Computational modelling of fatigue failure

Peerlings, R.H.J.

Published: 01/01/1996

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Dec. 2018
Computational Modelling of Fatigue Failure

Ron Peerlings
Report WFW 96.131

Eindhoven, October 1996
Contents

1 Introduction 2

2 Phenomena and mechanisms 3
  2.1 Cyclic behaviour ........................................ 3
  2.2 Crack initiation ........................................ 6
  2.3 Stable crack growth .................................... 7
  2.4 Final fracture .......................................... 9
  2.5 Complex loading ....................................... 9

3 Engineering practice 11
  3.1 Engineering approaches .................................. 11
  3.2 Total-life approach .................................... 12
  3.3 Defect-tolerant approach ............................... 15
  3.4 Material aspects ....................................... 16

4 Computational approaches 18
  4.1 Damage initiation ...................................... 18
  4.2 Numerical fracture mechanics ......................... 19
  4.3 Continuum damage mechanics ......................... 20

5 Conclusion 24

Bibliography 27
Chapter 1

Introduction

Mechanical failures of engineering structures and components often cannot be attributed to a single extreme load causing instantaneous fracture, but are due to repeated loading during a certain time interval. Material degeneration under these repetitive loads is called fatigue. It is difficult to formulate a comprehensive definition of fatigue failure, but the following features are characteristic of the process: a) loads and deformations are fluctuating, but loading levels remain well below those causing fracture under monotonic loading; b) the load variations cause an accumulation of microstructural damage, which may culminate in macroscopic cracks and complete fracture.

Estimates of the percentage of fractures that are due to fatigue range from 50% to 90%. It is encountered for instance in rotating machinery, ground vehicles and aircraft, but bridges and off-shore structures may also suffer from fatigue due to wind and wave loading. Fatigue fractures are often unexpected and their consequences in terms of financial damage or even loss of lives may be tremendous. For this reason, the subject has been studied extensively and is still a research topic. Current research is aimed at gaining insight in the underlying mechanisms, at developing materials with an improved resistance to fatigue and at developing tools and engineering approaches for design against fatigue.

The purpose of this report is to assess computational strategies for evaluating fatigue life. If accurate and efficient, numerical simulation techniques of fatigue failure may become valuable tools in engineering design, complementary to—and probably integrated with—other numerical simulation tools that are currently gaining interest in engineering practice. They can predict the fatigue resistance of several alternative designs without the need for prototyping and extensive full-scale testing, and can thus lead to considerable savings in terms of development cost and time. At an earlier stage, computational models can provide means of investigating fatigue phenomena in a systematic way, of developing engineering rules and of understanding and improving empirical relations.

The development of reliable numerical models of fatigue is impossible without at least a basic knowledge of the typical phenomena in fatigue and the (micro-)mechanical deformation mechanisms which are responsible for these phenomena. A resumé is therefore given in Chapter 2. Analogously, valuable insights can be gained from rules of thumb and empirical relations that have proved their value in decades of engineering practice. Some of these engineering aspects of fatigue are discussed in Chapter 3. The essentials of some fundamentally different approaches to numerical modelling of fatigue are presented in Chapter 4. Special attention is paid to the specific strengths and complications of these methods. Finally, some concluding remarks are made in Chapter 5.
Chapter 2

Phenomena and mechanisms

The nucleation and evolution of fatigue damage is governed primarily by mechanical loading, but it is affected also by a number of other factors. Most notable among these are temperature, chemical environment and the presence of material and surface defects. It is often accompanied by and it interacts with other physical processes, such as creep, corrosion, phase transformations, etc. These secondary effects are left out of consideration here. Also, the discussion is focussed on fatigue in metals and alloys, with only marginal attention being given to polymers, composites and ceramics.

The fatigue fracture process may be separated into three distinct stages, which can be identified afterwards on the fracture surface: crack initiation, stable crack growth and final fracture. In the initiation phase, a macrocrack nucleates at one, or sometimes at several locations. Subsequently, the crack propagates in a stable manner during a large number of load reversals. When it has reached a critical size, the remaining cross-section is traversed by fast, unstable crack growth, usually within one cycle. After an introduction of some general terminology (Section 2.1), these three stages of fatigue fracture are discussed below in separate sections for uniaxial, periodic loading. Some aspects of the generalisation to more complex loading conditions are discussed in Section 2.5.

2.1 Cyclic behaviour

In fatigue research, fluctuating loads are often idealised by cyclic loading conditions. The test specimen is subjected to a periodic stress or deformation of a certain frequency, mean value and amplitude. Standard nomenclature in this connection is explained in Figure 2.1, in which a stress cycle of mean value $\sigma_m$ and amplitude $\sigma_a$ has been plotted; similar definitions apply to cyclic strains. An additional parameter, which is often used to characterise the mean stress level, is the load ratio

$$R = \frac{\sigma_{\text{min}}}{\sigma_{\text{max}}}. \quad (2.1)$$

With this definition, $R = -1$ for fully reversed loading ($\sigma_{\text{min}} = -\sigma_{\text{max}}$) and $R = 0$ for zero-tension loading ($\sigma_{\text{min}} = 0$).

Fatigue failure has been found to depend primarily on the stress amplitude $\sigma_a$. Higher stress amplitudes lead to a shorter fatigue life, that is, a smaller number of loading cycles to failure. The influence of the mean stress $\sigma_m$ is secondary, but significant: the higher the mean stress level, the shorter the fatigue life.
Thus, tensile mean stresses are detrimental with respect to fatigue behaviour, while compressive mean stresses are beneficial. Under laboratory conditions, the fatigue life of metals is fairly independent of cycle shape and frequency up to about 100 Hz. This assumption is no longer valid for higher frequencies, at high temperatures and for polymeric materials, because viscous effects and hysteretic heating may then have a considerable influence.

The results of fatigue tests are often expressed in so-called S-N curves. S-N curves, also called Wöhler curves, are obtained by plotting the different stress amplitudes (or sometimes stress maxima) employed in a large number of otherwise identical fatigue tests against the fatigue life, \( N_f \). They usually represent uniaxial tests under fully reversed loading \((R = -1)\). Figure 2.2 shows a typical example of an S-N diagram. Notice that a logarithmic scale is used for \( N_f \), reflecting the wide range of possible fatigue lives. Fracture occurs after relatively few load cycles (typically 10-100) for high loading levels. In the extreme case of only one quarter of a cycle, the fatigue strength (i.e., the value of \( \sigma_a \) associated with a specific fatigue life) equals the fracture strength under monotonic loading. As indicated previously, the fatigue life increases with decreasing stress amplitude. The curve sometimes has a horizontal asymptote for large \( N_f \) (\( > 10^6 \)). Below the asymptotic stress amplitude, which is called the fatigue limit or endurance limit, the specimen has an infinite fatigue life. A fatigue limit is observed for low-strength steels, for which it is 35% to 50% of the tensile strength. Many high-strength steels and virtually all non-metallic materials do not exhibit a fatigue limit. For such cases an endurance limit
is assumed if fracture has not occurred after a very large number of cycles, usually between $10^7$ and $10^9$. The errorbars in Figure 2.2 give an indication of the scatter of fatigue life. For small $N_f$ (10-100) the uncertainty is in the order of 10%, but close to the fatigue limit variations by a factor 100 are not exceptional.

The high stress levels accompanying small $N_f$ generally cause appreciable macroscopic plastic deformation. Since these loads can be supported only for a small number of cycles, the behaviour is referred to as low-cycle fatigue. The corresponding local stress-strain response has been depicted schematically in Figure 2.3(a). It shows that the yield stress in compression is—in the absolute sense—lower than the maximum stress attained in the preceding tension part of the cycle, and vice versa. This phenomenon, which is called the Bauschinger effect, is believed to be caused by internal stresses which are accumulated during the forward deformation and assist the reversed flow. It allows for substantial plastic deformation in each cycle, often much larger than the elastic strain, and consequently in considerable energy dissipation.

Apart from hardening within a cycle, one may observe a gradual decrease of the maximum strain attained in subsequent constant stress range cycles (see Figure 2.4(a)). This so-called cyclic hardening is caused by multiplication and pile-up of dislocations; it is observed in well-annealed metals.
Work-hardened materials often exhibit the opposite effect, cyclic softening (Figure 2.4(b)), due to rearrangement of pre-strain induced dislocation networks. Cyclic hardening or softening is usually limited to the first 10 to 100 cycles, after which a stable hysteresis loop is attained.

A third phenomenon related to low-cycle fatigue is cyclic creep or ratchetting. This means that the plastic strain in the tensile loading portion of the cycle is not opposed by an equal amount of yielding in the compression part, causing an accumulation of mean plastic strain (Figure 2.4(c)). Cyclic creep is usually found in the presence of a non-zero mean stress, but it may also be the consequence of a pronounced yield anisotropy between tension and compression.

As opposed to low-cycle fatigue, the term high-cycle fatigue is associated with low stress amplitudes and thus with long fatigue lives. Plastic flow is limited to the microscale, that is, the size of the plastic zone is small compared with the crack length. The macroscopic stress-strain behaviour is essentially linear elastic, with negligible energy dissipation (Figure 2.3(b)). Consequently, phenomena such as cyclic hardening/softening and cyclic creep do not play a significant role in high-cycle fatigue. The transition from low-cycle fatigue to high-cycle fatigue is situated in the S-N curve between $10^3$ and $10^5$ cycles for metals. A clear demarcation does not exist; here, we will speak of high-cycle fatigue if the macroscopic plastic strain remains much smaller than the elastic strain, and of low-cycle fatigue if not.

### 2.2 Crack initiation

The initiation phase of fatigue cracks usually extends over only a small percentage of the fracture surface, but particularly in high-cycle fatigue it may constitute a large fraction (e.g. 80%) of the fatigue life. In fact, there is no macrocrack initiation at all for stresses below the fatigue limit. On the other hand, the presence of material defects, such as inclusions, voids or pre-existent microcracks, may drastically accelerate or even obviate the need of nucleation.

A commonly accepted definition of the exact moment of crack initiation does not exist. This definition can be said to depend on the scale of observation. For example, a mechanical engineer might consider a crack to have been initiated if it can be detected by the specific detection technique used, while a material scientist is likely to consider the nucleation of microstructural damage as the initiation stage. The latter point of view will be taken here, but even under this assumption the exact transition point from initiation to propagation remains somewhat arbitrary.

In metals that do not contain significant defects, the initiation of fatigue cracks is related to the development of so-called persistent slip bands. These are narrow (in the order of 10 μm) intragranular bands in which cyclic microplastic deformation is localised. They develop in sufficiently strained grains along a single slip system which is favourably oriented with respect to the direction of maximum shear stress. Plastic slip along individual gliding planes within the persistent slip bands is not fully reversed, so that the amount of net slip varies over the thickness of the bands. These slip offsets produce a surface roughening where the persistent slip bands emerge at the specimen surface (Figure 2.5). Within the roughened areas so-called extrusions (microscopic ‘hills’) and/or intrusions (‘valleys’) are formed. The latter act as micro-notches, the stress concentration promoting additional slip and thus causing them to grow deeper and eventually to develop into a (micro-)crack.

As indicated above, persistent slip bands can develop only in grains that are strained beyond a certain...
limit. Since their formation is a prerequisite for the crack nucleation process in defect-free metals, this threshold strain automatically implies the existence of a fatigue limit. The development of slip bands is retarded by the presence of grain boundaries, which is the explanation of the experimental observation that finer grained materials generally have longer initiation lives than materials consisting of larger grains. Analogously, materials that have a higher yield strength have a better resistance to fatigue damage initiation as a result of their improved resistance to the formation of slip bands.

The intrusion mechanism of crack initiation also establishes the fact that the initiation of fatigue cracks is primarily a surface phenomenon. In practice, the role of the specimen surface in fatigue life is even more important because of the presence of surface defects, such as scratches, machining marks, etc. The accompanying stress concentration effect causes cracks often to be initiated at such imperfections. On the other hand, fatigue cracks may also emanate from defects in the bulk of the material. Voids, gas entrapments or inclusions can act as internal surfaces, thus allowing for the formation of intrusions and eventually of microcracks. In materials containing hard particles, cracks may nucleate by debonding of the inclusion from the matrix. Debonding is an important initiation mechanism also in quasi-brittle materials (concrete, ceramics, etc.) and in composites. Crack initiation in polymers is usually due to crazing or shear-banding.

2.3 Stable crack growth

Immediately after their nucleation, fatigue cracks in ductile materials tend to grow along the persistent slip band in which they nucleated. The deformation mechanism responsible for the growth is essentially identical to the initiation mechanism: single slip in a plane of maximum shear stress. This so-called stage I growth is quite small, usually of the order of several grains. As cycling continues, other slip systems become active and crack growth occurs by simultaneous or alternating flow along two systems. This duplex slip mechanism, termed stage II, results in a planar, transcryalline crack path normal to the far-field tensile stress.
Stage II crack growth often comprises a large fraction of the fracture surface. In many engineering alloys this part is marked by striations. These are concentric ripples, whose spacing corresponds to the crack growth during one loading cycle. They are believed to be formed by alternate blunting and sharpening of the crack tip in the tensile and compressive portions of the loading cycle respectively. Striations can be made visible only with electron microscopy. They should not be confused with their macroscopic counterparts, clam shell marks (also called beach marks), which are visible with the naked eye. Clam shell marks represent periods of crack growth under different conditions rather than individual load excursions. In post-mortem analyses of fatigue fractures they can provide valuable information, such as the location where the crack was initiated and its growth direction.

The rate of growth of a fatigue crack (of length \(a\)) subjected to \((N)\) constant amplitude stress cycles is usually expressed in terms of the crack length increment per cycle, \(da/dN\). Experiments have shown that this quantity can be related to the stress intensity range, \(\Delta K\), which is defined as

\[
\Delta K = K_{\text{max}} - K_{\text{min}}.
\] (2.2)

\(K_{\text{max}}\) and \(K_{\text{min}}\) are stress intensities according to linear elastic fracture mechanics, representing the stress state at the crack tip under the far-field (mode I) stresses \(\sigma_{\text{max}}\) and \(\sigma_{\text{min}}\) respectively. If \(\sigma_{\text{min}} < 0\), \(K_{\text{min}}\) is usually set to \(K_{\text{min}} = 0\) to account for the fact that crack propagation is practically insensitive to negative loading excursions in constant amplitude testing. This insensitivity can be explained by the concept of crack closure: in the compressive part of the loading cycle the crack faces meet and the crack is closed, so that the driving force for crack growth vanishes. In fact, crack closure effects may occur already for positive stresses due to residual plastic deformation in the wake of the crack tip, roughness of the crack surface or the presence of a corrosion layer.

A typical log-log plot of \(da/dN\) versus \(\Delta K\) is shown schematically in Figure 2.6. A gradual increase of the crack growth rate with increasing intensity range is observed for a rather wide range of intermediate values of \(\Delta K\). In this regime crack growth occurs by the duplex slip mechanism termed stage II above and the growth rate is quite insensitive to microstructure and load ratio. For extreme values of \(\Delta K\), both low and high, there is a steep rise in crack growth rates with increasing \(\Delta K\). For low \(\Delta K\), crack growth

![Figure 2.6: Crack growth rate for zero-tension loading.](image)
is extremely slow, and for $\Delta K$ below the threshold value $\Delta K_{th}$ there is no growth at all. Crack growth in the near-threshold regime is governed by the stage I mechanism (single shear). It is strongly affected by crack closure effects, even for $R \geq 0$. The rapid increase of $da/dN$ for high $\Delta K$ is caused by the emergence of static fracture modes, in addition to stage II growth, as the maximum stress intensity during the cycle, $K_{max}$, approaches the critical stress intensity factor in monotonic fracture, $K_c$. Plastic effects become more important and an increasing amount of microvoid coalescence and/or cleavage is visible on the fracture surface. For $R = 0$, the crack growth curve is bounded exactly by the critical value $\Delta K = K_c$, because higher levels of $\Delta K$ cause instantaneous fracture (see Figure 2.6). Both for low and for high $\Delta K$ the growth rate is affected considerably by mean stresses and metallurgical variables, such as the yield strength, thermomechanical treatments and preferred orientations. Tensile mean stresses cause the growth rate to be higher and the threshold and critical values of $\Delta K$ to be lower. Generally speaking, more ductile materials (which often have a lower yield strength) have a better resistance to fatigue crack growth, because more energy must be dissipated during the fracture process. A significant sensitivity to grain size exists only in the low-$\Delta K$ regime, with larger grain sizes leading to smaller growth rates and a higher threshold $\Delta K_{th}$. A possible explanation for this seemingly anomalous behaviour can be found in the influence of roughness-induced crack closure.

It is important to note that the relation between $da/dN$ and $\Delta K$ is meaningful only if a linear elastic fracture mechanics description is appropriate. This condition is violated for instance for cracks whose length is comparable to the characteristic microstructural dimension (e.g. the grain size). Growth of these small flaws is dominated by microstructural effects, so that the continuum assumption made in fracture mechanics is no longer valid. Other examples are cracks for which the near-tip plastic zone is comparable to the the crack size and cracks which are engulfed by the plastic strain field of a notch. Small cracks (typically smaller than 0.5 mm) have been observed to propagate significantly faster than long cracks when subjected to the same nominal stress intensity, or even to grow at $\Delta K$ values for which they were expected to remain dormant (i.e. $\Delta K < \Delta K_{th}$; see Figure 2.6). This so-called small crack problem may lead to dangerous overestimates of fatigue lives if it is not taken into account.

2.4 Final fracture

At a certain stage in the crack propagation process the maximum stress intensity reaches the critical value for monotonic fracture, i.e. $K_{max} = K_c$. Crack growth then becomes unstable and the remaining cross section is traversed by dynamic fracture, usually within one cycle or a few cycles at the most. The corresponding fracture surface is much rougher than the stable growth region. Its size may vary from a small fraction of the cross section in relatively low stressed components to almost the entire area at high loading levels. The specific fracture mode by which the final fracture occurs (cleavage, ductile fracture or a combination of these) depends on the material ductility, stress state and frequency.

2.5 Complex loading

Up to this point, the stresses responsible for fatigue damage have been assumed uniaxial and periodic. The study of fatigue under these simple loading conditions provides valuable insight into the mechanical processes by which fatigue failure occurs. However, structural components used in practical applications are often subjected to multiaxial loading with varying amplitudes, mean levels and
frequencies. Both the complexity of stresses and the variation of, for instance, the stress amplitude may activate phenomena which cannot be observed in uniaxial, constant amplitude fatigue, but which may have a significant influence on fatigue life.

Multiaxial stress states are encountered for instance in shafts, which are often subjected to both torsional and tensile (or bending) cyclic loads. If the different components of the loading have the same frequency and are in phase, the term proportional loading is used. The ratio of the components then remains constant throughout the loading process. In case of non-proportional loading this ratio varies during a cycle, because the components are not in phase or even have different frequencies. Non-proportional loading has been found to lead to shorter fatigue lives as compared to proportional loading. The consequences of multiaxial loading are visible particularly in the crack growth phase. It results in a mixed-mode stress state at the crack tip, which may have a significant influence on the crack growth direction and—due to sliding friction of the faces of the crack surface—on the growth rate.

In bending, but also in the uniaxial loading of components which contain stress concentrators, a significant size effect is observed: larger specimens generally have shorter fatigue lives than smaller ones when subjected to the same nominal (bending) stress. This size effect is essentially of a stochastic nature. It is caused by the fact that the probability of a crack being initiated is higher in a larger specimen, since the smaller stress gradient in this specimen causes a larger fraction of the cross-section to be subjected to relatively high stresses.

An example of the influence of stress amplitude variations on fatigue life is the marked effect of incidental or periodic overloads among otherwise constant amplitude cycling. If it is applied early in the fatigue life, a tensile overload may cause premature crack nucleation and thus shorten the fatigue life considerably. In the crack growth stage, on the other hand, application of a tensile overload is often followed by retardation of crack growth over a certain distance, or even by complete crack arrest. This behaviour is controlled by residual deformation in the wake as well as ahead of the crack tip due to the overload cycle. In the wake of the tip, residual positive plastic strains induced during the loading part of the overload cycle cause crack closure to play a more important role. Ahead of the crack tip, a zone of residual compressive stresses is formed during the unloading portion of the overload cycle. These compressive stresses also have a retarding effect on crack growth. The distance over which growth retardation occurs after an overload depends on the size of the plastic zone induced by the overload, and thus on the severity of the overload cycle. After having traversed this plastic zone, the crack resumes its pre-overload growth rate. As opposed to a tensile overload, the application of a compressive overload can lead to acceleration of crack growth, but not to the same extent as the retardation after a tensile overload of the same magnitude. The underlying mechanisms are the opposite from those accompanying tensile overloads: reduced crack closure and residual tensile stresses ahead of the crack tip.

Transient growth phenomena such as those following overloads are called load sequence effects or load interaction effects. They are also observed when several blocks of different constant amplitude cycles are applied to a test specimen. The sequence of the respective blocks may then have a considerable effect on the fatigue damage induced. For instance, the application of a block of high stress amplitude after a block of low amplitude may result in a sudden burst of crack advance as a consequence of the abrupt increase in peak tensile stress. On the other hand, a high-low block sequence can cause pronounced crack retardation or even crack arrest.
Chapter 3

Engineering practice

Fatigue failure is an important criterion in the design of many engineering components. More than one century of fatigue research has yielded a multitude of strategies, diagram types, empirical relations, rules of thumb, etc., that can be of aid in designing against fatigue. These design tools can be classified in a number of approaches to fatigue failure and design against fatigue, which are laid out in Section 3.1. The two underlying philosophies, total-life design and defect-tolerant design, are subsequently elaborated upon in Sections 3.2 and 3.3. In Section 3.4 some attention is paid to material aspects of fatigue design.

3.1 Engineering approaches

The dimensioning and material choice of engineering structures and components essentially consists of balancing material and machining costs on the one hand and reliability, safety, and thus potentially much higher costs, on the other hand. A number of strategies have been developed to weigh these conflicting factors with regard to fatigue failure. The simplest and without doubt the safest among these is the so-called infinite-life philosophy. It requires design stresses to be safely below the fatigue limit, so that there will be no fatigue damage at all. Although rather conservative, this is a good design criterion for parts subjected to many millions of cycles, like for instance engine valve springs.

For many components service life is limited by factors other than fatigue. Requiring an infinite fatigue life is then unnecessary and economically undesirable. But even if the time to failure is expected to be governed by fatigue, designing for a finite life is usually quite satisfactory. The practice of designing for a finite fatigue life is known as safe-life design. It requires some knowledge of the load spectra which are imposed on a structural component in service. The component's fatigue life under these loading conditions is estimated by empirical rules, numerical analyses, laboratory tests, or a combination of these. This fatigue life, modified with a safety factor in terms of life or in terms of load, should exceed the anticipated service life. Safe-life design has found a widespread use in industry. Noteworthy examples of components to which it is applied are roller-bearings, shafts and pressure vessels.

Fail-safe fatigue design criteria were developed by aircraft engineers. They could not tolerate the added weight required by large safety factors nor the danger to lives implied by small safety factors.
Fail-safe design recognises that fatigue cracks may occur and arranges the structure such that cracks will not lead to failure of the structure before they are detected and repaired. Multiple load paths and crack arresters are some of the means used to achieve fail-safe design. These structural measures are complemented by periodic inspection (using reliable detection equipment) and repair or replacement of components when necessary.

A refinement of the fail-safe philosophy is called damage-tolerant design. It assumes that cracks will exist—caused either by processing or by fatigue—and uses fracture mechanics analyses and tests to check whether such cracks will grow large enough to produce failures before they are sure to be detected by periodic inspection. This approach is employed mainly in the aerospace industry.

### 3.2 Total-life approach

The classical fatigue design methods of infinite-life and safe-life are based on the expected total fatigue life to failure under the anticipated loading conditions. The total fatigue life incorporates the number of fatigue cycles to initiate a dominant crack and to propagate this crack until catastrophic failure occurs. Thus, no distinction is made between the different stages of fatigue damage development. In practice total-life design is often interpreted as design against damage initiation. Although strictly speaking incorrect, this simplification is justified in high-cycle fatigue by the fact that the crack initiation life constitutes a major component (up to 90%) of the total life.

The S-N curve, introduced in Chapter 2 (see Figure 2.2), is a typical exponent of the total-life approach. It gives the number of loading cycles to complete failure of a laboratory test specimen which is subjected to a constant stress amplitude. Although usually generated for smooth specimens, this data is also used in the analysis of notched parts with (non-singular) stress concentrations, by taking the local stress amplitude at the notch root as \( \sigma_n \). The corresponding fatigue life is then considered to represent the initiation and early (say stage I) propagation.

When drawn on a log-log scale, the high-cycle part of an S-N curve can be approximated quite adequately by a linear relationship. Thus, the relation between the stress amplitude and the number of load reversals (a constant amplitude fatigue cycle is usually composed of two load reversals) can be written in the form proposed by Basquin (1910):

\[
\frac{\Delta \sigma}{2} = \sigma_n = \sigma'_f (2N_f)^b,
\]

where \( \sigma'_f \) is the fatigue strength coefficient (which, to a good approximation, equals the fracture strength in monotonic tension for most metals) and \( b \) is known as the fatigue strength exponent or Basquin exponent. A similar relation, but in terms of the plastic strain amplitude \( \Delta \varepsilon_p / 2 \), was proposed independently by Manson (1954) and Coffin (1954) for low-cycle fatigue:

\[
\frac{\Delta \varepsilon_p}{2} = \varepsilon'_f (2N_f)^c.
\]

Here \( \varepsilon'_f \) is the fatigue ductility coefficient (which is experimentally found to be approximately equal to the fracture ductility in monotonic tension) and \( c \) is the fatigue ductility exponent.

As argued for instance by Manson and Hirschberg (1964), combination of (3.1) and (3.2) yields a relation which is applicable in both the high-cycle and low-cycle regime. For this purpose, the stress...
range in equation (3.1) is replaced by $\Delta \sigma = E \Delta \varepsilon$, with $\Delta \varepsilon$ the elastic strain range. Addition of the resulting expression for $\Delta \varepsilon$ and the Manson-Coffin relation (3.2) and noting that $\Delta \varepsilon = \Delta \varepsilon_e + \Delta \varepsilon_p$ then leads to

$$\frac{\Delta \varepsilon}{2} = \frac{\sigma_f}{E} (2N_t)^b + \varepsilon_f (2N_t)^c. \quad (3.3)$$

This equation forms the basis of the strain-life approach to fatigue design and has found widespread application in industrial practice. The first and second terms on the right hand side of (3.3) are the elastic and plastic components, respectively, of the total strain amplitude. These contributions have been plotted in Figure 3.1, along with the total strain amplitude, as a function of the number of load reversals. At short fatigue lives, the plastic strain amplitude is more dominant than the elastic strain amplitude and the fatigue life of the material is controlled by ductility. At long fatigue lives, the elastic amplitude is more significant than the plastic strain amplitude and the fatigue life is dictated by the fracture strength. In the transition region between these extremes, both contributions are relevant.

The aforementioned empirical relations for fatigue life pertain to fatigue loads where the mean stress level of the imposed fatigue cycle, $\sigma_m$, is zero. However, positive mean stresses are known to have a detrimental effect on fatigue resistance. Mean stress effects can be represented in terms of constant-life diagrams, in which different combinations of the stress amplitude and mean stress are plotted to provide a constant fatigue life; examples are Haigh-diagrams and Smith-diagrams. A number of empirical relations have also been developed to quantify the mean stress effect. The best-known among these is the—rather conservative—Goodman relation (Goodman, 1899)

$$\sigma_a = \sigma_{a0} \left( 1 - \frac{\sigma_m}{\sigma_f} \right), \quad (3.4)$$

where $\sigma_a$ is the fatigue strength for a non-zero mean stress $\sigma_m$ and $\sigma_f$ is the fracture strength in monotonic loading; $\sigma_{a0}$ denotes the fatigue strength (for a fixed life) under fully reversed loading ($R = -1$), which can be determined from an S-N diagram or from an empirical relation such as (3.1).

Total life approaches to design against variable amplitude fatigue usually involve the concept of cumulative damage. Each loading cycle or block of identical cycles in the component’s loading history is considered responsible for a specific amount of material damage. These contributions are added
and the component is expected to fail when this cumulative damage reaches a critical value. A variety of cumulative damage criteria have been proposed, but the oldest of these, the Palmgren-Miner law (Palmgren, 1924; Miner, 1945), still seems to be the most widely used. It states that \( N_i \) stress cycles of a specific amplitude and mean level consume a fraction \( N_i/N_{fi} \) of the total fatigue life. Here \( N_{fi} \) denotes the component’s fatigue life if it were subjected only to cycles of that specific amplitude and mean level; \( N_{fi} \) can be determined directly from constant amplitude data. Failure occurs when

\[
\sum_{i=1}^{n} \frac{N_i}{N_{fi}} = 1,
\]

that is, when the sum of the fatigue life fractions due to the respective cycle blocks reaches unity. It can easily be verified that failure is predicted after exactly \( N_f \) cycles under constant amplitude loading \((n = 1)\).

The Palmgren-Miner law is a linear damage accumulation rule: the expended fraction of the fatigue life increases linearly with the number of cycles. The order in which the different cycles/blocks are applied does not affect the fatigue life, so that (3.5) cannot account for load sequence effects. Empirical modifications of the Palmgren-Miner rule have been developed which do show a sequence effect. However, the inability to describe load interaction effects is not a very severe limitation in most design situations, since the exact magnitude of the different components of the future loading is seldom known, let alone their sequence.

In some applications the identification of loading cycles is rather straightforward. For instance, stresses are usually periodic in rotating machinery, so that a cycle corresponds to one revolution. However, stresses due to wind or wave loading, but also those acting on aircraft, automobiles, etc., are essentially of a random nature. Cycle-based techniques such as cumulative damage rules can then be applied only if the continuously fluctuating loads are replaced by an equivalent sequence of well-defined discrete cycles. A number of so-called cycle counting procedures have been developed to perform this reduction of complex loading histories to discrete cycles of a certain amplitude, mean value, etc. A simple example of a counting procedure is peak-cycle counting. It takes every peak in the stress history as \( \sigma_{\text{max}} \) and the subsequent minimum as \( \sigma_{\text{min}} \) of an equivalent stress cycle, as illustrated in Figure 3.2. The determination of the corresponding stress range and mean stress is then straightforward. The discrete cycles are assembled in blocks of (nearly) equal amplitude and mean value, to which a cumulative damage rule can be applied to compute the fatigue life. Peak-cycle counting and a number of equally simple methods generally result in overly optimistic life estimates. More accurate algorithms have

![Figure 3.2: Peak-cycle counting; cycles ordered by amplitude.](image)
been developed; these are more complicated, but they can usually be implemented quite efficiently in computer programs.

Total-life design methods are oriented highly on uniaxial stress states. Their application to components in which stresses are essentially multiaxial is far from straightforward. These situations are usually dealt with by defining an effective stress (or strain) quantity, in the same fashion as in plasticity models. This approach seems reasonable for the initiation stage of fatigue damage, because (cyclic) plastic deformation plays an important role in crack initiation. However, most effective stress definitions do not capture the effect of non-proportional loading on fatigue life properly. Furthermore, the effective stress concept does not explicitly account for the influence of multiaxial stresses on crack growth, in particular the significant effect of frictional sliding between the faces of the fatigue crack. One should therefore be careful when dealing with multiaxial stress states.

### 3.3 Defect-tolerant approach

In contrast with total-life approaches, in which the crack initiation stage is considered to form a substantial part of the entire fatigue life, the basic premise in damage-tolerant (and also fail-safe) design is that all engineering components are inherently flawed. The size of pre-existing flaws is determined by non-destructive (visual, ultrasonic, X-ray) flaw detection techniques. If no flaw is found in the component, the size of the largest (undetected) initial crack is estimated from the resolution of the detection equipment. The useful fatigue life is then defined as the number of fatigue cycles to propagate the dominant crack from the initial size to some critical size. The choice of the critical crack size for the fatigue crack may be based on the fracture toughness of the material, the limit load of the particular structural part or the permissible change in the compliance of the component.

The prediction of the crack propagation life involves empirical crack growth laws based on fracture mechanics. These are in fact approximations of the growth resistance curve of Figure 2.6. In the most widely used, the Paris law (Paris et al., 1961), the region of gradual growth, also called the Paris regime, is approximated by a linear relation (in the log-log space):

\[
\frac{da}{dN} = C \Delta K^m, \tag{3.6}
\]

with \(C\) and \(m\) material constants. Using this relation, the fatigue life can be determined by substituting the appropriate expression for \(\Delta K\) (in terms of \(\sigma_a\) and \(a\)) for the specific configuration and integrating the resulting equation from the initial to the critical crack size.

For high \(\Delta K\)-values the Paris law is no longer accurate; growth rates may be significantly higher than those predicted by it, see Figure 2.6. If (3.6) is applied nevertheless in this regime, dangerous underestimates of the fatigue life may result. A better approximation is obtained if the effect of mean stress is taken into account, for instance by using the Forman relation,

\[
\frac{da}{dN} = \frac{C \Delta K^m}{(1 - R) K_c - \Delta K}, \tag{3.7}
\]

(Forman et al., 1967). Comparison with (3.6) shows that the Paris relation has been modified by the term \((1 - R) K_c - \Delta K\), which decreases with increasing load ratio \(R\) and decreasing fracture toughness \(K_c\), both of which lead to higher crack growth rates at a given \(\Delta K\)-level. With this modification, a reasonable description is obtained in the Paris regime as well as for high \(\Delta K\). Numerous other growth
relations have been proposed which offer improved descriptions in specific $\Delta K$-regimes or for the complete $\Delta K$-range.

Mixed-mode stress conditions are characterised in fracture mechanics by three stress intensity factors. For fracture under monotonic loading, these three intensities can be combined to one parameter, which is then compared to a critical value to determine whether the crack will grow or not; the direction of crack growth is determined from the ratio of the three intensities. In principle, a similar approach can be followed for fatigue crack growth. The stress intensity ranges associated to the three fracture modes must then be reduced to an effective intensity range, which is considered to govern the crack growth rate. However, the definition of an appropriate effective intensity range is far from trivial, particularly in non-proportional loading. It is a topic of ongoing research, as is the determination of the crack growth direction under mixed-mode loading.

Variable amplitude loading can be taken into account in a quite natural fashion in a fracture mechanics description by performing the integration of the crack growth relation (e.g. (3.6) or (3.7)) cycle-wise or, in case of block loading, block by block. The crack length can then be said to act as a cumulative damage variable. In this form, the analysis cannot correctly describe load interaction effects. A number of empirical modifications of the procedure have been proposed, which do take these effects into account for overloads or block loading by defining an effective stress intensity range based on crack closure considerations. Application of the fracture mechanics approach to randomly fluctuating loads is possible if a proper counting procedure is used to reduce the random loading sequence to a finite number of discrete cycles (see Section 3.2).

Disregarding the crack initiation life is often overly pessimistic. For this reason, some attempts have been made to supply the fracture mechanics approach to fatigue design with an initiation mechanism. The initiation phase of these so-called two-stage models is usually based on total-life concepts, i.e. a crack is considered to have nucleated after a certain number of cycles or for a specific value of a cumulative damage parameter. This crack is then described with fracture mechanics in the second stage. These sophisticated models of fatigue fracture are quite attractive from a theoretical point of view, but they bear the disadvantage of containing a large number of (material)-parameters, which are difficult to quantify. Two-stage approaches are therefore probably too complicated for standard design purposes, where the uncertainties in for instance loading conditions are substantial and relatively large safety factors are used.

### 3.4 Material aspects

The choice of appropriate materials is an important aspect of the design of structures and components. Material properties have an important effect on the product’s appearance, manufacturing process and cost, but also on its strength, reliability and service life. In situations where fatigue life is an important criterion, the choice for a specific material is based primarily on the expected loading conditions. If a part is to be subjected to typical high-cycle loading, that is, to a large number of fluctuations of a relatively low amplitude, one may expect the fatigue life to be governed primarily by crack initiation. A strong material (i.e. one with a high yield strength) is then a natural option, since these materials have a better resistance against crack initiation. If crack growth is expected to constitute an extensive fraction of the fatigue life, on the other hand, a defect-tolerant approach is appropriate. It looks for materials with a high resistance to crack growth, i.e. with a high ductility. Unfortunately, ductile materials usually have a lower yield strength, while high-strength materials are less ductile. The choice
for a specific material therefore always involves careful balancing of crack initiation and crack growth characteristics.

It has been mentioned before that fatigue damage is frequently initiated at the component’s surface. For this reason, methods of improving the material’s resistance against crack initiation locally at the surface may be highly attractive. Most of these treatments rely on the (often combined) effects of an increased yield strength and compressive residual stresses. Well-known examples are thermal hardening, carburising, surface rolling and shot-peening.
Chapter 4

Computational approaches

The exponential growth and rapid spread of computing power during the past two decades has led to a considerable interest in numerical simulation techniques of all kinds of (mechanical) processes. The analysis of failure phenomena is currently an important focal point of these developments. The employment of powerful computers renders the empirical/analytical design criteria discussed in the previous chapter—which are limited to one or two dimensions, simple geometries and simple loading conditions—applicable to real engineering problems. Using numerical analysis tools such as the finite element method, stresses and deformations can be computed in complex configurations and under complicated loading. With the current computing technology, three-dimensional analyses are becoming more and more feasible, which is important particularly in failure analysis because failure is almost always essentially three-dimensional in practice.

Existing computational approaches to fatigue analysis seem to be highly oriented on the empirical/analytical design procedures discussed in Chapter 3. Damage initiation analyses are usually performed by feeding numerically computed deformation or stress fields into total-life design criteria. This approach is discussed in Section 4.1. Crack growth can be dealt with by numerical fracture mechanics techniques, which are presented in Section 4.2. But the availability of efficient non-linear numerical analysis methods also offers the opportunity to think of fundamentally new approaches to fatigue modelling. The essentials of one of these, Continuum Damage Mechanics, are laid out in Section 4.3.

4.1 Damage initiation

As mentioned in Section 3.2, total-life data (e.g. S-N curves) can be used for fatigue crack initiation analyses of notched parts if the local stress amplitude at the notch root is substituted for $\sigma_n$. In two- or three-dimensional analyses, an equivalent measure of stress or strain must be used. However, analytical solutions of these deformations or stresses can be obtained only for a small number of simple geometries. In more complicated situations, the analytical approach can only be applied if a reasonable estimate of the notch root stress is available. In these situations a numerical approach becomes highly attractive. The finite element method can generate accurate solutions of the deformation and stress fields in complicated configurations. These solutions can then be used directly in total-life criteria.
If the material behaviour is assumed to be linear elastic—as is commonly done in high-cycle fatigue—and the loading is proportional, one linear elastic finite element analysis suffices to relate the maximum (equivalent) stress amplitude in the structure to the load level amplitude. For constant amplitude loading, the maximum stress amplitude can then be related to fatigue life using for instance the S-N diagram or the Basquin relation (3.1). Alternatively, if a certain fatigue life is required, the maximum load level can be computed from the fatigue strength. For variable amplitude loading a cumulative damage law must be specified, but the procedure remains essentially the same.

As a refinement of the use of a cumulative damage law, a cumulative damage field variable is sometimes defined. The value of this variable in a material point reflects the fatigue damage induced by the deformation history of that point. Thus, this variable represents fatigue damage locally instead of at the structural level. Its evolution is defined in terms of the stress cycle amplitude (for instance based on (3.1)) or the strain amplitude (cf. relation (3.3)). The distribution of the damage variable at a certain stage in the loading process can be computed by integrating this evolution law over the loading history. When the damage reaches a critical value (which is usually defined as 1), a (macro-)crack is said to have been initiated.

In combination with linear elastic material behaviour, defining damage as a field variable does not seem to yield any substantial advantage for initiation analyses, since the location and value of the maximum stress amplitude remains the same throughout the loading process. However, particularly in low-cycle fatigue analyses these models are sometimes combined with sophisticated non-linear material models, which may affect the distribution of damage. These constitutive descriptions include cyclic plasticity models with (non-linear) kinematic hardening and—if the loading frequency has a significant effect on fatigue resistance—visco-plastic models. The growth of damage is usually related to the total or plastic strain rate in these models; as a result, the loading need not be composed of discrete cycles, but can vary arbitrarily in time without the need for a counting procedure.

4.2 Numerical fracture mechanics

The growth of fatigue cracks is commonly represented in a linear elastic fracture mechanics framework (see Section 3.3). The simulation of crack growth can then be performed step by step. First, the stress intensity range \( \Delta K \) is computed for the current configuration. This value is substituted in an appropriate growth law (e.g. (3.6)) to compute the crack extension. The crack tip is shifted by this crack increment, after which the procedure is repeated for the new configuration, etc.

For the calculation of stress intensity factor ranges, standard numerical techniques for fracture under monotonic loading can be used. A number of semi-analytical methods have been developed, but these are usually limited to rather specific configurations or loads. Fully numerical techniques are the boundary element method and the finite element method. We will concentrate on the finite element method, but developments in the boundary element method run quite parallel. The stress intensity factor can be determined directly from the numerically computed stress or displacement field by extrapolation to the crack tip. Alternatively, it can be determined indirectly by using its relation to other fracture mechanics quantities. Examples are virtual crack extension and compliance techniques to compute the energy release rate, but also Rice's J-integral method, which has gained interest lately. Accurate values of \( K \) can only be obtained if the stress field around the crack tip is described accurately. Due to the singularity of the stress field in the crack tip this would necessitate an extremely fine mesh of standard finite elements. For this reason, special crack tip elements are applied, which incorporate the
The simplest of these crack tip elements are quarter-point elements. These are standard quadratic elements with their mid-side nodes shifted half way in the direction of the crack tip (Barsoum, 1976).

The second component of the simulation procedure, the translation of the crack tip, is less standard. Static fracture analyses are usually limited to determining whether a crack of a certain size will grow or not, that is, to determining the stress intensity and comparing it with a critical value. The exact path and velocity of the subsequent unstable propagation are of less importance. In fatigue, however, it is the growth rate (and direction) in the stable growth phase which determines the fatigue life. After translation of the crack tip, a new element mesh must be generated to compute the stress intensity in the new configuration. Automatic re-meshing procedures have been developed for non-linear fracture mechanics analyses, but these are quite expensive and may—particularly in three dimensions—fail in certain circumstances.

Linear elastic fracture mechanics concepts can only be applied if the plastic zone at the crack tip remains small. Particularly in low-cycle fatigue, plastic straining occurs in quite a large zone, or even in the complete component. One would then have to use non-linear fracture approaches, combined with a suitable cyclic plasticity model. Some authors have suggested the use of the J-integral range or crack tip opening range as fatigue crack growth criteria. Alternatively, one might concentrate the non-linearity in a line, using the cohesive zone concept. In these type of models the actual crack is extended by a fictitious crack in which stresses can still be relayed. These stresses are a function of the (fictitious) crack opening. Neither type of model seems to have received much attention. The formulation of non-linear fracture approaches may be more complicated than the linear approach, the computational treatment is also far from straightforward. Because of the non-linearity in the constitutive relation, the analysis cannot simply be restarted after each crack increment. Instead, the field variables must be interpolated from the old mesh to the new mesh using a suitable transport algorithm (see for instance Segal, 1993).

A disadvantage of the fracture mechanics analysis of fatigue is that it concentrates on the growth of one dominant crack. In practice, a number of fatigue cracks may grow together, one crack may bifurcate, or a crack may be arrested (for instance due to the presence of crack arresters) while a new crack is initiated at some other location. These situations are difficult to deal with using numerical fracture mechanics. Furthermore, if the component is (nominally) defect-free, an initial crack must be defined to start the analysis. The location and direction of the initial crack may have an important effect on the fatigue life that is computed. It is often chosen quite arbitrarily or based on experience. The latter disadvantage can be avoided by using the two-stage approach mentioned in Section 3.3. In this approach, a cumulative damage variable is used to determine the moment and location of crack initiation. A crack of a certain initial size is then introduced, the growth of which is analysed using fracture mechanics. If the cumulative damage is updated also after the initiation of a crack, it may also be possible to simulate crack arrest and subsequent growth at some other location.

4.3 Continuum damage mechanics

Continuum damage mechanics shares its basic hypothesis with the cumulative damage models discussed in Section 4.1: material damage is represented by a macroscopic field variable, which represents the material damage—such as microcracks, voids, etc.—in a continuous fashion. The fundamental difference from the cumulative damage approach lies in the role played by this damage variable in the
constitutive model. In the cumulative damage approach, it is a derivative quantity, dependent on the strains or stresses, but it does not have an effect on the mechanical response. This simplification is removed in continuum damage mechanics: the damage variable appears in the constitutive model and has an explicit effect on the stress. As a consequence, the material model is highly non-linear even if the continuum damage concept is applied to linear elastic material behaviour.

The distinction between cumulative damage and continuum damage is not always made as rigourously in the literature as it is made here. The term continuum damage mechanics is sometimes used for models in which there is no effect on the constitutive law. The use of the term damage or damage variable is even more confusing; it is usually connected with continuum or cumulative damage models, but it is also used as a parameter representing the global loss of integrity or functionality at the component or construction level.

The classical interpretation of the damage variable in continuum damage mechanics is in terms of effective area reduction. Consider a material section with total area $A$ (Figure 4.1). If this segment contains microcracks, voids or other microstructural defects (represented by circles in Figure 4.1), only the remaining area $\tilde{A}$ is effectively occupied by the bulk material. The damage variable, denoted here as $D$, can then be defined as the surface fraction occupied by the defects, i.e.,

$$D = \frac{A - \tilde{A}}{A}. \quad (4.1)$$

With this definition, $D = 0$ characterises fully flawless material, while $D = 1$ represents the complete loss of material coherence.

Due to the presence of voids and cracks, only the effective area $\tilde{A}$ can participate in the transfer of stresses. It can therefore be argued that if a macroscopic stress $\sigma$ is applied to the segment, the net cross-section is subjected to the effective stress

$$\tilde{\sigma} = \frac{\sigma}{\tilde{A}}. \quad (4.2)$$

which, using (4.1), can also be written as

$$\tilde{\sigma} = \frac{\sigma}{1 - D}. \quad (4.3)$$

Figure 4.1: Interpretation of damage variable in terms of effective area reduction.
The effect of material damage on the stress state can be taken into account by replacing the stress $\sigma$ in the constitutive law by the effective stress $\overline{\sigma}$ according to (4.3). For linear elastic material behaviour, Hooke's law

$$\sigma = E \varepsilon$$

then becomes

$$\sigma = (1 - D) E \varepsilon.$$  \hspace{1cm} (4.5)

In this model damage can be interpreted as a stiffness degradation, with $(1 - D) E$ the effective stiffness for a specific value of the damage variable. For $D = 0$ the effective stiffness equals Young's modulus of the bulk material. When the damage variable reaches the value $D = 1$, all stiffness is gone and the material cannot sustain any stress.

The evolution of the damage variable is often related to the deformation. For simplicity, the discussion will be limited to the uniaxial, tensile case, i.e., $\varepsilon \geq 0$. It is assumed that damage can only be progressive if the strain exceeds a threshold value $\kappa$. This threshold may for instance be a function of the damage variable, but it is assumed constant here: $\kappa = \kappa_0$. Obviously, the threshold value $\kappa_0$ must be related to the fatigue limit. If the loading conditions are such that the strain remains below the threshold everywhere in a component, there will be no damage development and the component has an infinite fatigue life. For materials that do not exhibit a fatigue limit, the parameter $\kappa_0$ can be set to zero. As a second assumption, damage is considered to grow only with an increasing strain level. Thus, the damage evolution law can be written as

$$\dot{D} = \begin{cases} g(D, \varepsilon) \dot{\varepsilon} & \text{if } \varepsilon \geq \kappa \text{ and } \dot{\varepsilon} > 0, \\ 0 & \text{else}, \end{cases}$$

(cf. Paas (1990)). This relation must be complemented by an initial condition for the damage variable. The initial value is usually set to $D_0 = 0$, which represents initially defect-free material.

Notice that the evolution of damage is expressed in terms of the strain rate in (4.6) and not in terms of cycle amplitudes. Thus, the loading process need not necessarily be composed of well-defined cycles, which may be important when dealing with random fluctuations. If necessary, however, the evolution equation can quite easily be formulated in terms of cycles by integrating the damage evolution relation (4.6) over one cycle (Paas, 1990). This yields an evolution law of the type

$$\frac{dD}{dN} = G(D, \varepsilon_{\text{max}}, \varepsilon_{\text{min}}),$$

where $\varepsilon_{\text{max}}$ and $\varepsilon_{\text{min}}$ are the maximum and minimum strain attained in the cycle. The number of cycles $N$ is now treated as a continuous, time-like variable; the damage variable $D$ and the strain cycle parameters are considered continuous functions of $N$. Notice that relation (4.7) can also be written in terms of stresses by substituting $\varepsilon = \sigma / (1 - D) E$. For constant amplitude cycling, the damage can be related explicitly to the number of cycles by integrating evolution law (4.7) again. The resulting relation $D = D(N)$ can be regarded as the—generally non-linear—equivalent of a cumulative damage rule. Whether it satisfies the Palmgren-Miner rule for block loading depends on the specific choice for the evolution law. This relation between continuum damage models and total-life relations is often used to determine the parameters of damage models from for instance S-N curves.

The equivalence of the continuum damage and cumulative damage variables is only partial. Whereas the cumulative damage variable has no influence on the constitutive behaviour, damage evolution in the
sense of continuum damage mechanics inevitably results in cyclic softening. In the one-dimensional case, this softening is visible as a decrease of the stress amplitude in constant strain amplitude cycling. In two or three dimensions, the consequences are more far-reaching. Damage growth then results in a reduction of the load-carrying capacity and a redistribution of stresses. Completely damaged material \((D = 1)\) cannot sustain any stress, which represents a state of complete fracture. As a result of this feature, continuum damage mechanics is believed to be applicable also to crack propagation. A crack is then represented by a line on which \(D = 1\) and the zone in front of the crack where \(0 < D < 1\) represents the process zone. Notice that the condition \(D = 1\) in the crack implies that there is no stress singularity at the crack tip. Fracture and crack growth arise as a natural outcome of the deformation history in this approach, which contrasts with the usual approach to fracture analysis where the constitutive model and a fracture criterion are specified separately. The constitutive relations are usually of a phenomenological nature, but some authors have formulated them on the basis of micromechanical considerations (e.g., Lemaitre, 1996). As an intermediate form between cumulative damage and continuum damage models, the effect on the constitutive behaviour is sometimes neglected up to a certain critical value of the damage variable (e.g. \(D = 1\)). When this critical value is reached, the material stiffness is locally set to zero, so that the propagation phase is governed by a continuum damage mechanism.

The concept of continuum damage has been laid out here in a one-dimensional setting. However, the extension to the general, three-dimensional case is not complicated, in contrast with fracture mechanics. Also, the limitation to linear elastic material behaviour is made here only for simplicity. Using the effective stress concept, the damage mechanism can be—and has been—incorporated in a wide range of (non-linear) constitutive models, e.g. elasto- or visco-plasticity. In fact, there have also been developments with regard to thermomechanical processes and creep-fatigue interactions in a continuum damage framework.

Continuum damage models usually fit quite well in the general non-linear finite element methodology, so that the implementation in finite element codes does not seem to pose serious problems. However, damage models (and other constitutive descriptions of material degeneration) are known to suffer from mesh dependence under certain circumstances. At a certain level of damage the mathematical description may cease to be well-posed. In numerical analyses the deformation then localises in a band with the smallest width that can be represented by the spatial discretisation. As a consequence, the analysis becomes highly sensitive to the discretisation, i.e. to the element size and orientation in a finite element context. This complication, which is understood quite well for monotonic fracture, does not seem to have received much attention in fatigue. But even if the localisation of deformation is modelled mathematically correctly, the (finite element) interpolation must be sufficiently fine in order to capture the narrow deformation zones that may emerge. In practical analyses, particularly crack propagation analyses, adaptive meshing techniques or higher-order interpolation methods will therefore probably be necessary.
Chapter 5

Conclusion

As with analytical-empirical approaches to fatigue failure, the most effective numerical approach probably depends on the specific problem at hand. For standard design purposes an infinite-life or safe-life procedure is probably quite satisfactory. If, in addition, critical stresses can be estimated with a reasonable accuracy, there is no need to consider numerical modelling of the failure process. Numerical techniques do come into view for complicated geometries or loading conditions, and in situations where material cost, weight, and/or safety are so important that the simple total-life procedures are insufficient. Examples can be found in the aerospace industry, the construction of chemical, nuclear and off-shore installations, ship-building, etc. But a better understanding and quantification of failure, and thus a reduction of material cost, may also be profitable for less complicated products that are manufactured in mass-production.

In design against fatigue crack initiation, the use of numerical modelling is usually limited to computing stress distributions in complex geometries or under complex loading conditions. Standard, linear elastic finite element techniques often suffice for this purpose, but particularly in low-cycle fatigue more complicated non-linear models are sometimes employed. After the numerical calculation of the stress distributions in the component, the expected initiation life is computed analytically using empirical relations, or read from a diagram. As a somewhat more sophisticated approach, a field variable can be introduced which represents the distribution of cumulative damage in the component. However, this approach cannot represent the effect of the microstructural damage on the macroscopic material response, which may be considerable particularly in low-cycle fatigue.

The accepted treatment of fatigue crack growth is by fracture mechanics. Again, analytical solutions or estimates can be obtained only for relatively simple configurations, so that a numerical approach is often necessary. Numerical crack growth simulation is far more complicated than initiation analysis, even in a linear elastic fracture mechanics context. It is performed by alternately computing the stress field and translating the crack tip according to a suitable growth law. Special elements are needed at the crack tip to compute accurate stress distributions and (expensive) re-meshing must be performed in each growth increment. Non-linear effects (load-interaction effects, crack closure, etc.) are usually taken into account by empirical modifications of the growth law. Fracture mechanics analyses concentrate on one dominant crack. As a consequence, it is difficult to deal with situations where the response is governed by more than one crack path, for instance due to coalescence, bifurcation, or crack arrest and subsequent growth at some other location.
Crack initiation and crack propagation analyses by themselves are rather pessimistic in the sense that part of the fatigue life is neglected. If crack initiation is identified with failure, it is ignored that a cracked component may still have sufficient load-carrying capacity and thus may still function satisfactorily for quite some time. The fracture mechanics approach to fatigue, on the other hand, starts from the assumption that a crack is present in the structure. Thus, the initiation phase of fatigue damage development is ignored completely, whereas this phase may take up a substantial fraction of the fatigue life. These drawbacks of the classical fatigue analysis methods can be eliminated by the construction of models which can represent both the initiation and propagation of cracks. Two approaches can be discerned in this respect: two-stage approaches and models based on a continuous description of damage.

Two-stage models are obtained by coupling classical initiation and propagation procedures. For instance, the cumulative damage concept can be used to model the initiation phase and fracture mechanics to model the propagation phase. An advantage of this approach is that it is founded on well-known and generally accepted concepts for both stages. But this also implies that the method inherits the disadvantages of these fundamental approaches. Moreover, in practice the initiation and propagation phases cannot be separated as sharply as it is done in the model. In fact, they are partly governed by the same micromechanical mechanisms. The parameters related to the change of description are therefore somewhat arbitrary, whereas there may be some redundancy in the parameters of the two individual stages.

The discussion of continuous descriptions of damage focuses on the so-called continuum damage models here. However, most arguments also apply to similar models which may be formulated for instance in a plasticity framework. The quintessence of the continuum damage approach is the introduction of (continuous) field variables which represent material damage in an averaged (continuum) sense. These damage variables have a negative effect on the local constitutive behaviour (e.g., the stiffness), thus representing the degradation of the material properties due to the development of damage. For a certain critical value, the load-bearing capacity locally vanishes completely, which represents complete fracture. In contrast with the fracture mechanics approach, crack initiation and growth are natural consequences of the development and localisation of damage. The initiation and propagation stages of fatigue fracture are modelled in a unified fashion, with a gradual transition from the first to the second stage. There is no need to define an initial crack in defect-free materials, and arrest, coalescence and bifurcation of cracks follow in a natural fashion. The damage concept can be extended to include (visco-)plastic effects, for instance to model cyclic hardening, cyclic creep, load-interaction effects, etc. Since damage is represented in a continuum framework, continuum damage models can easily be cast in a finite element formulation, the implementation of which in (non-linear) finite element codes is rather standard. The theory is intrinsically three-dimensional, so that three-dimensional problems are not more complicated than two-dimensional analyses.

The principal advantage of the continuum damage approach to fatigue modelling, i.e. the representation of a gradual transition from diffuse micro-damage to a macroscopic dominant flaw by localisation of damage and deformation, is at the same time responsible for a major complication. The correct description of localisation of deformation can be quite difficult, both from a theoretical and computational point of view. The localisation problem is understood reasonably well (and solutions have been developed) in the context of monotonic fracture, but continuum damage theory for fatigue seems to be somewhat underdeveloped in this respect.

In conclusion, numerical analyses of fatigue damage initiation are almost standard design tools at present. Numerical (linear) fracture mechanics techniques for crack growth may reach the same status
quite soon, if the problem of re-meshing is solved satisfactorily. But the modelling of non-linear effects in crack initiation and propagation—which may have a significant effect on fatigue life—is still an area of active research, as is the development of models which can represent both stages of the failure process. A possible approach to solving these issues is to start from the classical descriptions of initiation and propagation. However, continuum damage mechanics may serve as a conceptually more attractive alternative to this strategy. It provides a framework in which nucleation and growth of flaws can be treated in a unified fashion and in which non-linearity can be introduced in a natural way. Moreover, although still a phenomenological approach, it seems to be more closely connected to the mechanical processes responsible for fatigue failure. But the practical feasibility of the continuum damage approach to fatigue critically depends on the availability of computational methods for the description of highly localised deformation phenomena. Although the subject of intensive research, these methods are presently available only to a limited extent.
Bibliography


