 (2005) 749–757.
[4.20] Antunes V, Ramalho A, Ferreira JM, Identification of fatigue crack propagation modes by means of roughness measurements, Int. J. Fatigue, 22 (2000) 781–788.

[4.21] Klotz U, Trzebiatowski O, Zgraggen M, König W, Winkler M. Eigenspannungen und Mikrostruktur von thermisch getrenntem Stahlblech, Materialwissenschaft und Werkstofftechnik, 32 (2001) 866–873 (in German).

CHAPTER 5

BENDING FATIGUE OF GEARS

5.1 Introduction

Two kinds of teeth damage can occur on gears under repeated loading due to fatigue; namely the pitting of gear teeth flanks and tooth breakage in the tooth root [5.1]. In this study, only tooth breakage is addressed and the developed computational model is used for calculation of tooth bending strength, i.e. the service life of gear tooth root.

The standardized procedures (DIN, AGMA, ISO, etc.) for the approximate determination of load capacity of gear tooth root are commonly based on the comparison of the maximum tooth-root stress with the permissible bending stress [5.2, 5.3]. Their determination depends on a number of different coefficients that allow for proper consideration of real working conditions (additional internal and external dynamic forces, contact area of engaging gears, gear's material, surface roughness, etc.). The traditional procedures are exclusively based on the experimental testing of reference gears, and they consider only the final stage of the fatigue process in the gear tooth root, i.e. the occurrence of final failure.

However, the complete process of fatigue failure in mechanical elements such as gears may be divided into the following stages [5.4, 5.5]: (1) microcrack nucleation; (2) short crack growth; (3) long crack growth; and (4) occurrence of final failure. In engineering applications, the first two stages are usually referred to as the *"crack initiation period",* while long crack growth is referred to as the *"crack propagation period".* An exact definition of the transition from initiation to propagation period is usually not possible. However, the crack initiation period generally accounts for most of the service life, especially in high-cycle fatigue. The complete service life of mechanical elements can then be determined from the number of stress cycles N_i required for fatigue crack initiation, 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.

One of the most convenient representations of fatigue crack growth is the Kitagawa-Takahashi plot of applied stress range required for crack growth, $\Delta\sigma$, against crack length, a, using logarithmic scales, as shown in Figure 5.1 [5. 6]. In the area of a constant value of the threshold stress intensity range ΔK_{th}, linear elastic fracture mechanics (LEFM) can be used to analyse fatigue crack growth. The threshold crack length a_{th}, below which LEFM is not valid, may be estimated approximately as [5.7]:

$$a_{th} \approx \frac{1}{\pi}\left(\frac{\Delta K_{th}}{\Delta\sigma_{FL}}\right)^2 \tag{5.1}$$

where $\Delta\sigma_{FL}$ is the fatigue limit, see Figure 5.1.

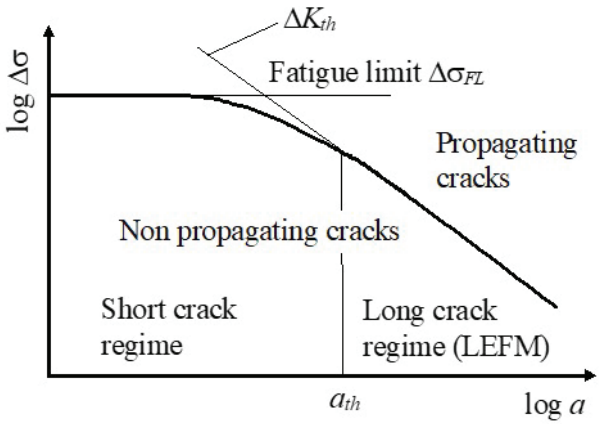

Figure 5.1. The Kitagawa-Takahashi diagram

The threshold crack length a_{th} thus defines the transition point between short and long cracks, i.e., the transition point between the initiation and propagation period in engineering applications. However, a wider range of values has been selected for a_{th}, usually between 0.05 and 1 mm for steels, where high strength steels take the smallest values [5.8].

Fracture mechanics has developed into a useful discipline for predicting the strength and life span of cracked gear tooth, and many authors have used this theory for calculating the tooth bending strength. A short review of earlier studies in this field is given in [5.9] and [5.10]. The authors have performed a comprehensive study to describe fatigue crack growth using

the fracture mechanics theory, where numerical procedures such as the Finite Element Method (FEM) and the Boundary Element Method (BEM) are normally used for that purpose. In some subsequent studies, authors tried to observe the specific effects on fatigue crack growth. Kato et al. [5.9] developed a method to simulate fatigue crack growth in a carburized gear tooth, where the effect of residual stresses is taken into account. In their model, the crack initiation period was neglected and the entire service life of the gear tooth was assumed to be the process of crack propagation. Blarasin et al. [5.10] studied the problem of fatigue crack propagation in specimens similar to gear teeth, considering the influence of different surface treatments on the service life. Lewicki and Ballarini [5.11] investigated the effect of gear rim thickness on crack propagation in a gear tooth root. In their study, the crack initiation period was determined experimentally. The effect of crack closure on crack propagation in a gear tooth root was studied by Guagliano and Vergani [5.12]. The results of their investigations show that the crack closure effect may be significant when the load applied is lower.

Recently, some authors focused on the appropriate analytical approach for the determination of low and high cycle bending fatigue in carburized gears [5.13, 5.14]. In their models, the strain-life approach was proposed for low and high cycle fatigue regimes. The hardness method and the multilayer method were used to acquire strain-life fatigue properties of material layers. Through the rule of mixture, average cyclic stress-strain curves for carburized specimens were obtained. In addition, an approach was suggested for translating axial to bending fatigue data through Neuber's rule and modifying factors. The authors concluded that 93% of the predicted bending fatigue lives fall within the scatter factor of three when compared to the experimental results, while overall good correlation is observed between predicted and experimental data. For future work, the authors suggested additional analytical bending fatigue life investigations of carburized specimens considering various carburizing depths.

As mentioned in [5.15, 5.16], Powder Metallurgy (P/M) can be a useful tool in manufacturing parts such as gears. Namely, the P/M technology is suitable for high volume production with very little wastage of material. Powder metal gears were initially used only for light-duty applications such as toys and power tools. Today, powder metal gears are a cost-efficient alternative for machined gears in larger series in the automotive industry (synchronizer gears, oil pump gears, engine gears, etc.). The next step should be power transmission gears [5.17]. A critical review regarding the

application of P/M-sintered gears for transmissions and machinery was presented by Dizdar [5.18]. In his work, he pointed out that P/M-sintered gears can reach a relatively high level of dynamic strength in comparison to wrought steel gears. However, they also offer highly sustainable production, low cost and full recycling for a range of applications in the automotive, power tools and home appliances industry. In the last few years, high-performance sintered steel gears have been extensively investigated by researchers, who focused on new technologies (surface densification, gear rolling, burnishing, shot peening, high-density pressing, warm compaction, warm die pressing techniques, etc.) and their influence on gear characteristics [5.19–5.21].

The main purpose of this study is to present a complete computational model for the determination of the service life of a gear tooth root, where both, crack initiation and crack propagation period are analysed using appropriate numerical models. The proposed model is used on a real spur gear pair made from high strength alloy steel (through-hardened) without additional surface treatment (case-hardening or shot-peening). In fact, residual stresses on the surface layer are smaller in comparison with case-hardened gears and are not considered in this study. Furthermore, the material parameters used in the computational analysis have been determined previously by means of appropriate test specimens, whereby the material and thermal treatment (through-hardening) were the same as with the treated gears.

5.2 Fatigue crack initiation

The initiation of fatigue cracks represents one of the most important stages in the fatigue process. The position and mode of fatigue crack initiation depend on the microstructure of a material, the type of applied stress, and the micro- and macro-geometry of the specimen. The initiation phase of fatigue life in a virgin material is often assumed to constitute the growth of short cracks up to the size a_{th} (see Figure 5.1), which is the transition length of short cracks into long cracks.

Fatigue crack initiation includes the early development of fatigue damage and is strongly dependent on the size scale of observation. For example, scientists are likely to consider the nucleation of flaws along with persistent slip bands (PSB) as the initiation stage of fatigue damage, whilst mechanical engineers may associate the resolution of crack detection with the threshold for crack nucleation. Between this wide range of view-points lies a variety

of failure mechanisms that are affiliated with the inception of microscopic flaws at grain boundaries, twin boundaries, inclusions, as well as microscopic and macroscopic stress concentrations [5.8]. It is quite difficult to find full agreement on what is meant by the term short crack, but for steels, this might be all cracks less than 1 mm [5.22].

The model for fatigue crack initiation presented here is based on the continuum mechanics approach, where it is assumed that the material is homogeneous and isotropic, i.e., without imperfections or damages. Methods for fatigue analyses are usually based on the Coffin-Manson relation between deformations (ε), stresses (σ) and the number of loading cycles (N_i). However, the strain-life method (ε-N_i) is usually applied in order to determine the number of stress cycles N_i required for fatigue crack initiation, where it is assumed that a crack is initiated at the point of the largest stresses in the material. The total cyclic strain range Δε comprises two components (elastic and plastic cyclic strain range $\Delta\varepsilon_e$ and $\Delta\varepsilon_p$) and can be described as [5.23]:

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\Delta\sigma}{2E} + \varepsilon_f' \left(\frac{\Delta\sigma}{2\sigma_f'}\right)^{\frac{1}{n'}} = \frac{\sigma_f'}{E}(2N_i)^b + \varepsilon_f'(2N_i)^c \qquad (5.2)$$

where Δσ is the applied stress range, E is Young's modulus, n' is the cyclic strain hardening exponent, 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, see Figure 5.2. The number of stress cycles N_i required for fatigue crack initiation can then be solved iteratively from equation (5.2) for the applied stress range Δσ and the appropriate material parameters E, n', σ_f', ε_f', b and c.

It is known from practical applications that fatigue failures on gears are usually nucleated at the surface and so surface conditions become an extremely important factor influencing fatigue strength. Normally, scratches, pits, machining marks, etc. influence fatigue strength by providing additional stress raisers that aid the process of crack nucleation. Broadly speaking, high strength steels are more adversely affected by a rough surface finish than softer steels. Therefore, the influence of the surface finish on fatigue strength is strongly related to the tensile strength of the material. The surface finish correction factor C_{sur} is presented in Figure 5.3 in dependence on the surface roughness Ra and tensile strength of the material R_m [5.5]. Based on this assumption, the actual service life of

gears may be reduced with regard to the appropriate value of C_{sur}, which is calculated with the following equation:

$$\Delta\sigma_{FLr} = \Delta\sigma_{FR} \cdot C_{sur} \tag{5.3}$$

where $\Delta\sigma_{FLr}$ is the real fatigue limit and $\Delta\sigma_{FL}$ is the fatigue limit of polished laboratory specimen.

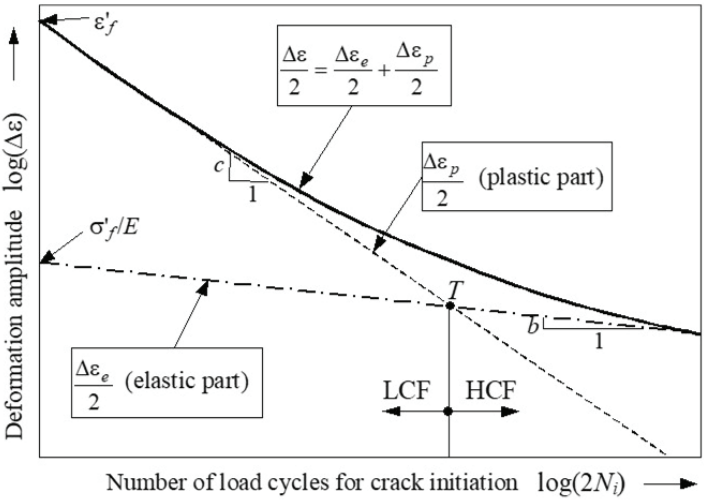

Figure 5.2. Strain-life (ε-N_i) method for the fatigue crack initiation

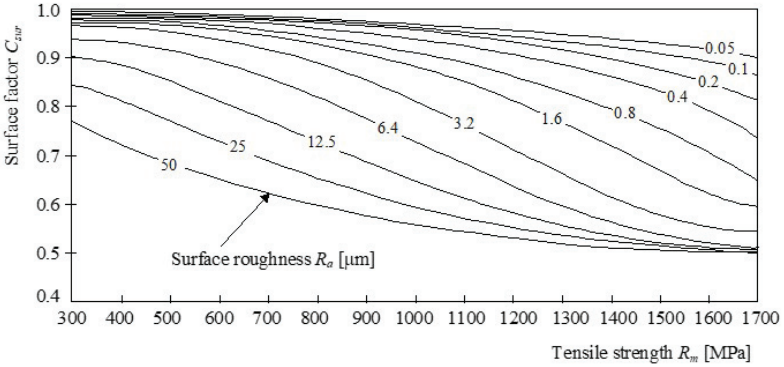

Figure 5.3. Surface finish correction factor C_{sur}

5.3 Fatigue crack propagation

The application of LEFM to fatigue is based upon the assumption that the fatigue crack growth rate, da/dN, is a function of the stress intensity factor range $\Delta K=K_{max}-K_{min}$, where a is the crack length and N is the number of loading cycles. In this study, the simple Paris equation is used to describe the crack growth rate [5.24]:

$$\frac{da}{dN} = C[\Delta K(a)]^m \tag{5.4}$$

where C and m are the material parameters. With the integration of eq. (5.4) one can obtain the number of loading cycles to the crack propagation N_p:

$$\int_0^{N_p} dN = \frac{1}{C} \cdot \int_{a_{th}}^{a_c} \frac{da}{[\Delta K(a)]^m} \tag{5.5}$$

Equation (5.5) indicates that the required number of loading cycles N_p for a crack to propagate from the initial length a_{th} to the critical crack length a_c can be explicitly determined, if C, m and $\Delta K(a)$ are known. C and m are material parameters and can be obtained experimentally, usually by means of a three-point bending test according to the standard procedure ASTM E 399 [5.25]. For simple cases, the dependence between the stress intensity factor and the crack length K=f(a) can be determined using the methodology given in [5.24]. For geometry and loading cases that are more complicated, it is necessary to use alternative methods. In this work, the Finite Element Method in the framework of the program package FRANC2D [5.26] was used for the simulation of fatigue crack growth. In this approach, the determination of the stress intensity factor is based on the displacement correlation method using singular quarter-point, six node triangular elements around the crack tip, Figure 5.4.

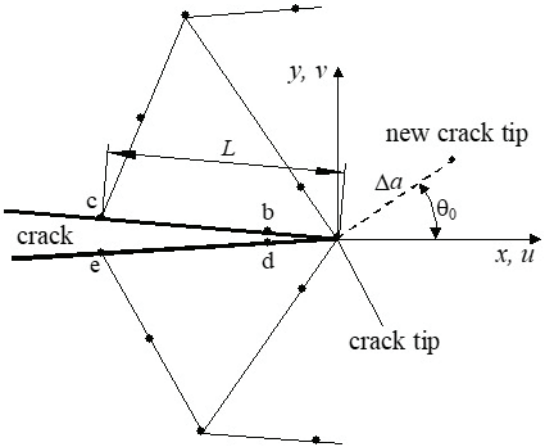

Figure 5.4. Triangular quarter-point elements around crack tip

The stress intensity factor in mixed mode plane strain condition can then be determined from the nodal displacements as:

$$K_I = \frac{2G}{(3-4v)+1} \cdot \sqrt{\frac{\pi}{2L}} \cdot [4v_d - v_e - 4v_b + v_c]$$
$$K_{II} = \frac{2G}{(3-4v)+1} \cdot \sqrt{\frac{\pi}{2L}} \cdot [4u_d - u_e - 4u_b + u_c] \tag{5.6}$$

where G is the shear modulus of the material, ν is the Poisson ratio, L is the finite element length on crack face, u and v are displacements of the finite element nodes b, c, d and e, see Figure 5.4. The combined stress intensity factor is then:

$$K = \sqrt{(K_I^2 + K_{II}^2) \cdot (1 - v^2)} \tag{5.7}$$

The computational procedure is based on incremental crack extensions, where the size of the crack increment is prescribed in advance. In order to predict the crack extension angle, the maximum tangential stress criterion (MTS) is used. This criterion proposes that the crack propagates from the crack tip in a radial direction in the plane perpendicular to the direction of greatest tension (maximum tangential tensile stress). The predicted crack propagation angle (see Figure 5.4) can be calculated by:

$$\theta_0 = 2\tan^{-1}\left[\frac{1}{4}\cdot\frac{K_I}{K_{II}} \pm \sqrt{\left(\frac{K_I}{K_{II}}\right)^2 + 8}\right] \tag{5.8}$$

A new local remeshing around the new crack tip is then required. The procedure is repeated until the stress intensity factor reaches the critical value K_c, when the complete tooth fracture is expected. Following the above procedure, one can numerically determine the functional relationship K=f(a).

5.4 Practical example

The presented model has been used for the computational determination of the service life of a real spur gear, with the complete data set given in Table 5.1. The gear is made of high strength alloy steel 42CrMo4 (0.43 %C, 0.22 %Si, 0.59 %Mn, 1.04 %Cr, 0.17 %Mo) with Young's modulus E=2.1·10^5 MPa and Poison's ratio ν=0.3. The gear material is thermally treated (through-hardening) as follows:

– flame heated at 810 °C; 2 min,
– hardened in oil; 3 min
–tempered at 180 °C; 2 h.

Table 5.1: Basic data of a treated spur gear

Profile	Involute
Normal module	m_n = 4.5 mm
Number of teeth	z = 39
Pressure angle on pitch circle	α_n = 24°
Coefficient of profile displacement	x = 0.06
Tooth width	B = 28 mm
Gear material	42CrMo4

5.4.1 Fatigue crack initiation

The strain-life method (ε-N_i) in the framework of the FEM program package MSC/FATIGUE [5.27] has been used to determine the number of stress cycles N_i required for fatigue crack initiation. The material parameters n'=0.14, σ'_f=1820 MPa, ε'_f = 0.65, b = – 0.08 and c = – 0.76 according to

eq. (5.2) have been taken from the material database available in [5.27]. Based on the gear data set given in Table 1, the finite element model shown in Figure 5.5 has been constructed for the further numerical calculation of the stress-strain field in a gear tooth root for the plane strain conditions. The gear tooth was loaded with normal pulsating force F (different values of F have been considered in the numerical computations, see Table 5.2), which acts at the outer point of single tooth contact. Computational analyses have been performed at the point where maximum principal stresses occur in a gear tooth root, see Figure 5.5. The influence of the surface finish on fatigue strength has been considered with the surface finish factor C_{sur} as described in section 5.2 for different surface roughness (R_a= 0.8, 3.2 and 6.4 μm). The results of these numerical computations are presented in Table 5.2.

Table 5.2: Computational results for the fatigue crack initiation period depending on the surface finish of gear teeth flanks

Loading F [N/mm]	Maximum principal stress in a gear tooth root σ [MPa]	Number of stress cycles for the fatigue crack initiation N_i		
		R_a = 6.4 μm	R_a = 3.2 μm	R_a = 0.8 μm
800	527	$1.368 \cdot 10^7$	$5.270 \cdot 10^7$	$2.327 \cdot 10^8$
1000	659	$2.327 \cdot 10^6$	$6.732 \cdot 10^6$	$2.049 \cdot 10^7$
1200	790	$4.370 \cdot 10^5$	$1.144 \cdot 10^6$	$3.141 \cdot 10^6$
1400	922	$9.811 \cdot 10^4$	$2.292 \cdot 10^5$	$5.677 \cdot 10^5$
1600	1050	$3.151 \cdot 10^4$	$6.504 \cdot 10^4$	$1.446 \cdot 10^5$

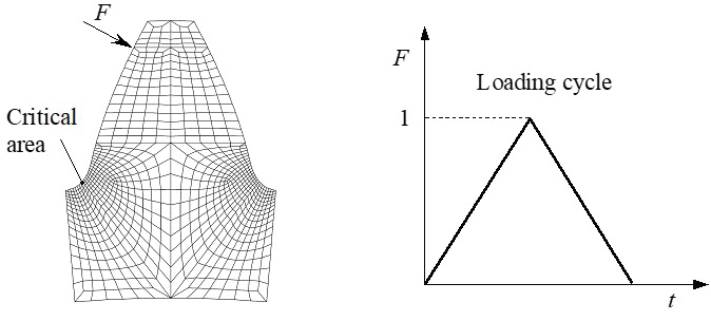

Figure 5.5. Finite element model with the pattern of loading cycle for the subsequent fatigue analysis

5.4.2 Fatigue crack propagation

The FEM-programme package FRANC2D, described in section 5.3, was used for the numerical simulation of fatigue crack growth. The initial crack was placed perpendicularly to the surface at the point where crack initiation had been determined previously, see Figure 5.6. In numerical computations, it was assumed that the initial crack corresponds to the threshold crack length a_{th}, see Section 5.1. Considering the material parameters $\sigma_{FL} \approx 550$ MPa and $K_{th} \approx 269$ MPa$\sqrt{}$mm [5.28], the threshold crack length is equal to $a_{th} \approx 0.1$ mm. The fracture toughness $K_{Ic} \approx 2620$ MPa$\sqrt{}$mm, and the material parameters $C=3.31 \cdot 10^{-17}$ mm/cycl/(MPa$\sqrt{}$mm)^m and $m=4.16$ were determined previously by means of three-point bending samples according to the ASTM E 399-80 standard and for the same material as used in this study [5.29].

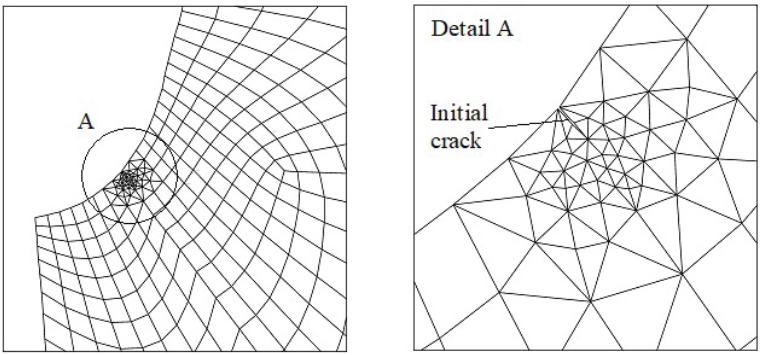

Figure 5.6. Finite element mesh around an initial crack in a gear tooth root

The tooth loading was set in the same way as in the numerical analysis of fatigue crack initiation, see Section 5.4.1. During numerical simulations, the crack increment size Δa was 0.2 mm up to the crack length $a = 4$ mm, and after this 0.4 mm up to the critical crack length a_c, see Figure 5.4. To be able to determine the number of loading cycles N_p required for the crack to propagate from the initial crack length a_{th} to the critical crack length a_c according to equation (5.5), it is necessary to determine the function $\Delta K=f(a)$ first. Figure 5.7 shows the functional relationship between the combined stress intensity factor K and crack length a, where K is obtained with equation (5.7) using numerically determined values of K_I and K_{II}. Numerical analyses have shown that the K_I stress intensity factor is much higher if compared with K_{II} (K_{II} was less than 5 % of K_I for all load cases

and crack lengths). Therefore, fracture toughness K_{Ic} can be considered as the critical value of K and the appropriate crack length can be taken as the critical crack length a_c. The loading cycles N_p for crack propagation to the critical crack length can be estimated using equation (5.5), see Table 5.3. Figure 5.8 shows the numerically determined crack propagation path in a gear tooth root.

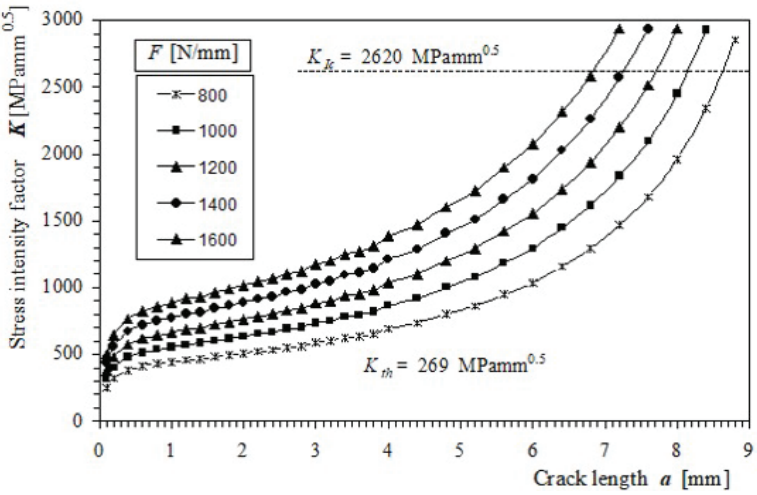

Figure 5.7. Functional relationship between the stress intensity factor K and crack length a

Table 5.3: Gear tooth crack propagation life of the cracked tooth

Loading F [N/mm]	Critical crack length a_c [mm]	Number of cycles N_p
800	8.6	$1.160 \cdot 10^6$
1000	8.2	$4.372 \cdot 10^5$
1200	7.7	$2.005 \cdot 10^5$
1400	7.3	$1.047 \cdot 10^5$
1600	6.9	$7.206 \cdot 10^4$

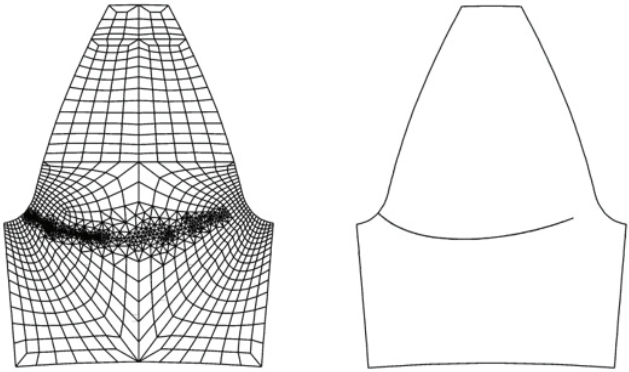

Figure 5.8. Predicted crack propagation path in a gear tooth root

On the basis of the computational results for crack initiation (N_i) and crack propagation (N_p) period in Table 5.2 and Table 5.3, the complete service life of gear tooth root can be obtained as a sum $N = N_i + N_p$, see Figure 5.9. It is evident from Figure 5.9, that the ratio between periods of initiation and the end of propagation (i.e., final breakage) depends on the stress level. At a low stress level, almost all service life is spent in crack initiation, but at high stress levels the significant part of the life is spent in crack propagation (note that the abscissa of the diagram has a logarithmic scale). The computational results were compared with the available experimental results [5.28], which were obtained using the FZG-test machine, for the same material and thermal treatments as used in this study. The comparison of computational and experimental results shows a reasonable agreement.

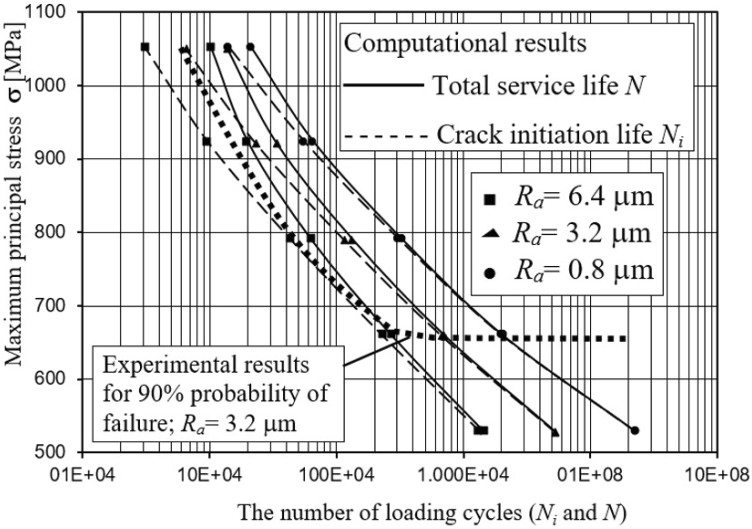

Figure 5.9. Comparison between computational and experimental results

5.5 Conclusions

The study presents a computational model for determining the service life of gears with regard to bending fatigue in a gear tooth root. The fatigue process leading to tooth breakage in a tooth root is divided into the crack initiation (N_i) and the crack propagation (N_p) period, which enables the determination of the total service life as a sum $N = N_i+N_p$. The crack initiation period is based on stress-strain analysis in the framework of the Finite Element Method, where it is assumed that the crack is initiated at the point of maximum principal stress in a gear tooth root. The displacement correlation method is then used for numerical determination of the functional relationship between the stress intensity factor and crack length K=f(a), which is necessary for a consequent analysis of fatigue crack growth.

The proposed model is used to determine the complete service life of spur gear made from high strength alloy steel 42CrMo4 (through-hardened). The computational analysis is performed for variable loading and different surface finishes of gear teeth flanks, which is characterized by the surface finish correction factor C_{sur} as a function of surface roughness and tensile strength of the material. The results of the computational analysis are shown

in Figure 5.9, where two curves are presented for each surface finish of gear teeth flanks: the crack initiation curve and the curve of the end of crack propagation, which at the same time represents the total service life. The results show that at low stress levels near the fatigue limit, almost all service life is spent in crack initiation. It is very important to determine the service life of real gear drives in engineering applications, because the majority of them actually operate with loading conditions close to the fatigue limit. The computational results correspond well with the available experimental data [5.28].

The estimated bending fatigue life of the gear may deviate from real service life, because certain effects, such as non-homogenous materials and possible causes of retardation of the crack propagation (crack closure), were not taken into account in the numerical analysis. Therefore, the model can be further improved by means of additional theoretical and numerical research, although additional experimental results will be needed in order to provide the required material parameters.

References

[5.1] ISO 6336, Calculation of load capacity of spur and helical gears, International Standard, 2019.

[5.2] Podrug S, Glodež S, Jelaska D, Numerical Modelling of Crack Growth in a Gear Tooth Root, *Strojniški vestnik*, 57 (2011) 579–586.

[5.3] Podrug S, Jelaska D, Glodež S, Influence of different load models on gear crack path shapes and fatigue lives, *FFEMS*, 31 (2008) 327–339.

[5.4] Glodež S, Flašker J, Ren Z, A new model for the numerical determination of pitting resistance of gear teeth flanks, *Fatigue Fract. Eng. Mater. Struct.*, 20 (1997) 71–83.

[5.5] Glodež S, Šraml M, Kramberger J, A computational model for determination of service life of gears, *International journal of fatigue*, 24 (2002) 1013–1020.

[5.6] Kitagawa H, Takahashi S, Applicability of fracture mechanics to very small cracks or cracks in the early stage, Proceedings of the 2nd International Conference on the Behaviour of Materials, Boston, (1976) 627–631.

[5.7] Bhattacharya B, Ellingwood B, Continuum damage mechanics analysis of fatigue crack initiation, *International journal of fatigue*, 20 (1998) 631–639.

[5.8] Suresh S, Fatigue of Materials, Cambridge University Press, Cambridge, 1998.
[5.9] Blarasin A, Guagliano M, Vergani L, Fatigue crack growth prediction in specimens similar to spur gear teeth, *Fatigue Fract. Engng Mater. Struct*., 20 (1997) 1171–1182.
[5.10] Kato M, Deng G, Inoue K, Takatsu N, Evaluation of the Strength of Carburized Spur Gear Teeth Based on Fracture Mechanics, JSME International Journal, 36 (1993) 233–240.
[5.11] Lewicki DG, Ballarini R, Rim thickness effects on gear crack propagation life, *International Journal of Fracture*, 87 (1997) 59–86.
[5.12] Guagliano M, Vergani L, Effect of crack closure on gear crack propagation, *International Journal of Fatigue*, 23 (2001) pp. 65–73.
[5.13] Čular I, Vučković K, Žeželj D, Glodež S, Analytical approach for low and high cycle bending fatigue life prediction of carburized gear steel specimens, *Engineering Failure Analysis*, 108 (2020), 1–12.
[5.14] Vučković K, Galić I, Božić Ž, Glodež S, Effect of friction in a single-tooth fatigue test, *International journal of fatigue*, 114 (2018) 148–158.
[5.15] Glodež S, Šori M, Vučković K, Risović S, Determination of service life of sintered powder metallurgy gears in regard to tooth bending fatigue, *Croatian journal of forest engineering*, 39 (2018) 129–137.
[5.16] Glodež S, Šori M, Verlak T, A computational model for bending fatigue analyses of sintered gears, *Strojniški vestnik*, 60 (2014) 649–655.
[5.17] Flodin A, Brecher C, Gorgels C, Rothlingshofer T, Henser J, Designing powder metal gears, *Gear Solutions*, August (2011) 26–35.
[5.18] Dizdar S,. High-performance sintered-steel gears for transmissions and machinery: A critical review, *Gear Technology*, August (2012) 60–65.
[5.19] Sudhakar KV. Fatigue behaviour of a high density powder metallurgy steel, *International Journal of Fatigue*, 22 (2000) 729–734.
[5.20] Sonsino CM, Mueller F, Mueller R. The improvement of fatigue behaviour of sintered steels by surface rolling, *International Journal of Fatigue*,14 1992, 3–13.
[5.21] Koide T, Ishizuka I, Takemasu T, Miyachika K, Oda S, Load bearing capacity of surface-rolled sintered metal gears, *Int. J. of Automation Technology*, 2 (2008) 334–340.

[5.22] Aberšek B, Analysis of short fatigue crack on gear teeth, Ph.D. thesis, University of Maribor, Faculty of Mechanical Engineering, Maribor, 1993.
[5.23] Dieter GE, Mechanical Metallurgy, McGraw-Hill Book Company (UK), Metric Edition, 1988.
[5.24] Ewalds HL, Wanhill RJ, Fracture Mechanics, Edward Arnold Publication, London, 1989.
[5.25] ASTM E 399-80, Standard Test Method for Linear-Elastic Plane-Strain Fracture Toughness K_{Ic} of Metallic Materials, American Standard.
[5.26] FRANC2D, User's Guide, Version 2.7, Cornell University.
[5.27] MSC/FATIGUE, Quick Start Guide Version 8, The McNeal-Schwendler Corporation, Los Angeles.
[5.28] Niemann G., Winter H., *Maschinenelemente - Band II*, Springer Verlag, 1983 (in German).
[5.29] Aberšek B., Analysis of short fatigue crack on gear teeth, Ph.D. thesis, University of Maribor, Faculty of Mechanical Engineering, Maribor, 1993.

CHAPTER 6

PITTING PHENOMENON ON GEAR FLANKS

6.1 Introduction

As already mentioned in Chapter 5, two kinds of teeth damage can occur on gears under repeated loading due to fatigue; namely the pitting of gear teeth flanks and tooth breakage in the tooth root [6.1, 6.2]. In this Chapter, only the pitting phenomenon is addressed. The process of surface pitting can be visualised as the formation of small surface-breakings or subsurface initial cracks, which 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 pit [6.3–6.5].

Although surface pitting is a well-known problem in engineering and many hypotheses have been proposed to-date, a general theory, which would describe the complicated mechanism realistically and completely, has yet to be established. Some of the early attempts to apply fracture mechanics to the study of pit formation mechanism [6.6–6.8], led to some very comprehensive computational models presented in [6.9–6.16]. These models assume that gear tooth pitting may be surface or sub-surface initiated. The former is observed in gears with rough surfaces and poor lubrication, since it is strongly influenced by surface roughness and damage, like machining marks, large notches, etc. The sub-surface pitting initiation is common in gears with smooth contact surfaces and good lubrication, where sub-surface stress during contact initiates a sub-surface crack.

The presented model for the simulation of surface pitting is based on the theory of short fatigue crack growth, where the crack growth rate da/dN is proportional to the crack tip plastic displacement δ_{pl} [6.17]:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C_o \cdot \left(\delta_{pl}\right)^{m_o} \tag{6.1}$$

where C_o and m_o are material constants, which can be determined experimentally. 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 is expressed in the following form [6.18]:

$$\delta_{pl} = \frac{2(1-\nu)}{G\sqrt{\pi}} \cdot \frac{\sqrt{1-n^2}}{n} \cdot K(a) \cdot \sqrt{a} \tag{6.2}$$

where G is the shear modulus and ν is Poisson's ratio. The parameter $n=a/c$ describes the relative position of the crack tip to the grain boundary (see Figure 6.1), where a is the crack length and c is the crack length together with the length of the plastic zone extension ahead of the crack, which always extends to the grain boundary.

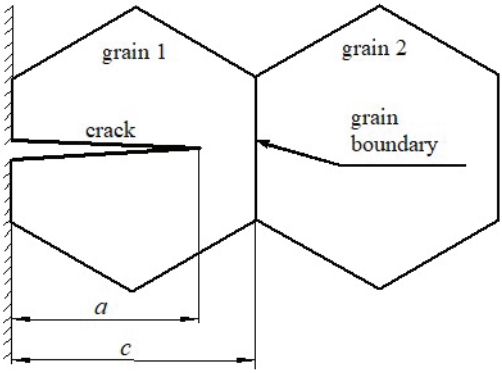

Figure 6.1. Short crack inside a crystal grain

The used computational model attempts to account for different parameters influencing the pitting process: Hertzian contact pressure, friction between contacting surfaces, EHD-lubrication, moving contact of gear flanks, fluid trapped in the crack, and, residual stresses due to heat treatment of the material. Computational simulation of the crack growth leading to pitting starts from the initial surface-breaking fatigue crack, which is a consequence of mechanical or heat treatment of the material as well as a consequence of the running in process.

6.2 Parameters influencing fatigue crack growth

6.2.1 Normal and tangential contact loading

The normal contact pressure between meshing gear flanks has been determined using the Hertzian theory [6.19], where the distribution of normal contact pressure $p(x)$ can be determined analytically by (see Figure 6.2):

$$p(x) = \frac{2F_N}{\pi b^2}\sqrt{b^2 - x^2} \tag{6.3}$$

where F_N is the normal force per unit gear width and b is the half-length of the contact area, which can be determined from

$$b = \sqrt{\frac{8F_N R^*}{\pi E^*}} \tag{6.4}$$

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

$$R^* = \frac{R_1 \cdot R_2}{R_1 + R_2} \tag{6.5}$$

$$E^* = \frac{2E_1 \cdot E_2}{E_2(1 - \nu_1{}^2) + E_1(1 - \nu_2{}^2)} \tag{6.6}$$

Here, R_1, E_1, ν_1 and R_2, E_2, ν_2 are the curvature radii, Young's modulus and Poisons ratio of the contacting cylinders, see Figure 6.2.

The maximum contact pressure p_0=$p(x$=0) can then easily be determined as

$$p_0 = \sqrt{\frac{F_N E^*}{2\pi R^*}} \tag{6.7}$$

The distribution of tangential contact loading $q(x)$ due to the relative sliding of the gear flanks is determined here by utilising Coulomb's law of friction:

$$q(x) = \mu \cdot p(x) \tag{6.8}$$

where μ is the coefficient of friction between the meshing gear flanks.

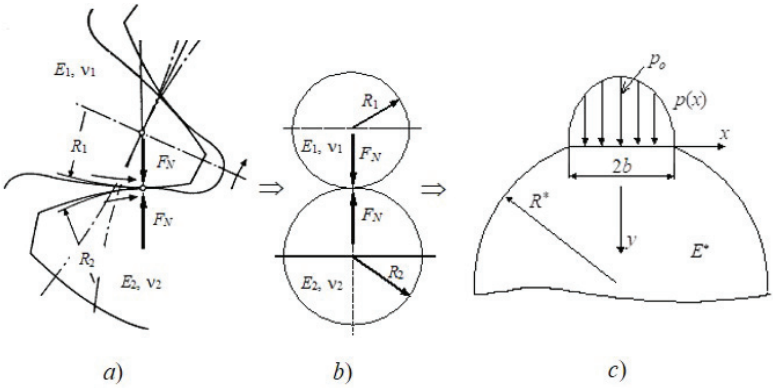

Figure 6.2. Schematic presentation of contact of two bodies (*a*) actual contact of gear flanks, (*b*) substitute model of two cylinders, (*c*) equivalent Hertzian contact model

6.2.2 Influence of the EHD-lubrication

In the proposed computational model, the normal contact loading distribution $p(x)$ also considers the influence of the Elasto-Hydro-Dynamic (EHD) lubrication conditions, which appear in the contact of lubricated gear flanks. Under such conditions, the contacting surfaces are separated with a thin lubricant film. However, the presence of the viscous lubricant in the contact area of the sliding surfaces affects the contact pressure distribution in a way that is illustrated in Figure 3. This pressure distribution can be determined experimentally or with the use of appropriate computational models [6.20, 6.21]. A significant pressure spike develops in the outlet contact region and strongly depends on the lubricant's pressure-viscosity characteristic. In computations reported herein the dimensionless pressure spike amplitude Y and dimensionless pressure spike location X (see Figure 6.3) have been determined using the following empirical equations [6.22]:

$$Y = 0.267W^{-0.375} \cdot U^{0.174} \cdot G^{0.219} \tag{6.9}$$

$$X = 1 - 2.469W^{-0.941} \cdot U^{0.206} \cdot G^{-0.848} \tag{6.10}$$

Here, W, U and G are dimensionless parameters that are determined as follows:

$$W = \frac{F_N}{R^* E^*} \tag{6.11}$$

$$U = \frac{\eta_o \cdot \mu}{R^* E^*}; \ \eta_o = \rho \cdot \nu \tag{6.12}$$

$$G = \alpha \cdot E^* \tag{6.13}$$

where F_N is the normal force per unit gear width, R^* is the equivalent radius, E^* is the equivalent Young's modulus, η_o is the dynamic viscosity at the atmospheric pressure, $u=(u_1+u_2)/2$ is the mean surface velocity and u_1 and u_2 are respective surface velocities, ρ is the lubricant density, ν is the kinematic viscosity and α is the pressure-viscosity coefficient.

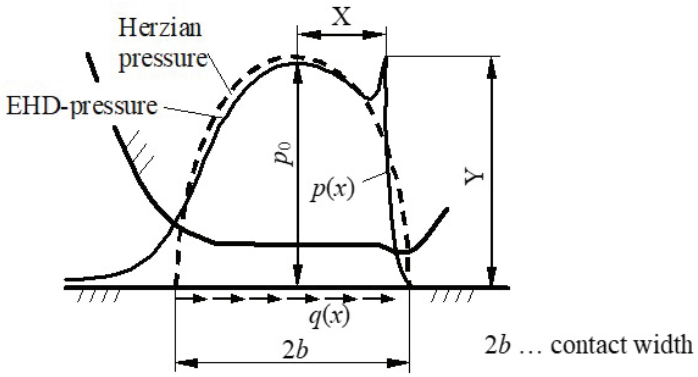

Figure 6.3. EHD contact pressure distribution

6.2.3 Influence of moving contact and lubricant trapped in the crack

For a more realistic simulation of fatigue crack growth, it is necessary to consider the moving contact of gear flanks. The moving contact can be simulated with different loading configurations (load cases), as shown in Figure 6.4. In all configurations, the normal $p(x)$ and tangential $q(x)$ contact loading distributions are of the same magnitude; however, they are acting at different positions with respect to the crack.

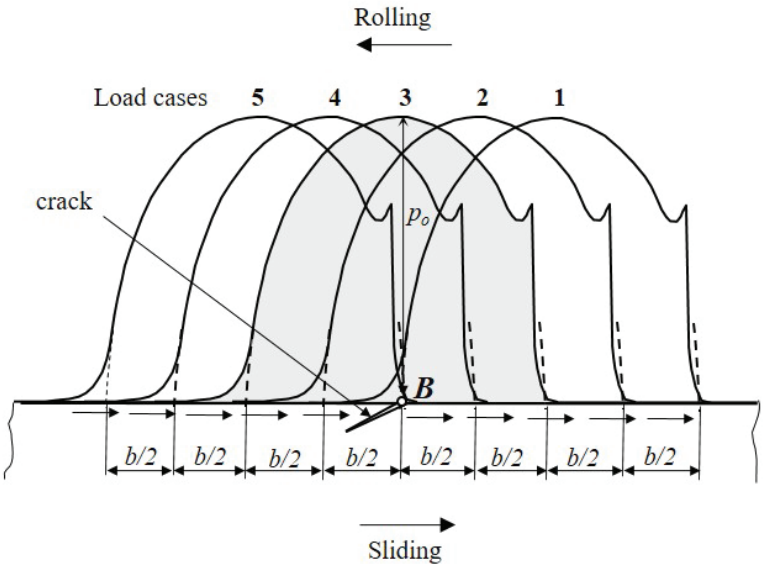

Figure 6.4. Moving contact simulated with different load cases

The simulation of surface initiated fatigue crack propagation should also consider the influence of lubricant pressure acting on the crack faces. The lubricant pressure is not constant but depends on the contact loading position, i.e., the contact pressure distribution position with respect to the crack, see Figure 6.5. For example, if the contact pressure p_2 is acting at the position of an existing surface crack, then the crack faces are also loaded with the same lubricant pressure; the same is valid for contact pressures p_3 and p_4. It should be pointed out that this effect does not appear in the case of subsurface initiated cracks.

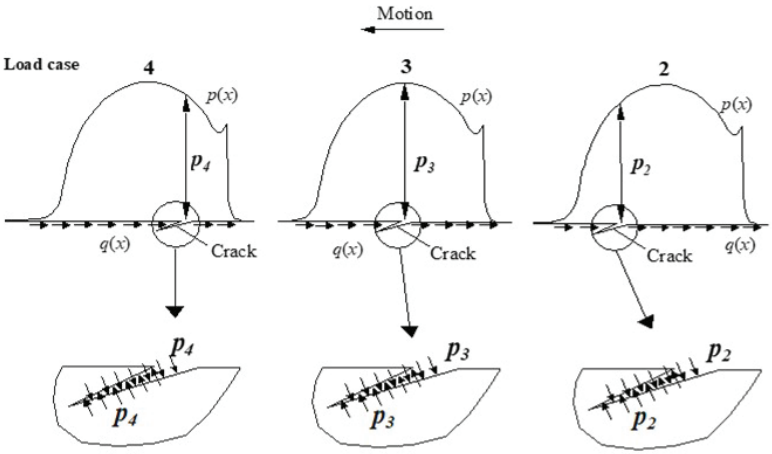

Figure 6.5. Lubricant pressure acting on the crack faces

6.2.4 Influence of residual stresses

To determine the residual stresses due to heat treatment of the gear teeth flanks, the analytical model described in [6.23] has been used. The model is based on the hardness distribution H along the depth D under the gear flank (see Figure 6.6):

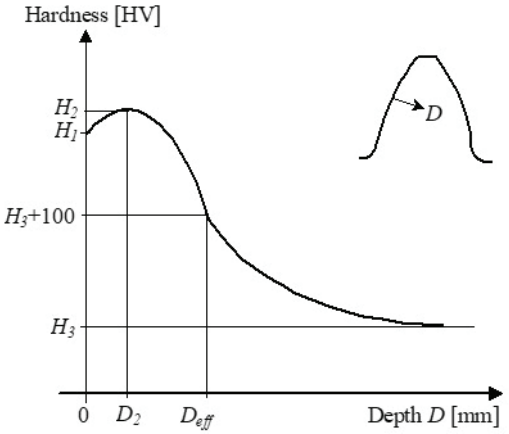

Figure 6.6. Hardness distribution H along depth D

$$H = (H_2 - H_3) \cdot \exp[-A(D - D_2)^2] + H_3 \tag{6.14}$$

$$A = -\frac{1}{{D_2}^2} ln\left[\frac{H_1 - H_3}{H_2 - H_3}\right] \qquad D \leq D_2 \tag{6.15a}$$

$$A = -\frac{1}{\left(D_{eff} - D_2\right)^2} ln\left[\frac{550 - H_3}{H_2 - H_3}\right] \quad D > D_2 \tag{6.15b}$$

where H_1, H_2 and H_3 are surface hardness, maximum hardness and core hardness, D_{eff} is the effective case depth with a hardness of 550 HV and D_2 is the depth of the maximum hardness. Based on the hardness distribution, the carbon content C (%) in the surface layer of the gear teeth flanks can be estimated as:

$$C = 0.103 \cdot \exp(3.5 \cdot 10^{-6} H^2) \qquad H \geq 700\ HV \tag{6.16a}$$

$$C = 0.131 \cdot \exp(3.0 \cdot 10^{-6} H^2) \qquad H < 700\ HV \tag{6.16b}$$

The specific volumes of martensite, austenite and gear material before heat treatment are then a function of the carbon content. The residual stresses are assumed to be caused only by the difference in the volume expansion of the material in the case and in the core, and can easily be estimated using the procedure described in [6.23].

6.3 Numerical simulation of fatigue crack growth

For the purpose of fatigue crack growth simulation, the virtual crack extension (VCE) method in the framework of the finite element method (FEM), has been applied. The VCE-method, as proposed by Hellen [6.24], is based on the criterion of released strain energy dV per crack extension da (G =dV/da), which serves as a basis for determining the combined stress intensity factor K around the crack tip for the plane strain conditions. The complete procedure for determining the stress intensity factor K using the VCE-Method is fully described in [6.5].

Assuming the validity of the maximum energy release criterion, the crack will propagate in the direction corresponding to the maximum value of G, i.e., in the direction of the maximum stress intensity factor K. The computational procedure is based on incremental crack extensions, where the size of the crack increment is prescribed in advance. For each crack

extension increment, the stress intensity factor is determined in several different possible crack propagation directions (see Figure 6.7) and the crack is actually extended in the direction of the maximum stress intensity factor, which requires local remeshing around the new crack tip. The incremental procedure is repeated until the crack reaches a critical length. Following the above procedure, one can numerically determine the functional relationship $K=f(a)$.

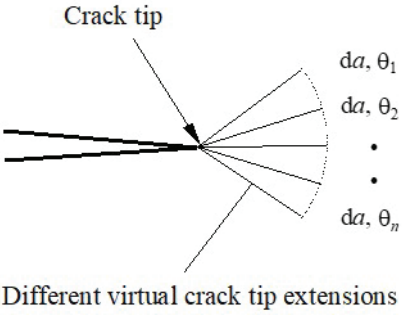

Figure 6.7. Virtual crack extensions of the crack tip

6.4 Practical application

The proposed model has been used for the simulation of surface fatigue crack growth on a spur gear pair, which has also been experimentally tested. The gear pair is made of carburized steel 16MnCr5 (according to the ISO standard) with Young's modulus $E=2.06\cdot10^5$ N/mm^2 and Poisson's 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), with the equivalent radius of gear teeth flanks $R^*=10$ mm.

6.4.1 Parameters influencing fatigue crack growth

Using eqs. (6.3) to (6.7), the half-length of the contact area of meshing gear flanks 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. (6.3). For all computations, the coefficient of friction $\mu=0.04$ has been used, which is the average value for well-lubricated gears [6.25]. Therefore, the tangential loading $q(x)$ has been determined using eq. (6.8).

The influence of EHD-lubrication on the normal loading distribution $p(x)$ has been estimated using eqs. (6.9) to (6.13) for the lubricant oil ISO-VG-220, with the kinematic viscosity $\nu_{40} = 220$ mm^2/s, density $\rho_{15} = 0.9$ kg/dm^3 and pressure-viscosity coefficient $\alpha = 0.18 \cdot 10^{-7}$ m^2/N. The mean surface velocity of the contacting surfaces has been taken as a constant value u=5 m/s, which is a common value for gears [6.26]. Using these parameters, the dimensionless pressure spike amplitude Y and the dimensionless pressure spike location X (see Figure 6.3) are equal to X= 0.9462 and Y= 0.8146, respectively.

The hardness distribution H and the carbon content C (%) in the surface layer of the gear teeth flanks have been determined using eq. (6.14) to (6.16), where the following values have been measured previously: H_1=765 HV, H_2=770 HV, H_3=450 HV, D_{eff}=1.25 mm, D_2 =0.1 mm. The distribution of residual stresses in the surface layer has then been estimated using the analytical model described in [6.23], and is shown in Figure 6.8 for 10 % retained austenite on the surface. In the numerical model, the calculated residual stresses are approximated by appropriate thermal loading.

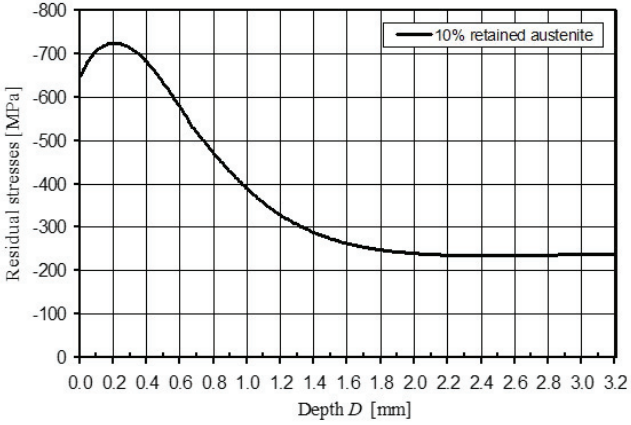

Figure 6.8. Distribution of residual stresses

6.4.2 Numerical simulations

The finite element mesh shown in Figure 6.9, and the boundary conditions as described above, have been used in the subsequent numerical analyses.

The simulation of the pitting phenomenon has been studied for surface and subsurface initiated fatigue crack growth.

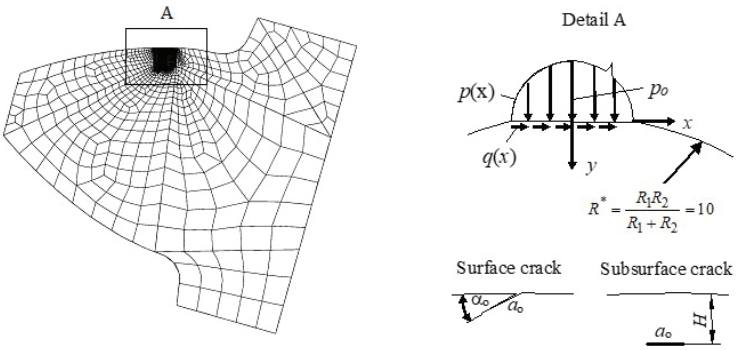

Figure 6.9. FE discretisation and configuration of the initial crack

Numerical simulation of surface initiated fatigue crack growth.

For the configuration of the initial crack on the surface 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 $\alpha_o = 22°$ [6.27]. In numerical computations, the crack increment was of size Δa = 1.5 μm. The stress intensity factor K was estimated in each crack increment for 30 different virtual crack tip extensions. Five different loading configurations have been considered in each computation for the purpose of simulating the effect of the moving contact of contacting surfaces (see Figure 6.4). For each crack increment, the crack was actually extended in the direction of the recorded K_{max} from all calculated load cases. 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 as micro pitting. But it is not the final and most critical surface failure. Further operation of the mechanical components results in the formation of larger pits, and consequently progressive pitting. In this respect, similar numerical simulations as described above have been continued for two more steps, as shown in Fig. 6.10. In these calculations, it was assumed that the initial crack of length a_o=7.5 μm started from the bottom of the existing surface pit. For all computations, the crack increment was Δa=2 μm.

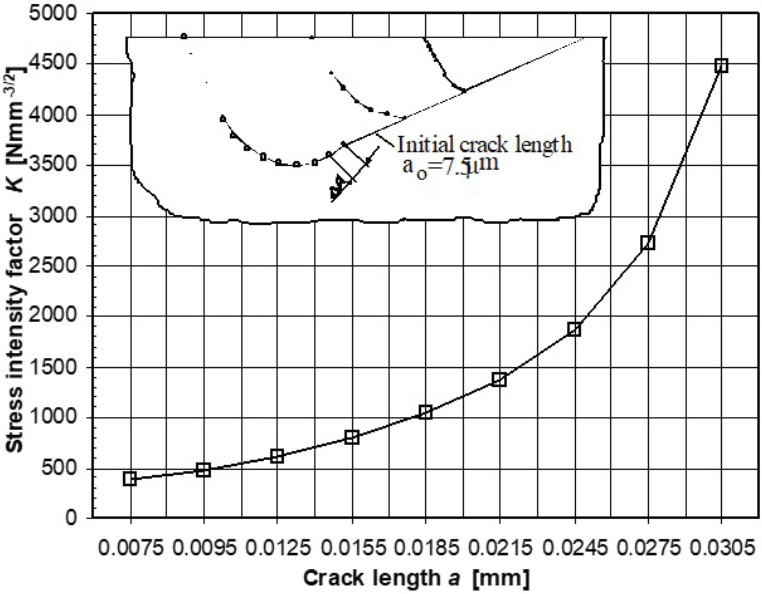

Figure 6.10. Stress intensity factor for crack propagation from an existing micro pit

Numerical simulation of subsurface initiated fatigue crack growth.

An initial crack length of a_o=D= 0.05 mm has been positioned at the point of maximum equivalent stress, i.e., at the depth of H=0.192 mm under the contact surface, see Figure 6.9. This follows the assumption that in mechanical components with smooth surfaces and good lubrication, the largest contact stresses appear at a certain depth below the surface. Therefore, fatigue cracks here are usually initiated under the surface along a persistent slip band inside a crystal grain. In this case, the average length of the initial crack is equal to the grain size D, i.e., a_0=D.

Numerical analyses have shown that upon initial crack appearance, the stress intensity factor is much higher in the crack tip, which coincides with the direction of applied friction force, which is the right crack tip for the treated case, see Figure 6.11. Therefore, it is assumed that the crack will start to propagate in this direction. At the present, the computational model can follow only the growth of a single crack tip and it was necessary to

assume that the crack will grow only in one direction until it reaches the surface.

Figure 6.11 illustrates the simulation of fatigue crack growth for the data set p_o=1550 MPa and R^*=10 mm. The local FE discretization around the crack is shown for some typical crack extensions. Step 1 shows the initial crack, while steps 4-11 illustrate the right crack tip propagation towards the surface. During the FE analysis, the computed SIF is very small at the beginning but later increases as the crack propagates towards the contact surface. At the moment when the crack breaks through to the contact surface, the SIF in the other crack tip exceeds the critical SIF for high quality steels. This implies that when the subsurface crack first reaches the contact surface, the corresponding crack length can be taken as the critical crack length a_c. The functional relationship between the SIF and crack length is shown in Figure 6.12.

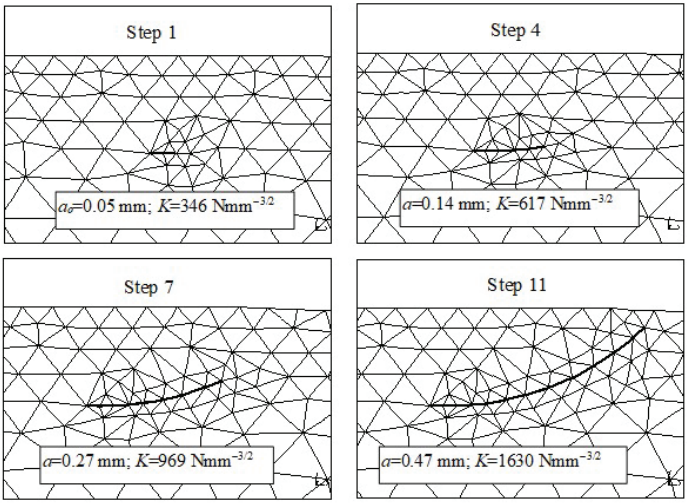

Figure 6.11. Numerical simulation of subsurface initiated fatigue crack growth

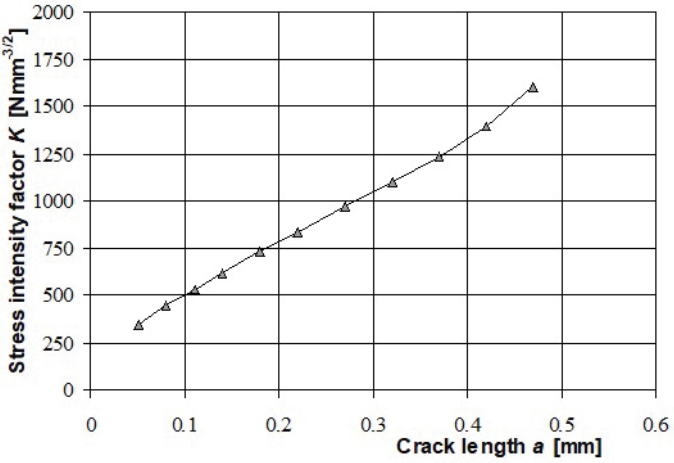

Figure 6.12. Stress intensity factor for subsurface initiated crack growth

6.4.3 Comparison between numerical and experimental results

Experimental testing of the spur gear pair has been performed on a FZG-pitting test machine according to the DIN 51354 standard. The tested gears have been subjected to the same operating conditions and loading parameters as used in the numerical computations. Figure 6.13 shows that the shape and magnitude of numerically determined pits correspond well with available experimental data [6.26, 6.28].

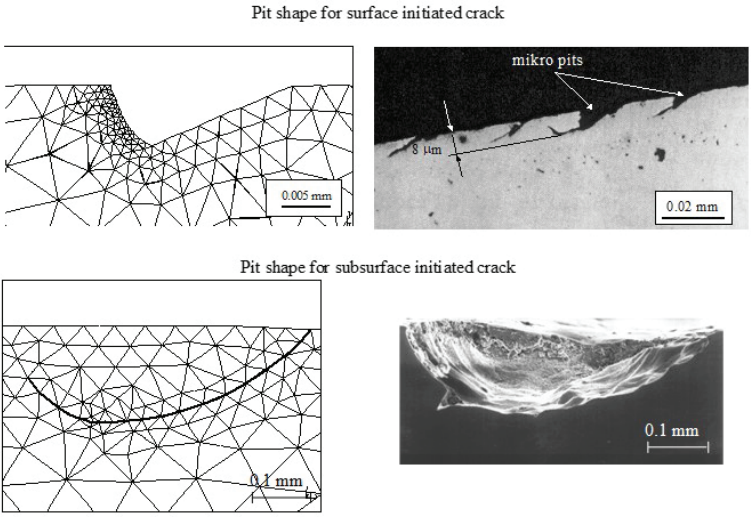

Figure 6.13. Comparison between numerically and experimentally obtained pit shapes

6.5 Conclusions

Rolling contact fatigue cracks are one of the most common forms of failure in gears, railway tracks and bearings. There are two different types of contact fatigue crack: the crack may either be initiated at the contacting surfaces, and thereafter propagate at a shallow angle to the surface; alternatively, the cracks may be initiated at large non-metallic inclusions below the surface, in the region of the maximum cyclic shear stress. In both cases, the cracks continue to propagate in this region before either detaching lumps of surface material, or changing direction to cause the final failure.

A computational model based on fracture mechanics has been developed to demonstrate the concept of predictive damage modelling for contacting mechanical components like gears, bearings, wheels, etc. A simple equivalent model with applied Hertzian boundary conditions has been used for simulation of surface and subsurface initiated fatigue crack growth under rolling–sliding contact conditions. The model also considers EHD-lubrication conditions, the moving contact of mechanical elements, residual stresses due to heat treatment of contacting mechanical elements, and for surface initiated cracks, the fluid trapped in the crack.

Simulation of surface and subsurface fatigue crack propagation from the initial crack up to the formation of the surface pit, and the dependence of the stress intensity factor on crack length, are determined by the finite element method, where the required functional relationship between the stress intensity factor at the crack tip and the crack length is determined using the virtual crack extension method. Based on the computed relationships between the stress intensity factor and crack length, and through a consideration of some particular material parameters, the service life of a mechanical component can then be determined using the theory of short fatigue crack growth as described in Section 6.1.

The treated gear pair has also been experimentally tested according to the appropriate standardized procedure. Comparison of numerically predicted and experimentally recorded pit shapes shows that they are in very good agreement. However, the model could be further improved by means of additional theoretical, numerical and especially experimental research, since it relies mostly on experimentally determined material parameters.

References

[6.1] ISO 6336, Calculation of load capacity of spur and helical gears, International Standard, 2019.

[6.2] Podrug S, Glodež S, Jelaska D, Numerical Modelling of Crack Growth in a Gear Tooth Root, *Strojniški vestnik*, 57 (2011) 579–586.

[6.3] Glodež S, Winter H, Stüwe HP, A fracture mechanics model for the wear of gear flanks by pitting, *Wear*, 208 (1997) 177–183.

[6.4] Glodež S, Flašker J, Ren Z, A new model for the numerical determination of pitting resistance of gear teeth flanks, *Fatigue & fracture of engineering materials & structures*, 20 (1997) 71–83.

[6.5] Glodež S, Ren Z, Flašker J, Simulation of surface pitting due to contact loading, *International journal for numerical methods in engineering*, 43 (1998) 33–50.

[6.6] Keer LM, Bryant M.D, A pitting model for rolling contact fatigue, *ASME Jour. Lubr. Tech.*, 105 (1983) 180–205.

[6.7] Miller GR, Keer LM, Cheng HS, On the mechanics of fatigue crack growth due to contact loading, *Proc. Roy. Soc. London*, 397 (1985) 197–209.

[6.8] Zhou RS, Cheng HS, Mura T, Micropitting in rolling and sliding contact under mixed lubrication, *ASME J. Tribology*, 111 (1989) 605–613.

[6.9] Xu X, Lai J, Lohmann C, Tenberge P, Weibring M, Dong P, A model to predict initiation and propagation of micro-pitting on tooth flanks of spur gears, *International journal of fatigue*, 122 (2019) 106–115.
[6.10] Li S, Kahraman A, A micro-pitting model for spur gear contacts, *International journal of fatigue*, 59 (2014) 224–233.
[6.11] Zhang B, Liu H, Zhu C, Li Z, Numerical simulation of competing mechanism between pitting and micro-pitting of a wind turbine gear considering surface roughness, *Engineering failure analyses*, 104 (2019) 1–12.
[6.12] Sekerciogla T, Kovan V, Pitting failure of truck spiral bevel gear, *Engineering failure analyses*, 14 (2007) 614–619.
[6.13] Moorthy V, Shaw BA, An observation on the initiation of micro-pitting damage in as-ground and coated gears during contact fatigue, *Wear*, 297 (2013) 878–884.
[6.14] Aslantas K, Tasgetiren S, A study of spur gear pitting formation and life prediction, *Wear*, 257 (2004) 1167–1175.
[6.15] Shahani AR, Davachi R, Babaei M, The crack propagation path under multiple moving contact loads in rolling contact fatigue, *Theoretical and applied fracture mechanics*, 100 (2019) 200–207.
[6.16] Ancellottti S, Benedetti M, Dallargo M, Fontanari V, The role of the second body on the pressurization and entrapment of oil in cracks produced under lubricated rolling-sliding contact fatigue, *Theoretical and applied fracture mechanics*, 91 (2017) 3–16.
[6.17] Sun Z, de los Rios ER, Miller KJ, Modelling small fatigue cracks interacting with grain boundaries, *Fatigue Fract. Engng Mater. Struct*., 14 (1991) 277–291.
[6.18] Navarro A, de los Rios ER, Short and long fatigue crack growth, *Philosophical Magazine A*, 57 (1988) 15–36.
[6.19] Johnson K L, *Contact mechanics*, Cambridge University Press, 1985.
[6.20] Williams J A, *Engineering tribology*, Oxford University press, 1994.
[6.21] Oster P, Simon M, Messung und Berechnung von Gleit-Wälzkontakten an Scheiben und Zahnrädern in Bereich der Elastohydrodynamik, *Antriebstechnik*, 27 (1988) 51–56 (in German).
[6.22] Hamrock BJ, Lee RT, Houpert LG Parametric study of performance in elastohydrodynamic lubricated line contact, *Fluid film lubrication–Osborne Reynolds centenary* (1987), 199–206.
[6.23] Tobe T, Kato M, Inoue K, Takatsu N, Morita I, Bending Strength of Carburized SCM42OH Spur Gear Teeth, *JSME*, 29 (1986) 273–280.

[6.24] Hellen TK, On the method of virtual crack extensions, International Journal for Numerical Methods in Engineering, 9 (1975) 187–207.
[6.25] Winter H, Knauer G, Einfluss von Schmierstoff und Betriebstemperatur auf die Grübchentragfähigkeit einsatzgehärteter Zahnräder, Antriebstechnik, 29 (1990) 65–84. (in German).
[6.26] Glodež S, The fracture mechanics model of gear flanks fatigue, Ph.D thesis, Faculty of Mechanical Engineering, University of Maribor, 1996.
[6.27] Knauer G, Zur Grübchentragfähigkeit einsatzgehärteter Zahnräder, Ph.D. thesis, TU Munich, 1988 (in German)

CHAPTER 7

FUTURE CONSIDERATIONS OF FRACTURE AND DAMAGE MECHANICS

7.1 Introduction

In all mechanical systems, mechanical failure may occur for many different reasons. Environment, loading and stress, defects in materials and poor-quality design are counted as causes of mechanical failure. Thus, study, research and investigation on mechanical failures of parts and structures seem necessary to avoid extra cost and downtime. Successful early applications of fracture mechanics [7.1] [7.2] [7.3] supported this new field in the engineering community. During the last three decades, demands on reliability and safety of technical systems are significantly growing. This led to many types of research in the area of fault detection, failure analysis, and prediction of remaining service life. Although fatigue crack propagation and fracture cause a large part of failure events in industrial practice, fracture mechanics in failure analysis seems to be still a side issue. Much more pronounced than in the design stage, the benefit of fracture mechanics in failure analysis depends on its accuracy. This is limited by both, intrinsic factors of the method and the availability and quality of the input information.

The loss of integrity and adverse effect on mechanical properties can be concluded as attributing micro/macro-mechanics damage in structures, especially in composite structures or porous materials. Damage as a progressive degradation of material continuity in engineering predictions for any aspects of initiation and propagation, requires to be identified by a trustworthy mechanism to guarantee the safety of structures. Beside the materials design, structural integrity and health are usually prone to be monitored clearly. One of the most powerful methods for the detection of damage and the prediction of service life is artificial intelligence (AI) [7.4] [7.5] [7.6]. This chapter presents state-of-the-art AI methods and their

applications in structural damage and prediction. Popular AI methods are identified, and the performance and future trends are discussed.

7.2 The history of AI in fracture and damage mechanics

Artificial intelligence has been studied for decades and is still one of the most elusive subjects in computer science. This is partly due to how large and nebulous the subject is. AI ranges from machines truly capable of thinking to search algorithms used to manage our lives. It has applications in nearly every way in which we use computers in the society.

A Greek myth speaks of Prometheus who earned the wrath of Gods of Olympus when he sought to steal for the benefit of the human race, not only fire, but also the gift of intelligence or *nous* (the rational mind). A human effort to gain knowledge constitutes a transgression against Gods and is deeply ingrained in the Western civilization thought; see, for example, works of Dante, Milton, and Shakespeare. The belief that the desire for knowledge must ultimately leads to a disaster has survived the Renaissance, the Age of Enlightenment, and even the scientific and technological advances of the 19th and 20th centuries. Instead of dispelling the ancient fear of intelligence and knowledge, a modern technology has only made these consequences appear more imminent. The legend of Prometheus has been retold many times in the language of a technological society. Thus, it is not surprising that artificial intelligence is the subject of controversy in academic, intellectual, and popular circles.

Although potentially extremely interesting and rewarding, a detailed study of the history of scientific and philosophical thought leading up to a current research in artificial intelligence is beyond the scope of this book. However, it is essential to have at least a glimpse of the history because it is impossible to appreciate where we are without some knowledge of how we got there.

7.2.1 Philosophical landmarks of AI

- Aristotle (384–322 B.C.) distinguishes matter from form (e.g., a sculpture of Aristotle is made from the material bronze and has the form of a human), thereby laying the seeds of abstracting the medium from its representation which is at the heart of modern computer science.
- 350 B.C. – Panini develops a formal grammar for Sanskrit laying the foundations of syntactic models that led to Chomsky's theory of syntactic structures in 1956.

- 825 – Al Khowarizmi introduces to the west the eastern mathematical tradition which largely consists of mathematical recipes, i.e., algorithms in his text explaining the Indian system of numeration, which was translated into Latin under the title *Algorithmi de numero Indorum*, as well as the Arabic algebra.
- Hobbes (1588–1679) proposes that thinking is a rule-based computational process analogous to arithmetic.
- Descartes (1596–1650) discounts sensory experience as untrustworthy and justifies his own existence in terms of thought: *Cogito ergo sum* (I think, therefore I am). He establishes the notion that the structure of ideas about the world is not necessarily the same as the structure of their subject (matter); an idea which underlies much of the methodology of AI, epistemology, psychology, mathematics, and modern literature.
- Leibniz (1646–1716) seeks a general method by which all truths will be reduced to a kind of calculation.
- Boole (1815–1864) puts forth his study of logic and probability as an investigation into the laws of thought; as the inventor of what is now called the Boolean logic, which became the basis of the modern digital computer, he is regarded as the founder of a field called computer science [7.7].

7.2.2 AI is industry (1980–present)

- Mid 1980s – Some of the failures of rationalist/logical approaches to AI lead to a renewed interest in biologically inspired neural networks and evolutionary models, which, in turn, lead to modest successes at solving some old problems (e.g., pattern recognition).
- Mid-1990s – A progress in algorithmic models of learning begins to offer promising and practical alternatives to the knowledge engineering, and AI technologies begin to be used in critical components of large software systems. The most successful approaches incorporate elements of both, the rationalist/logical/symbolic tradition and the existential/phenomenological/non-symbolic tradition. Proposals for reconciling the two approaches begin to appear; the maturity of several subfields of AI, such as vision, language processing, knowledge representation, planning, etc., leads to insights on the capabilities as well as limitations of the techniques that were developed, and redirects the attention to the problem of building intelligent agents as opposed to subsystems.
- 1982 – McDermott develops the DEC R1 expert system.

- 1980 – Neural networks are back: Werbos introduces the backpropagation algorithm, an abbreviation for "backward propagation of errors", which is a common method of training artificial neural networks – a network learns to identify a desired output from exemplary inputs, similar to the way a child learns to identify a dog from examples of dogs.
- 1984 – The Internet is rolled out.
- 1989–1991 Berners-Lee invents the world wide web (www).
- Mid 1990s–present – AI technologies continue to find applications in adaptive information retrieval, data mining and knowledge discovery from databases, customizable software systems, smart devices (e.g., home appliances), agile manufacturing systems, autonomous vehicles, healthcare systems, medical informatics, etc. A slow but steady progress in fundamental AI research problems continues as part of the fourth industrial revolution (Industry 4.0) and the super-smart society (Society, 5.0) [7.7].

7.2.3 AI as science today

A synthesis of traditional logic-based systems, soft and adaptive computing technologies (e.g., neural networks, probabilistic models, fuzzy logics, etc.), and a synthesis of software agents and multi-agent systems leads to the emergence of nouvelle AI which views the intelligence as an emergent behaviour resulting from interactions (e.g. a communication, coordination and competition) among large numbers of autonomous or semi-autonomous entities (neurons, computer programs, or individuals) that are situated in the world. Topics of its active research are displaying a structure and organization at multiple spatial and temporal scales, interacting with the world through sensors and effectors, and hosting fundamental computer programming problems, such as designing individual agents in terms of their communication, coordination, and organizations [7.8].

The long-term goal of research in the field of AI, which still today seems unattainable, is to create an artificial system that will reach or even exceed human intelligence through independent learning. Here is Minsky's definition of AI: "Artificial intelligence is the *science of making machines* do things that would require intelligence if done by men." [7.9] AI expects an intelligent system (e.g., a computer, machine, robot) to be extremely intelligent (or capable) not only in one aspect of the intelligence. It is expected to be complex and intelligent in all fields of intelligence needed by human when solving problems: perception, reasoning, inference, decision-making, learning, judging, evaluating, etc. Research in the field of

AI involves developing a system that operates more or less intelligently and is capable of indenpendently solving relatively demanding problems. These methods are often based on *imitating human problem solving*. AI areas, except for machine learning and deep learning, are related to knowledge presentation, speech recognition, automated deduction and theorem proving, logic programming, qualitative and cognitive modelling, game playing, expert systems, heuristic problem solving, artificial perception and robots [7.10].

Machine learning algorithms play an important role in all AI areas. By using learning techniques, the system is able to learn and improve its perception (it becomes adaptive), speech recognition, deduction etc. [7.11]. The field of logic programming is also closely related to inductive logic programming used in the development of logic programmes, e.g. for destination settings (such as GPS devices). Machine learning can be used for developing expert systems to create databases from examples of problem solving. Intelligent robots or cybernetic physical systems have to improve their procedure for problem solving with the help of learning. Finally, cognitive modelling cannot take place without using learning algorithms.

Ever since the invention of electronic computers, no significant progress has been noted in pursuing the ultimate goal – *to create an intelligent machine on the basis of machine learning algorithms*. Nowadays, the following are important steps (among others) toward this goal:

- Lenat's Automated Mathematician – an interesting system for discovering new concepts in mathematics.
- A great success of computers in the field of complex games, such as checkers, chess, or backgammon.
- Artificial neuron networks for modelling cognitive processes in the brain.
- Anderson's ACT-R modular cognitive model of human brain processes.

Therefore, the question is, have machines already crossed boundaries drawn between them and humans? According to the results of current research, machine intelligence falls within the domain of systems, which were designed by developers for end users. However, this is, for the most part, a poorly researched area, and what has been investigated (results included) is usually regarded as a highly protected secret and not readily available [7.12] [7.13].

Intelligent approaches and systems have been applied in a wide range of industries and commercial fields. Expert systems performed in certain limited domains such as failure analysis and mechanical fault diagnosis. Three decades ago, the proficiency of the AI approach and particularly the expert system for solving problems in the area of fracture mechanics, has been discussed [7.4]. Researchers described an idea of creating an expert system in the domain of linear elastic fracture mechanics and exhibited the validity of the technique.

7.2.4 Components of artificial intelligent systems

Any computer-assisted tool, especially an artificial intelligent system (AIS), must consist, in general, of two parts:

1. *Heuristics*. The term generally refers to experience-based techniques for problem solving, learning, and discovery. Heuristic methods are used to speed up the process of finding a satisfactory solution, where an exhaustive search is impractical. Examples of this method include using a “rule of thumb”, an educated guess, an intuitive judgment, or common sense. More precisely, heuristics are strategies using readily accessible, though loosely applicable, information to control problem solving in human beings and machines [7.14].
2. *Epistemology*. The term refers to the branch of philosophy concerned with the nature and scope (limitations) of knowledge. It addresses the following questions: What is knowledge? How do we know what we know? How is knowledge acquired? The kind of knowledge usually discussed is *propositional knowledge*, also known as "knowledge that" which differs from "knowledge how". For example, I know *that* 2 + 2 = 4, but I also know *how* to add 2 and 2. Some philosophers think there is an important distinction between "knowing that" and "knowing how", with epistemology primarily interested in the first [7.9] [7.5].

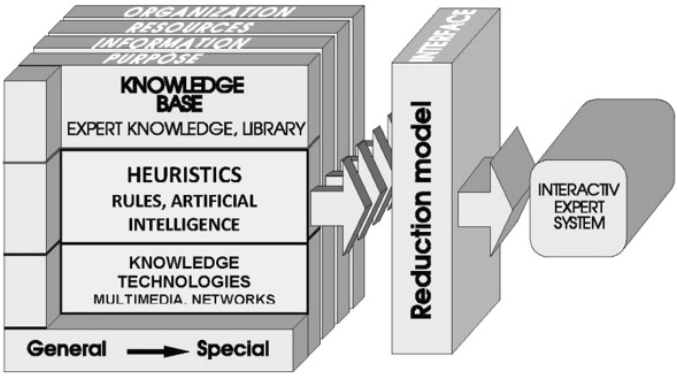

Figure 7.1: Configuration of an expert system [7.4].

At the concrete level, AIS are interactive computer programs, which incorporate expertise and provide advice on a wide range of tasks, typically consisting of the following three basic components [7.4] [7.5]:

- *behaviour* of the problem domain;
- *context* as a workspace for the problem (constructed by the inference mechanism from the information provided by the user and the knowledge base);
- *inference mechanism* (monitoring the execution of a program by using the knowledge base to modify the context).

In addition, AIS should also be provided with a user-friendly communication interface, as shown in Figure 7.1:

- user interface;
- explanation facility;
- knowledge acquisition module.

Although AI has a long history, it is constantly used and actively grown. Over the years, AI applications indicated not only that AI can contribute major aspects to mechanical engineering, but also that investigation of AI on various domains can develop and improve in fracture mechanics. AI methods are applied in this field to reduce the failures and solve the problems. However, there are still unsolved problems and improvement opportunities.

Investigations on fracture mechanics, including performing experimental tests and computational methods are not always an easy task and are quite time consuming, but also a high level of technical expertise is required. Indeed, fracture mechanics and failure analysis are one of the extreme complex engineering fields. There are many challenges for detecting failures and faults of mechanical machinery, parts and systems. The main issues based on the existing works in fracture mechanics domain is as follows:

- identifying failure mode,
- identifying failure mechanism,
- monitoring of fault diagnosis process,
- monitoring of damage and failure diagnosis process,
- early detection of faults,
- early detection of damages and failures,
- predicting of fracture parameters and service life.

Therefore, in order to achieve the goals and objectives set out in this book, keywords have been searched for in scientific databases, based on these challenges in fracture mechanics. According to these challenges, different AI methods are able to support solving the complex mechanical fracture problems. Material toughness characterisation, damage detection, fault diagnosis, failure mode identification and crack detection are various aspects of fracture mechanics, which are surveyed by different approaches of AI. To address these issues, an analysis of literature was performed, based on a classification of only three earlier methods and techniques of AI, which have also been used in this research:

- artificial neural networks (ANN),
- genetic algorithms (GA), and
- fuzzy logic (FL).

In this chapter, the main concepts and mechanisms of these methods are briefly highlighted. Moreover, it is explained how they are applied in sub-domains of fracture mechanics: failure mode and failure mechanism identification, fault and error detection and diagnosis, damage and failure detection and diagnosis, mechanical fracture and fracture parameters, and calculation of service life or remaining service life. This contribution could provide some ideas on future perspectives of research in this field.

7.3 Artificial intelligence approaches in fracture mechanics

During the past decades, many structural failures have been predicted by the fracture mechanic approach. Research on fracture and failure showed that saving time and money is possible by focusing on two major areas, which are materials and structures. In fact, predicting and preventing failure load have been interesting topics of research in the past years [7.4] [7.15]. There are different AI methods applied in mechanical engineering, while methods presented here are implemented in the domain of fracture mechanics.

For technology in mechanical engineering, not only the supervision of processes, but also the study, investigation and analysis of different sub-domains of engineering fracture mechanics of the systems, are important. These parameters are able to improve process and system performance. As it can be seen in Figure 7.2, the three mentioned AI methods are reviewed which are applied in sub-domains of fracture mechanics.

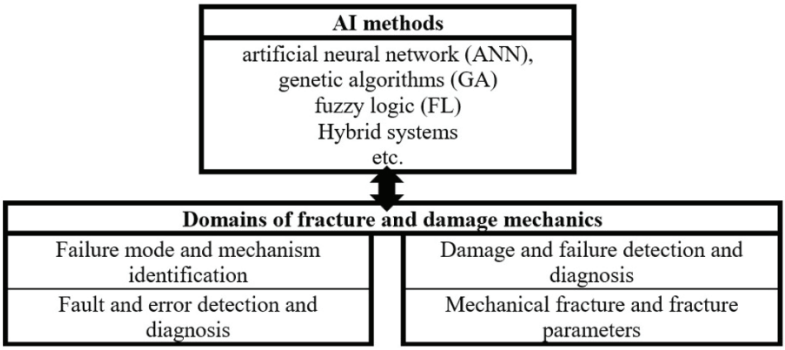

Figure 7.2: AI applications in fracture mechanics.

The four sub-domains of fracture mechanics according to Figure 7.2 have been actively researched over the past years, and they are also the research domain of AI applications. The definitions of these sub-domains are as follows:

- *Failure mode and failure mechanism*: a mechanical failure mode is defined as a physical process or processes that take place or combine

their effects to produce a failure [7.17] In this regard, failure mechanism is defined as a detailed description of a failure mode.

- *Damage and failure*: mechanical damage is a physical characterization that presents an existing defect on the object. Damage can occur on a micro-, meso- or macro-scale and can lead to a mechanical failure.
- *Fault and error*: in mechanical components or systems, fault and error are defined as the loss of ability to perform the required mechanical action. In other words, they are a deviation from the normal (expected) value and an occurred condition.
- *Mechanical fracture and fracture parameters*: fracture mechanics deal with solid containing planes of displacement discontinuities (cracks) [7.18]. Fracture is the separation or fragmentation of a solid body into two or more parts [7.19]. Several parameters such as crack propagation, fracture energy, fractography and fracture toughness are explored as part of mechanical fracture.

In this chapter, three most promising AI methods are reviewed which are used in the above mentioned sub-domains of engineering fracture mechanics.

7.3.1 Artificial neural network

ANN is a computational model. Marvin Minsky defined AI as follows: "Artificial intelligence is the *science of making machines* do things that would require intelligence if done by men." [7.9]. Over the past few decades, various types of ANNs have been applied in different areas of science and numerous other systems, including fault detection and failure analysis. An ANN contains three main sections, an input layer, a hidden layer and an output layer, which represent input parameters, the learning process and a solution of the problem, respectively.

The ability of ANNs to learn directly from examples is one of their most important properties. On the other hand, ANNs are able to regain and extract information from noisy, incomplete or poor data. All these properties make ANNs a powerful method for solving problems in the area of fracture mechanics. Nowadays, ANNs can be considered as a mature branch in computational science and engineering with lots of publications and commercial software.

As a branch of AI, ANNs are introduced and applied in various mechanical engineering problems, including in fracture mechanics. ANNs deal with

problems, which are difficult to analyse with the help of other methods. Applications of ANNs in fracture mechanics include optimization, identification and damage detection.

7.3.1.1 Artificial neural networks in failure mode and mechanism identification

ANNs can be used in the prognosis of fracture in steel cold forming operations. They receive input data on stress and strain histories and predict fractures for a wide range of bulk forming operations. The systems are trained by stress and strain data, and the prediction of fracture occurrence is defined as an output of the system. With the aim of best performing architecture, an ANN can be carried out in a MATLAB environment. Different configurations can be tested for the network to obtain a satisfactory result, consisting of different hidden layers. Such architecture is characterised by a successful training performance, which means square error is very little. The MATLAB ANN toolbox produces results with low values of mean error, which proves that the ANN method is effective in predicting failure load [7.10].

7.3.1.2 Artificial neural networks in fault and error detection and diagnosis

ANNs have been met with acceptance in many fields for modelling complicated real-world problems; an example is a neural network for fault detection. Numerous mechanical failures on machinery parts often show warnings in advance. These warnings are usually physical conditions that specify that a failure is near to occur. Examples of such warnings in ANN-based system for fault and damage detection include changes in natural frequencies and mode shapes. Based on the conducted research, it can be concluded that the most significant advantage of ANNs in comparison to other methods is their capability to predict crack growth direction in the construction under the influence of two or more cracks. Another advantage is that with the use of an ANN, it is not necessary to make prior assumptions about the behaviour of the concerned material. In addition, in structural damage identification, the use of ANN has shown that local damage is better identified when a structure is divided into sub-structures and each sub-structure analysed separately.

Many types of machinery use rotary parts such as different types of bearings, gears, motors and shafts. Over the past few decades, ANNs have been extensively adapted for use in fault diagnosis of rotating parts and

machinery. Numerous applications of ANNs include Back Propagation Feed Forward (BPFF), Feed Forward Neural Network (FFNN), Back Propagation Neural Network (BPNN), Multi-Layer Perceptron (MLP), Radial Basis Function Networks (RBF), Recurrent Neural Network (RNN) and Learning Vector Quantization (LVQ). All of these neural network methods are applied for fault detection of rotating parts and machinery.

7.3.1.3 Artificial neural network in damage and failure detection and diagnosis

ANNs are often used for damage detection in different structures. In [7.20], vibration data is analysed by an ANN in order to detect the location and severity of damages in beam-like structures. In fact, natural frequencies and mode shapes are used as the input data for the ANN to predict damage. Modelling and simulation is performed and vibration data is analysed, and these are first three natural modes for damaged and undamaged beams. To find the most effective ANN, different neural networks can be designed using, for example, the MATLAB toolbox. To check the robustness, artificial random noise is generated numerically and added to the noise-free data during the training of the ANNs. In the experiments, modal parameters are obtained by surface-bonded strain gauges, which are connected at the tip of the steel beam. An increase in the noise level leads to more accurate results in the prediction of damage location in comparison to prediction on the severity of damage. Moreover, experimental procedures and measuring devices play an important role in the prediction of the location and severity of the damage. The application of ANN for damage detection in beam like structures has been developed by Li, He, Ji, Wang & Hao [7.21]. The proposed system predicts the location and severity of the crack damage.

In a multiple crack analysis performed by an ANN, not only material properties (for instance, the modulus of elasticity or yield strength), but also a wide range of crack position parameters (such as crack size and crack offset distance), can be used as input data. The network structure and its operational parameters are selected mostly by using the trial and error method. It has been demonstrated that best predictions made by ANNs occur when a sigmoid activation function with two hidden layers is used. Although different fracture criteria for determining the direction of crack initiation exist, results obtained from experiments with ANNs show good agreement with experimental results. ANNs also employ the finite element model (FEM), and a combination of natural frequencies and mode shapes as a dataset for training and as an input parameter. Modal testing and

experimental modal analyses can be performed on both, undamaged and damaged structures. For example, the ratio of damage depth to the height of the beam, and the ratio of damage location to the length of the beam, are defined as damage severity and damage location indices, respectively. Both, mean square error and absolute error, are considered as an evaluation of the designed ANN.

In general, ANNs are a feasible and efficient technique for damage identification on different structures.

7.3.1.4 Artificial neural network in mechanical facture and fracture parameters

Due to catastrophic failures, mechanical engineering design criteria are developed based on fracture toughness. Advances aim to increase safety in the design of structures by predicting fracture toughness in specific working conditions. Fracture toughness is a fundamental material property, which is influenced by the microstructure of the material. The study, investigation and prediction of fracture toughness in different materials have been the subject of considerable research interest for many years. Theocaris and Panagiotopoulos [7.22] demonstrated the prospect of neural networks applications for fracture mechanics by adapting computational fracture analysis methods to the neural network computing environment. The following three problems were considered in more detail:

- a crack with classical interface condition,
- cracks with detachment and
- parameter identification problem for a cracked body.

To take into account these problems, the neural network approach and the 4th-order Runge-Kutta method are applied, in order to obtain a numerical solution of the system. Due to the appearance of inequality subsidiary conditions, problems are at different levels of difficulty, and the applied neural network shows good convergence. The obtained results and numerical experiments have led researchers to conclude that with respect to inequality problems, the ANN method is superior in comparison to the trial and error approach and traditional optimization. In addition, a relationship is developed between fracture toughness and crack geometry, specimen dimensions and operating temperature. An important property of ANNs is their ability to determine the contribution of parameters which impact the average fracture toughness.

Another way of using ANNs is to predict fracture parameters of certain heterogeneous materials such as concrete, composites or porous materials. Such parameters may include the critical stress intensity factor, and the critical crack tip opening displacement. The required data is available from specialist literature, i.e., from different test methods that had been performed in various laboratories. This data is then presented to the ANN, after which the trained system (ANN) shows results, which are very close to experimental results. The obtained results have an accuracy that is acceptable for most design applications. A study on the effects of compressive strength and maximum aggregate size on fracture parameters proved that ANNs can be used to perform parametric studies in the area of fracture mechanics.

In comparison to other methods, an ANN can be built directly from experimental data without simplifying assumptions or using self-organizing capabilities. Another advantage of such a system is that the results of ANNs can be easily adapted to other fracture models which also propose two parameters for fracture modelling.

7.4 Genetic algorithms

Evolutionary computing is a research branch of artificial intelligence that mimics models of nature, whereby through mutation, crossing and selection, only the fittest survive. Basic evolutionary-based techniques include:

- genetic algorithms (optimization methods),
- genetic programming (for program development), and
- evolution strategies.

The basic element (object) of *genetic algorithms* is the *gene*, which offers a possible solution of the problem for a single parameter. Genes are joined into a string to form a *chromosome*, corresponding to one solution of the problem. Genes can be:

- binary vectors,
- whole numbers, or
- real numbers.

Each new repetition of the evolutionary cycle produces a better solution of the problem. In genetic programming, an entire population of computer

programs is subjected to evolution. The goal is to choose the program that best solves the given problem. The evolutionary cycle consists of the following steps:

- generation of initial population;
- assessment of the last population (selection);
- the formation of a new generation based on genetic operations (crossing, mutation).

It is expected that each new repetition of the last two steps in the evolutionary cycle will provide a better solution to the problem. The problem that is being solved by genetic algorithms determines the form of the population and the assessment function that selects the best specimen. This method is used to solve optimization problems.
Studies of evolution simulation are related to a version of *genetic algorithms* as developed by John Holland [7.23] in his research on the nature of adaptive systems.[1] Holland simulated three processes critical to biological development:

- the *inheritance mechanism*, which can produce offspring similar to the parents;
- a procedure introducing *diversity to the process of reproduction*;
- the process of *differentiation*.

In the standard image of biological evolution, the inheritance mechanism is related to *chromosomes* (located in the genes). Diversity is obtained through gene splicing or *mutation* and differentiation reproduction through natural selection. In genetic algorithms, strings of symbols act as chromosomes, operations (as the recombination and mutation of these symbols) represent diversity (variation), while strings are reproduced and the readiness function performs selective reproduction. In this way, genetic algorithms produce new generations over and over again. In every successive (new) generation, the most successful string is selected to become the parent of the next generation. Once the "health" of the offspring of this generation has been checked, the best (most successful) strings become the parents of the following generation. Over the recent years, advanced research has focused mainly on the understanding of biological systems and processes [7.4] [7.5]. Different measured responses could be used to determine structural damages. For instance, strain mode shapes, natural frequencies and mode

[1] See also Grossdberg [7.24].

shapes are measured and used in the detection and identification of different types of structural damages. Like other AI methods, genetic algorithms are also applied in the area of fracture mechanics. In this chapter, engineering applications of GA in damage and failure detection, which is a sub-domain of fracture mechanics, are reviewed.

7.4.1 Genetic algorithms in damage and failure detection and diagnosis

Much like the ANN approach, genetic algorithms, too, are used for mechanical damage identification. Damage detection, and particularly the determination of the location and intensity of damage in mechanical systems or structures, is performed by utilizing the measured vibrational data on GA. Identification and location of damage in linear elastic structures can be determined from measured natural frequencies and mode shapes data based on GA. Applications of GA in damage detection include the implementation of a method with a GA-based residual force vector which is tested on a truss-type structure of gear and gearing to detect microscopic structural damage. Such structures are simulated with the finite element model, and GA-operated with exact mode shapes and natural frequencies. In order to simulate an experimental analysis, three different problems are assumed, and random noise is added to the value of natural frequencies and mode shapes. The gears transmition is also simulated numerically by FEM. The damage evaluation problem can be formulated as an optimization problem, or as an objective function. Different numbers of modes are used to compute the value of the objective function. Results show that the proposed technique is more suitable in comparison to conventional methods that use residual force vector for damage identification. Furthermore, accurate identification is recorded when multiple damages are present in the structural model. Since the modal parameters of the structures are functions of the physical parameters, the presence of any type of damages could be proved by changes in the modal properties of the engineering structures. Different simulated damage scenarios are defined by means of different locations and different degrees of damage. The FEM shows a decrease in stiffness of each finite element presented in the structure. For the validation of the proposed technique, experimental data is used.

Taking into consideration the various applications of GA in damage detection of engineering structures, it can be concluded that this method is an effective, easy-to-implement method. Due to its formulation, this technique can be used in different engineering structures. Moreover,

experience shows that no special requirement regarding the initial values of unknown parameters is needed. It should also be noted that the quality of measurement plays an important role in the success of this method.

Over the past years, researchers have suggested a combination of a GA with a compatible method to produce a hybrid computational method in order to accelerate convergence and obtain more accurate results. The application of GA in hybrid intelligent techniques for fracture mechanics and fault and failure analysis is discussed later on.

7.4.2 EXAMPLE: Optimization gears assemblies with GA

The primary goal when designing gear assemblies is to ensure their functionality for a certain service-life. To reduce the production cost with a rational use of the material, it is often required that the overall dimensions of the assembly as well as the dimensions of its parts are optimised with respect to the minimum weight. One needs to take into account the geometric, kinematic, mechanical, tribological, technological and economic constraints. A generalised optimisation procedure consists of:

- the definition of functionality, constraints and selection of the optimisation criteria,
- the identification of independent variables and definition of the gear assembly model,
- an iterative search process to provide a theoretically ideal gear assembly design based on a comprehensive model analysis.

Assembly configuration, operating modes, transmitted powers, prescribed angular velocities, speed ratios, centre distances and other prescribed features of the gear assembly have to be specified in the first step. The purpose of optimisation is to find a proper combination of gear modules, number of teeth, teeth widths and tooth helix angles, where the objective function reaches the minimum value.

7.4.3 Genetic optimization algorithms

Several gear pair optimization procedures, as well as optimization procedures for selected types of multi-stage gear assemblies have been proposed in the past [7.4] [7.7]. They are based on deterministic (usually gradient) optimization methods. Several optimization criteria can be used,

most commonly the criterion of minimum sums of volumes of pitch cylinders of gears:

$$F = \sum_{i=1}^{n} (m_{ni} z_i \cos(\beta_{oi}))^2 \, b_i + \sum_{i=1}^{m} P_i \tag{7.1}$$

Equation 7.1 is aimed at finding such a combination of parameters (module m_n, number of teeth z, tooth width b, angle of helical tooth β_o, materials and heat treatments), which decreases the value of the selected function F as much as possible. Symbol P_i, represents the constraint functions, which take into account a number of additional parameters (profile adjustment coefficients, material properties, etc.) Many parameters, which considerably affect the design of the gear assembly (the number of teeth, modules, centre distances), are only allowed to take the integer values, standardized or some explicitly prescribed values. In this way, the objective function becomes non-uniform and has a typical staircase shape.

However, gradient methods are not best suited for solving such problems because the derivatives of the objective function cannot always be determined analytically. Therefore, the optimization procedure needs to be carried out in several stages that always include random search patterns. In any case, it is necessary to evaluate the objective function in a significant number of parameter points. This shows that the classical methods of optimization cannot be implemented in this specific field using their full ability. This implies that the use of some alternative method is necessary.

Conventional calculations are limited only to one pair of gears in engagement and are not possible to take into account the side effects, such as the service life of bearings, shafts, and other parts of gear assemblies. Therefore, the presented optimization model includes an optimizer based on *a genetic algorithm*. A genetic optimization algorithm is an optimization procedure, which is based on the imitation of the principles of natural evolution. Important characteristics of the genetic algorithms are:

- the procedure is governed by the probabilistic transition rules and not the deterministic principles;
- characteristic binary coding of the parameter set is used, and parameters are not directly involved in computations;
- the search process does not begin from a single point (preliminary design), but from a population of points;

- in every point only the value of the objective function is required, without the need to determine gradients or any other information's.

In principle, genetic optimization algorithms require many objective function evaluations. Therefore, their efficiency cannot be directly compared with deterministic methods, where the objective functions are smooth, unimodal and well defined. The main advantage of genetic optimization algorithms is their robustness. This means, that non-uniform and multi-modal objective functions affect the ability of finding the global extreme on a much smaller scale, if compared to the gradient methods. A genetic algorithm's characteristic feature is *parameter coding*. The simplest and most commonly used is binary coding. A set of independent parameters of the objective function is coded in a form of binary strings which are recorded sequentially one after the other. The resulting binary strings are called the chromosome, which fully describes the related design. In this way, the discrete nature of variable description becomes very convenient for the optimization of gear assemblies.

Genetic optimization algorithms simulate the population development of living organisms. At the start, a group of points-members (the primary population) is randomly selected, which is analogous to biological systems. Individual members are represented with a single point. The fitness value is then assigned to every member, which determines the ability of the member to survive. In optimization terminology, it can be said that every point is being assigned an objective function value. The population size is usually prescribed at the beginning and does not change during the optimization process. The genetic algorithm tends to improve the current population. The criteria of improvement is usually the average of member fitness values, while the standard statistical deviation from the average value is also considered as important information. During optimization, the average population fitness should improve and at the same time, sufficient deviation needs to be maintained. The evolution process comprises of the members' selection according to some selection mechanism and then their transformation with the genetic operators into the next population generation.

The selection of mechanism is very important for the proper functioning of genetic optimization algorithms. Their fundamental theorem implies that the probability of a member selection is larger, if its fitness value is also large. The most popular and well performing is the method of stochastic remainder sampling without replacement. It has been shown that the probability of a member selection should not be directly proportional to its

fitness. That would allow for an exaggerated influence of just a few fittest members and would result in a premature convergence to local peaks. Such domination may be prevented by using the fitness scaling function to artificially reduce the fitness differences between members. Several scaling procedures may be used; however, practical experience shows that the ranking method is the best [7.4].

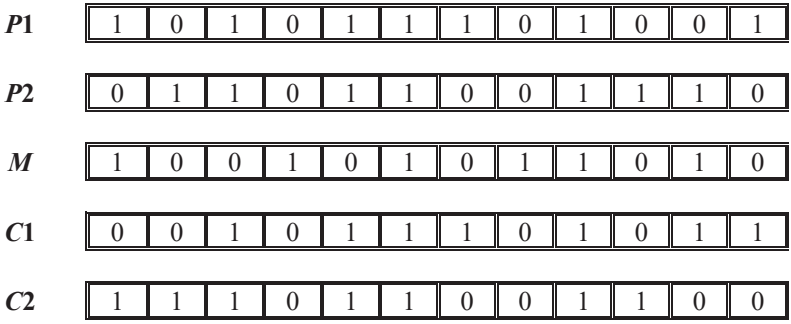

Figure 7.3: Example of uniform crossover.

Using this method, the members are sorted according to their fitness. Corrected member fitness is now represented with its rank. The convergence speed may also be controlled by using the additional linear rank correction function with variable gradient. The fundamental genetic operators are the reproduction, crossover and mutation. Reproduction is essentially the implementation of Darwin's well-known principle of natural selection, which predicts survival of only the fittest members. It is easily implemented by copying the selected members to the next generation without any modifications. This ensures continuous growth of the population average fitness; however, no improvements are made to the fittest individuals. If only this operator is used, it results after a few generations in an unbalanced population consisting only of the copies of the best-fitted individual from the primary population. The crossover operator produces new members by mating two selected individuals. Every newly born member inherits some properties from both parents. This is the motive process of the genetic algorithm because it generates new, better solutions that are based on best existing solutions.

7.4.4 The optimization {XE "optimization"} procedure

The uniform crossover [7.4] operator provides the best practical results.

Using this technique, two parents $P1$ and $P2$ are selected from the current population. Then the crossover bit mask M is randomly generated in a way, where the probability of each bit being set to I is equal to 50%. The crossover operator may be generally written in the following form:

$$C_I = P1 \overset{e}{\underset{\mathrm{V}}{}} \left(\left(P1 \overset{e}{\underset{\mathrm{V}}{}} P2 \right) \wedge M \right)$$
$$C2 = P1 \overset{e}{\underset{\mathrm{V}}{}} \left(\left(P1 \overset{e}{\underset{\mathrm{V}}{}} P2 \right) \wedge \bar{M} \right) \tag{7.2}$$

The crossover operator preserves the features (bits) that are common to both parents, while the remaining features are mixed randomly. Figure 7.4 illustrates the best optimization algorithm, which has been determined through numerical experimentation with various algorithms.

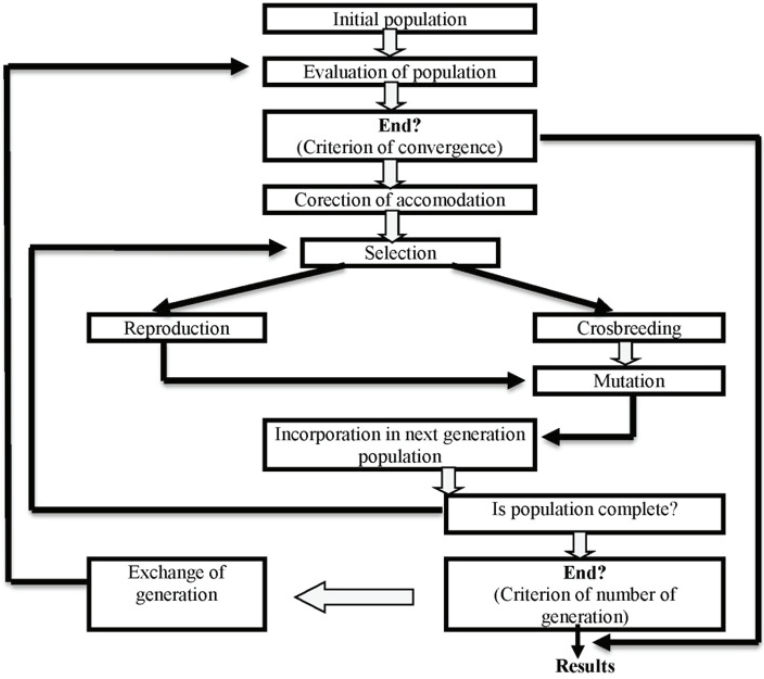

Figure 7.4: Flowchart of the optimization procedure

In the first step, all parameters are considered to be real numbers. This means that each parameter is coded with 32 bits. The algorithm of geometric dimension minimization tends to the use of the best quality materials, thermal treatments and manufacturing procedures. Therefore, in the first phase the best materials and procedures are prescribed, whereby a list of allowable values is taken into account.

Constraints are considered with limiting functions, which need to be carefully chosen. If the rate of the limiting function is too high, the fitness value of any member inhabiting the forbidden design area is too low to make its survival possible. The next generation therefore possesses no information about the forbidden area and the evolution is dependent only on the slope in the allowed design area. When the objective function is monotonous enough, the best results are obtained when the objective function is mirrored over the limiting point, which, in the case of limiting the lower value of the centre distance, is expressed as follows:

$$P_i(a) = F\big(a_{min} - (a_{min} - a)\big) - F(a) = F(2a_{min} - a) - F(a) \quad (7.3)$$

A relatively small population (20 to 50 members) is used when applying the genetic algorithm, and the procedure is repeated many times (3 to 10 times). The best solutions of individual iterations are stored in the database. The primary population is generated randomly before the first iteration. In the following iterations, the starting population is generated from the database and the rest are generated randomly. After each iteration, a convergence analysis is performed and the control parameters are adjusted, if necessary. If a too large diversity of population is observed in the last stage, the probability of mutation and crossover is lowered. With that, the direct reproduction rate is increased. Premature convergence and stagnation of the procedure in the last generations is prevented with a change of the scaling function slope.

The effectiveness of a genetic algorithm applied to typical gear assemblies is, at this stage, worse in comparison to traditional methods of optimization. However, a genetic algorithm performs much better in the case of a larger number of independent parameters and unusual design configurations. Further improvements can be achieved through the use of hybrid genetic operators, which need to be properly investigated. In the next phase, discrete parameter values are used. It is not necessary to narrow the optimization domain, since the genetic algorithm obtains information about the global extreme positions from the primary population members. The primary

population consists of the best members from the first phase of the process. Due to a change in the coding, it is necessary to convert the database accordingly. This phase of the process fully demonstrates all the advantages of genetic algorithms. The solution domain around the global extreme is quite wide, which necessitates the use of different optimization criteria in the second phase. In the first phase, it is advisable to use the criterion of the minimum sum of pitch cylinder volumes, which results in optimal assembly mass. In the second phase, criteria such as the rational use of material and manufacturing cost consideration are more appropriate. Genetic algorithms also allow for a multi-criteria optimization.

For solving equation 7.1, however, the mentioned genetic algorithm is suitable primarily because it does not need any other information about a selected function, except its value in selected points. Genetic algorithms take into account the total effect of reproduction, crossbreeding and mutation. By combining the three influencing variables, the basic theorem of genetic algorithm is obtained:

$$m(H,t+1) \geq m(H,t)\frac{f(H)}{\bar{f}}\left[1 - p_c\frac{\delta(H)}{l-1} - o(H)p_m\right] \tag{7.4}$$

At this point, the optimization process can be summed up as follows:

1. a small initial population is selected at random;
2. by using the genetic algorithm the convergence of the population with respect to a selected local criterion is affected;
3. a new population is selected including the most successful members of the old population as well as new members selected at random;
4. the procedure is stopped if the convergence criterion[2] is won, or we return to point 2.

7.5 Fuzzy logic

The fuzzy set theory (FST) was developed and applied by Lotfi Zadeh [7.25] for making better decisions when dealing with the uncertainty of daily life. The FST is more general than traditional logical systems. A fuzzy

[2] *Criterion of convergence*: the convergence of particular solutions to the true solution is based on the behavior of errors which measure how close the iterations are to the true solution. The difference between two consecutive results is calculated, and when this difference is close to 0, the convergence criterion is reached.

logic system (FLS) provides a valuable flexibility for reasoning by setting its rules in natural language. Over the past few decades, it has found many engineering and scientific applications and its achieved results have increased the amount of interest of this AI method. The FST is often used in different areas of mechanical engineering such as design, modelling and optimization, as well as in different fields of fracture mechanics. The FLS process has three main phases, fuzzification, inference and defuzzification. In fuzzy systems, the human knowledge is represented in terms of simple if–then rules with a condition and a conclusion [7.26] [7.27].

A FL algorithm is initialized qualitatively by defining the linguistic variables and terms, and quantitatively by contrasting the membership functions (μF(x)). Therefore, by utilizing these membership functions, the input data is converted to fuzzy values. Afterwards, in the inference phase, all the rules are evaluated, and the results of each rule are combined. Finally, the output data is converted to non-fuzzy values by utilizing the membership functions.

7.5.1 Fuzzy logic in failure mode and mechanism identification

In recent years, different kinds of knowledge and experience have been utilized and applied in order to characterise and predict mechanical failure. Historical data has been used to develop a FL-based model that is able to predict mechanical and operational failures. More specifically, the researchers using the MATLAB fuzzy logic tool box considered not only different parameters, but also critical factors affecting failure. The prediction effectiveness of the model is validated mathematically. Moreover, the obtained results were compared to models which were developed in previous studies through different methods. The FL approach has predicted failure equally correctly as other techniques.

In the implementation of FL systems, the generated fuzzy rules are based on engineering knowledge and use logic rules, provided by experts or by standard specifications. This method can be used also very effectively in the maintenance of various engineering structures.

7.5.2 Fuzzy logic in fault and error detection and diagnosis

The fuzzy set approach is applied successfully in mechanical fault classification [7.28] and Mechefske experimentally studied the feasibility of FST in integrated machine fault diagnosis. In order to provide

information related to the dynamic behaviour of the machine, vibration sensors are installed on the housing of the gearbox and bearings support. The rotating parts produce a unique frequency and a change of this frequency is defined as a fault. Complete information about the dynamic condition of the machine is defined, after which the Fast Fourier Transform Spectrum (FFTS) is utilized to extract useful information from sample data and convert it into the frequency domain. Fault clustering and fault assignment as two stages for fault diagnosis are considered. The output of fault clustering is set, and the fault assignment appoints input data to an existing pattern cluster. Fuzzy relations are defined with respect to the FST, equivalence class under a certain threshold is founded. The algorithm can be implemented in different symbol programming languages. Mechanical vibration signals are used extensively for fault diagnosis by means of different methods.

7.5.3 Fuzzy logic in damage and failure detection and diagnosis

Since the failure of engineering structures can lead to irreversible damage and fatality, structural damage detection by AI methods is a critical issue for researchers. Through their research work, Chandrashekhar and Ganguli [7.29] demonstrated the ability and accuracy of FL in mechanical damage detection. They modelled a Euler-Bernoulli beam, whereby natural frequencies were determined by FEM. The reduction on the element stiffness represents the damage on the structure. The measurement displacements are the input of the FL system, and location and size of the damage are the outputs. In this way, fuzzy sets with Gaussian membership functions are used for input variables. In addition, fuzzification of the numerical values, which are achieved through the finite element analysis, is used to obtain rules for the fuzzy system. The FL system could be tested by measured data and high percentage of success is achieved. Sometimes in real-life situations, parts of measurements are missing or faulty. The researchers considered this issue and tested the implemented FL system with a modal vector missing. The proposed damage detection algorithm was found to be quite accurate in damage classification with missing measurements as well.

Damage in engineering structures is an important theme, a topic of interest in research. As part of fuzzy mechanics in crack detection [7.16], a FL control method could be applied. More specifically, a fuzzy controller with a hybrid membership function could be formulated. As the existence of cracks is able to change the modal data, this data can be used to determine

crack parameters. Therefore, the first three natural frequencies and the first three mode shapes are employed and defined as input parameters of the system. Then, the desired parameters of the crack are defined as an output of the system. On the other hand, the MATLAB toolbox could be used, and several fuzzy rules are learned in order to achieve the results. A comparison between the obtained results from the MATLAB and the FEM proved that the implemented fuzzy system is able to estimate the location and depth of the crack accurately.

The review of the published papers indicated that in the researches which are used FL approach for damage diagnosis based on experimental data, achieving higher precision is possible by more measurement tests and results. It should be noted that, environmental noise or incorrect locations of the sensors have high impact on modal parameters like mode shapes and accuracy of the results. Furthermore, it is concluded that the FL is beneficial in the decision safety systems while response time is important.

7.5.4 Evolutionary optimization of service life

Performing an optimiziation process refers to the sequence of targeted actions based on a study of the elements of the optimization process in a particular order. This sequence is a string of optimization elements that make up units. Thus, a series of fuzzy oriented graphs is created.

A variety of options are available: interconnecting different elements, combining related structure, applying different time optimization methods, or using different resources that are required in order to perform the optimization process. The formation of the optimization process, therefore, can be seen as a search for the most optimal implementation of this process. Here is a demonstration of the basic principles through the example of time optimization for a specific mechanical part. In order to do this, the matrix of the relation must be taken into account:

$$Y = F(Y, Z) \tag{7.5}$$

where

$X = (x_1, x_2, K, x_n)$ is the vector of the input variable, marking the external environment characteristics that affect the functioning of the model;

$Z = (z_1, z_2, K, z_r)$ is the vector of the internal variable, marking the condition of the individual elements of the model;

$Y = (y_1, y_2, K, y_m)$ is the vector of output variables, which marks the characteristics of the model in general, and determines the stage at which the intended purpose is achieved.

One of the most critical elements in the sequence of optimization is definitely service life. To determine the time, internal variables and their connections have to be taken into account and be logically evaluated. To simplify the above formulation, let us consider the optimization element as a constant. Therefore, for the calculation of time, the elements of the model according to Equation 7.5 are as follows:

Y is the time needed to study a single element;
X is the vector determining the structure of the element;
Z is the vector determining the relations between the internal variables of the individual element of a structure and the structure as a whole (i.e., it describes the appropriate matrices).

In this case, the dependency relation is determined according to the following equation:

$$Y = \sum_{i=1}^{m} \sum_{j=1}^{m_i} m_{ij} \tag{7.6}$$

where m is the number of elements, and m_i is the number of elements for calculating in the j-th optimization.

However, the above expression fails to consider the conditions that affect the calculation of the time required, i.e., whether the two related elements occur simultaneously, consecutively, or parallel. Thus, Equation 7.6 is reformulated as:

$$Y_{12} = m_1 + m_2 - \Delta_{12} \tag{7.7}$$

where Δ_{12} is the connection between the optimization elements.

During the study of content-related element E1 and E2, however, several other, e.g. p optimization units can also be performed. A generalized equation can be derived for determining the study of two optimization elements, which are related in terms of their form:

$$Y_{12} = m_1 + m_2 - \Delta_{12} \tag{7.8}$$

$$\text{where } \Delta_{12} = \begin{Bmatrix} 0, if \{l_1\} and \{l_2\} = 0 \\ \sum_{i=1}^{n} \Delta_i, \ if \ if \{l_1\} and \{l_2\} \neq 0 \\ -\sum_{i=1}^{p} n_i \ \zeta_i, if \ |Y(m_2) - Y(m_1)| > 0 \end{Bmatrix}$$

and l_1, l_2 is the number of elements for optimization in the calculating element E1– m_1, or E2 – m_2.

Therefore, the full vector is $Z = \{Z_1, Z_2\}$ where Z_1 is the vector determining the possible sequences of the optimization elements during calculation, and Z_2 defines the matrix of content relations. It needs to be taken into account that the vector is

$$Z_1 = \{z_{1i} \rho(i+1) > \rho(i)\} \tag{7.9}$$

where $\rho(i)$ is the local level *I*.

In order to use the matrix further according to Equation 7.5, the following changes need to be made:

1. On the basis of the matrix for content-related connections $M = \| \mu_{Kij} \|$ we move to the binary matrix $M^* = \| m_{ij} \|$ according to the conversion formula:

$$m_{ij} = \begin{cases} 1 \ if \ \mu_{Kij} > 0, \\ 0 \ everywhere \ else \end{cases} \tag{7.10}$$

2. Matrix *M** is converted into matrix $T = \| t_{ij} \|$ that represents the mutual temporal relations between the learning units NB:

$$t_{ij} = \Delta_{ij} \tag{7.11}$$

where Δ_{ij} is determined according to equation 7.8.

According to Equations 7.8 and 7.11, the time can be calculated according to the following expression:

$$Y_{i,i+1} = m_i + m_{i+1} - \Delta_{i,i+1} \tag{7.12}$$

To study the connections between the optimizing element E, the following function is derived:

$$Y = \sum_{i=1}^{m-1} Y_{i,i+1} \rightarrow min \tag{7.13}$$

The problem regarding the optimization of time needs to be approached in accordance with the limits defined by the sequence vectors of the optimization elements, with different E's - $Z_1(LE1), Z_2(LE2)$ and with the matrix for content-related connections *M*.
Such a fitness function[3] (survival function), which is the result of these relationships, is described as follows:

$$Y = F\ (LE1, LE2, m, Z_1(LE1), Z_1(LE2), M) \tag{7.14}$$

The optimal solution to the problem of time calculation according to Equation 7.13, coupled with the fitness function according to Equation 7.11, generates a set of tailored solutions, which consists of the vectors X_1, X_2, K, X_p, where p is the number of elements with a non-zero value in the matrix, which constitute integration vectors. A genetic algorithm is applied to solve this problem. The sought values are defined according to the theory of genetic algorithms[4]. A chromosome is a vector in the causal sequence. Each chromosome has a conditional binary structure, which is schematically shown in Figure 7.5.

[3] The *fitness function* is a function which takes a candidate solution to the problem as input and produces as output how "fit" our how "good" the solution is with respect to the problem in consideration.
[4] A *genetic algorithm (GA)* is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection.

1	0	1	0	1	1	0	...	0	0	0	1	0	1	0	1	...	1
0	1	1	0	0	1	0	...	0	0	1	0	0	1	0	0	...	0
1	0	1	0	0	0	1	...	0	0	1	1	0	1	0	1	...	0
1	0	0	1	0	1	0	...	0	0	1	1	0	0	1	0	...	0
0	0	1	0	1	1	0	...	0	0	1	0	0	1	0	0	...	0
1	0	1	1	1	0	0	...	0	0	1	1	0	1	1	1	...	0

Figure 7.5: The structure of a string of chromosomes

Each chromosome $X_i | \, i = \overline{1,p}$ consists of n genes (bits), and, in addition, it is valid that $l_1 + l_2 = n$. Each gene is either present (1) or absent (0) in a certain learning situation. Moreover, the position (locus) $X_i(j) | \, j = \overline{1,n}$ determines the number of optimization units in the course of the optimization task. It can be concluded that a genetic algorithm (GA) is the most appropriate way to optimize the formalization of the problem in question [7.27].

The goal of time optimization is to find the next sequence of chromosomes, which ensure minimal training times during the entire execution of the optimization process, taking into account the constraints arising from the internal subject structure of the individual subject.

A block diagram of changes occurring in a GA is a result of modifying the parameters is shown in Figure 7.6.

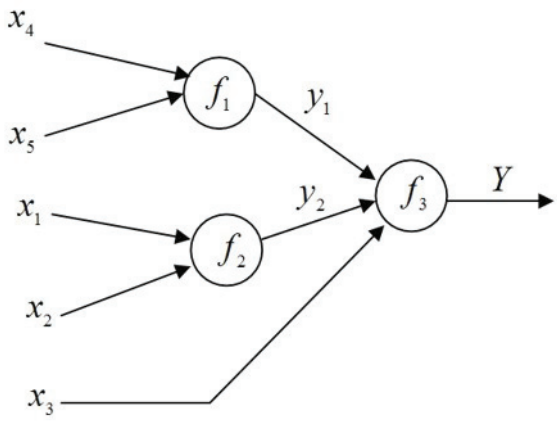

Figure 7.6: Block diagram of GA

For an overall analysis of the effectiveness of the optimization process, it is also important to define the degree of distance from the set optimization goals. The problem is considered in more detail in the following section.

7.5.5 Hierarchic fuzzy reasoning for effective optimization

It is a fact that in order to achieve a high level of optimization effectiveness in the process of calculation, it is also extremely important to provide the designer with feedback information regarding their achievements. Below, a short example is used to demonstrate how such a requirement can be met. Since there are no reliable experimental data available for this process of determining the effectiveness of optimization, the model will employ a system of fuzzy reasoning. Considering the main elements to define a model for evaluating the effectiveness of optimization, and taking into account the "input-output" principle, this model can be expressed as follows:

$$Y = F(x_1, x_2, x_3, x_4, x_5) \tag{7.15}$$

where

$Y = C$ – is the indicator of the degree of distance from goals;
$x_1 = U$ – is the level of absorption of the optimizing unit NE;
$x_2 = A$ – is the indicator of the level of abstraction of the NE;
$x_3 = y$ – is the indicator of the level of awareness about the assimilation of NE;
$x_4 = K_3$ – is the coefficient of NE absorption; and

$x_5 = K_H$ – is the coefficient of knowledge absorption for the individual optimizing unit NE.

If a variable is defined to evaluate the application of *four* conditions, then the maximum number of rules for creating relations according to Equation 7.15, in a single subject, will be $4^5 = 1024$. For a comprehensive acquisition of knowledge, it is recommended that a concept of hierarchical knowledge be applied. Moreover, using a model of hierarchical fuzzy knowledge, the problem of high-dimensional matrices can be avoided [7.27] [7.10].

7.6 Hybrid intelligent techniques

Hybrid intelligent systems are used in a wide variety of applications [7.4]. In an intelligent hybrid system, two or more AI methods are combined to overcome the limitations of an individual method. In other words, hybrid intelligent systems are computational systems with integrated different intelligent approaches. The benefits of hybrid intelligent systems for different applications are increased. For example, rule-based reasoning does not perform very good in optimizing and in working with poor or noisy data, but it is very good for explanations of the reasoning process. On the other hand, artificial neural networks are not very good for explanations of the reasoning process, but they act very well in learning and in working with noisy and poor data. Therefore, a hybrid intelligent system combining rule-based reasoning and ANNs has advantages in comparison to the individual approaches. However, many attempts are performed in order to predict and consider all the possibilities at the design stage. This section provides a review of hybrid intelligent systems employed for mechanical fault detection and diagnosis.

In the early current century, the application of neuro-fuzzy networks in vibration monitoring was proposed by Meesad and Yen (2000). The system consists of an input layer, a hidden layer and an output layer based on a standard FL system. If-then rules are reduced by using a neural network. For the evaluation and illustration of the performance of the proposed hybrid system, Fisher's Iris data set and the Westland vibration data set are used. For the Westland vibration data set, several accelerometers are mounted on different locations of the power transmission of the helicopter. The data is recorded for different types of fault conditions and one correct (no fault) condition. On the Westland vibration data set, two classifiers are used and compared with the implemented neuro-fuzzy classifier. The

obtained data showed that the proposed neuro-fuzzy network has a higher accuracy and performance in comparison to other classifiers.

The neuro-fuzzy technique is also used for crack identification in engineering structures. In the early current century, the neuro-fuzzy technique began to be employed for crack identification by Shim and Suh (2002). Since crack properties such as location and depth of the crack have a significant influence on the dynamic behaviour of structures, identification of these parameters has been widely investigated. In order to predict the size and location of cracks in structures, a Continuous Evolutionary Algorithm (CAE) can be employed based on the Adaptive-Network-Based Fuzzy Inference System (ANFIS) architecture. ANFIS uses the depth and location of the crack as an input, and structural eigenfrequencies as an output. Crack parameters can be determined from measured eigenfrequencies of the structure. The neuro-fuzzy-evolutionary technique estimates the size and location of cracks with high accuracy. The method could be generalized for general boundary conditions and structures.

7.6.1 Example: A neuro-fuzzy model for cracks

Technological development requires a change of the working methods, means and forms in order to meet the demands of tomorrow's society. Just as it is no longer possible to imagine modern-day production without the use of flexible manufacturing systems and robots, perhaps tomorrow it will be impossible to envisage production that would not employ some other, new artificial intelligent technological tool or method of work. The foundation of such an innovative approach to society includes a variety of adaptive (intelligent) programs, which manage society and direct it to increase productivity, something enabled especially by contemporary computer and information systems [7.30]. Such an intelligent system should be able to, among other, perform the processes of information channelling, problem solving, and adapting basic parameters such as the depth and location of cracks [7.31]. The focus should not be limited only to the pace at which this process unfolds, but also to the content and the methods used to deliver this content.

Based on these considerations, the possibilities of using intelligent technologies in the automation of the optimizing process (OP) S_p, will be explored, and their application demonstrated on simple, yet still informative enough examples. As a theoretical starting point *Industry 4.0* will be used, introducing structural and factor analysis to the system. The structural

elements of the optimization model must define the objectives of the functioning of the OP, as well as a series of influence factors that comprise such an OP and, above all, the mutual interactions between these factors. An OP consists of techniques, forms, and learning tools. A formalized description of the process of functionalizing an OP, together with a macro-model of the optimizing process, is shown in Figure 7.7 [7.27].

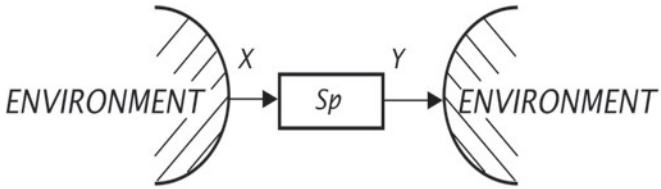

Figure 7.7: A macro-model of the optimization process

The optimization system OP *(Sp)* is a converter of state *X* to state *Y,* whereby *X* is the data input and *Y* is the result characterizing the achieved level of optimization. By the term optimization environment, the *object of management* can be labelled*,* which is the source for the target object. In the widest sense of the word, this "object" can stand for the entire structure, which is placed on a particular structural part, and thereby setting appropriate standards and requirements. Here is an example in which the object of investigation is the society, or the economy, for example, Industry 4.0, which defines its target requirements *Z,* and thereby defines the necessary output parameters of the OP.

$$Y \in Z \tag{7.16}$$

The optimization system in this case is an environment, in which the elements in manage and perform the optimization process with the use of technology. In order to achieve this, the elements within the structure, consisting of two interwoven groups, need to be defined:

- a group of structural elements, and
- a group of elements representing the optimization technology that ensures the resolving of a structural problem.

The structure of the optimization system in relation to OP is shown in Figure 7.7.

Below is a simple example, a single-element OP. On the basis of a suitable structural and functional analysis, objectives for this case must be set, and the basic elements comprising this OP must be defined. The latter can include a series of mechanical $\{m_1, m_2, K, m_k\}$ structural $\{f_1, f_2, K, f_l\}$ functional methods $\{s_1, s_2, K, s_m\}$, for system configurations $\{ds1, ds2, K, dsn\} \in OP$, and elements of structure $\{le_1, le_2, K, le_n\}$. Such is the preliminary study for the current structural element le_{n+1}, as anticipated by an OP designed in this way. The connections between the elements comprise a set of coefficients, which generate the appropriate choice of the intermediate layer of neuron network (NN), and the choice of a string of binary values to define the single-element OP:

A formal sample of the OP is obtained in this way, based on the description of these structural parameters, which are founded on the knowledge of experts and on relevant research. This OP model can now be formalized and presented in the form of a multi-layered NN. The input layer thus consists of groups of optimization system features:

$$\sum_{i=1}^{n} le_i = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i + \sum_{i=1}^{n} c_i + \sum_{i=1}^{n} K_{H_i}, \tag{7.17}$$

where

n is the number of elements anticipated by the OP;
a is the level of abstraction of the elements ($a \in \{1, 2, \ldots 4\}$);
b is the level of assimilation of the elements ($b \in \{1, 2, \ldots 4\}$);
c is the level of attention absorption and
K_H is the coefficient of automation, i.e. the acquired competences ($0 \leq K_H \leq 1$).

where OP is the name of the variable for adding one of the values, corresponding to a subject-based system. However, since the results of research serving as a basis for a certain OP cannot be defined uniquely (i.e., they cannot be formalized), this model proposes a system of fuzzy rules in order to present these results. For the presented example from the perspective of defining the goals in the OP and formalizing the optimization situation, a neuro-fuzzy model NN is chosen. The main advantage of this model in comparison to similar ones is that it combines the benefits of the NN and at the same time allows for the replacement of deterministic rules by applying fuzzy rules. Apart from this, a synthesis of these models enables the initial NN model to be performed on the basis of expert evaluations and assessments of the proper functioning of such a system.

The following criteria should be considered in order to illustrate this; level of assimilation, speed of operation, complexity, costs, system stability (result dispersion), and time of learning. In the case of unsatisfactory values of these performance indicators, the rules can be defined by the system itself. The neuro-fuzzy model of building an OP for concrete training is the basis for creating a single element system. Through a sequence of combinations, however, a multiple-element OP can also be created, while the rules for meeting specific structural and functional conditions can be generated as fuzzy knowledge. A model developed in this way, combined with performing neuro-fuzzy activities, enables a comprehensive automation even of the rather complex and lengthy procedure of selecting the most suitable OP.

In the building of information sub-systems that make up a single element system, various professional and semi-professional tools can be used. With a view to demonstrating the possibilities of such an approach in modelling an OP, but also on the account of accessibility and relative simplicity of use, the *MATLAB* software package is used. By using the ANFIS fuzzy NN[5], Figure 7.8 shows both the structure and the values of the input variables *a, b, c,* K_H, on the basis of which the OP was built. As part of a logical implementation, the method of minimal values is used in combination with the method of maximum value. As a method of reasoning (i.e., of activating reasoning for each fuzzy rule), the method of minimum value was selected. The function point method was used as a principle for selecting the appropriate optimization method.

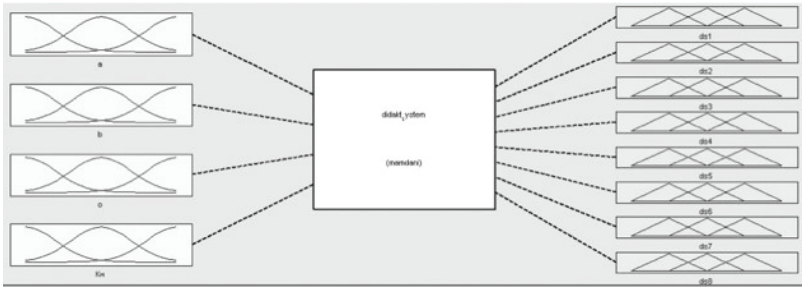

Figure 7.8: Structure of a neuro-fuzzy NM for an OP [7.27]

[5] ANFIS is an *adaptive neuro-fuzzy inference system* or *adaptive network-based fuzzy inference system,* a kind of artificial neural network that is based on the Takagi–Sugeno fuzzy inference system.

The obtained fuzzy rules are based on data acquired through preliminary structural research performed by experts (knowledge base). Figure 7.9 thus shows one of the possible results obtained by means of the presented model. The results are based on a general analysis, an estimate of the impact of changing the input values of fuzzy (soft) variables.

A model designed in this way, i.e., based on fuzzy logic, enables the automation of one of the most demanding stages of creating an optimization system. The evaluation of its effectiveness is based predominantly on fuzzy rules, by means of which the missing information between the input and the output data is provided.

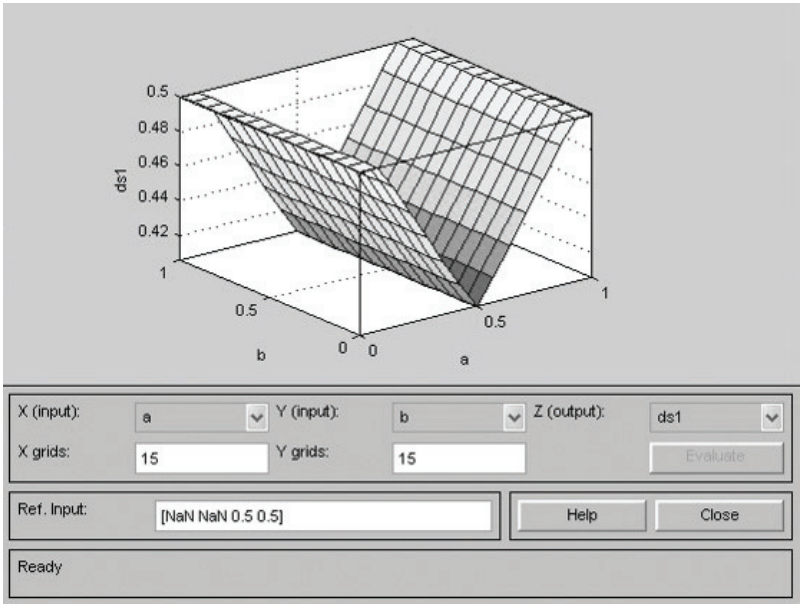

Figure 7.9: Fuzzy model surface at the end of model development [7.27].

7.7 Conclusion

In this chapter, the main concepts and mechanisms of AI approaches in fracture mechanics are briefly highlighted. Moreover, it is explained how they are applied on sub-domains of fracture mechanics: failure mode and failure mechanism identification, fault and error detection and diagnosis, damage and failure detection and diagnosis, mechanical fracture and

fracture parameters, and calculation of service life or remaining service life. The contribution could provide some ideas for future perspectives of research in this field.

References

[7.1] Griffith, A. A., "The Phenomena of Rupture and Flow in Solids," *Philosophical Transactions of the Royal Society of London.* A221, pp. 163-197. 1921: and "The Theory of Rupture," *Proceedings of the First International Conference of' Applied Mechanics,* Delft (1924).

[7.2] Muskhelisvili, N. I., *Some Basic Problems in the Mathematical Theory of Elaslicity,* Nordhoff, The Netherlands (1954).

[7.3] Gordon J.E., "The New Science of Strong Materials, or Why You Don't Fall through the Floor", *Penguine,* New York, (1976)

[7.4] Aberšek, B., Flašker, J. (2004). *How gears break*. Southampton: Witt Press.

[7.5] Aberšek, B., Popov, V. (2004). Intelligent tutoring system for training in design and manufacturing, *Advanced engineering software*, 35, pp. 461–471.

[7.6] Nasiri, S., Khosravani, M.R. and Weinberg, K. (2017). Fracture mechanics and mechanical fault detection by artificial intelligence methods: A review *Engineering Failure Analysis, 81. 270-293.*

[7.7] Aberšek, B., Borstner, B., & Bregant, J. (2014). *Virtual teacher: cognitive approach to e-learning material*. Newcastle upon Tyne: Cambridge Scholars Publishing.

[7.8] Honavar, V. (2006). Artificial Intelligence: An Overview. Ames: Iowa State University.

[7.9] Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction.* Oxford: Blackwell Publishers.

[7.10] Aberšek, B. (2018). Problem Based Learning and Proprioception. Newcastle upon Tyne: Cambridge Scholars Publishing.

[7.11] Kordigel Aberšek, M. & Aberšek, B. (2020). Society 5.0 and Literacy 4.0 for the 21st Century. New York: Nova Science Publishing.

[7.12] Kurzweil, R. (2005). *The Singularity is Near*. New York: Viking Press.

[7.13] Deutsch, D. (2012). The Beginning of Infinity: Explanations That Transform the World. Harlow: Penguin Books.

[7.14] Pearl, J. (1983). Heuristics: Intelligent Search Strategies for Computer Problem Solving. New York: Addison-Wesley.

[7.15] Aberšek, B. and Flašker, J., Numerical methods for evaluation of service life of gear, *International Journal for Numerical Methods in*

Engineering, Vol. 38, 2531-2545, 1995.
[7.16] Glodež S, Jezernik N, Kramberger J, Lassen T. Numerical modelling of fatigue crack initiation of martensitic steel, Advances in Engineering Software, 41 (2010), 823–829.
[7.17] Kurz, M. Handbook of Materials Selection, Wiley, New York, 2002.
[7.18] Prawoto, Y. Solid Mechanics for Materials Engineers, Lulu Enterprises Inc., USA, 2013.
[7.19] Astakhov, V.P. Metal Cutting Mechanics, CRC Press, USA, 1999.
[7.20] Sahin, S., Shenoi, R.A. Quantification and localisation of damage in beam-like structures by using artificial neural networks with experimental validation, Eng. Struct. 25 (2003) 1785–1802.
[7.21] Li, H., He, C., Ji, J., Wang, H., Hao, C. Crack damage detection in beam-like structures using RBF neural networks with experimental validation, Int. J. Innov. Comput. Appl. Inf. Control 1 (2005) 625–634.
[7.22] Theocaris, P.S., Panagiotopoulos, P.D. Neural networks for computing in fracture mechanics. Methods and prospects of applications, Comput. Methods Appl. Mech. Eng. 106 (1993) 213–228.
[7.23] Holland. J.H. (1975/1992). Adaptation in Natural and Artificial Systems. *Cambridge: The MIT Press.*
[7.24] Grossberg, S. (Ed.) (1988). *Neural Network and Neural Intelligence*. Cambridge: The MIT Press.
[7.25] Zadeh, L.A. Fuzzy sets, Inf. Control, 8 (1965) 338–353.
[7.26] Mendel, J. Fuzzy logic systems for engineering: a tutorial, Proceedings of the IEEE, 1995, pp. 345–377.
[7.27] Mazurok, T. (2013). *Learning Management Systems.* Odessa: PNPU. K. D. Ushinsky,
[7.28] Mechefske, C.K. Automatic Machinery Fault Detection and Diagnosis Using Fuzzy Logic, University of West Ontario, Canada, 1998.
[7.29] Chandrashekhar, M., Ganguli, R. Structural damage detection using modal curvature and fuzzy logic, Struct. Health Monit. 8 (2009) 267–282.
[7.30] Meesad, P., Yen, G.G. Pattern classification by a neurofuzzy network: application to vibration monitoring, ISA Trans. 39 (2000) 293–308.
[7.31] Shim, M.B., Suh, M.W. Crack Identification Using neuro-fuzzy-evolutionary technique, KSME Int. J. 16 (2002) 454–467.

INDEX