Background and contents

The “Prediction of lifetime of railway wheels” project was initiated at Chalmers Solid Mechanics in 1994 and was placed under the Competence Centre CHARMEC in 1995. The aim of the project was to predict fatigue life with the help of mathematical modelling. Such a predictive model was deemed to be useful in fatigue design and in preparations of tenders where guarantees of operational life may be requested. These research results should also be of use in optimization of maintenance of railway wheels as well as in practical work to avoid wheel failures.

The present work focusses on the initiation of fatigue cracks in the tread of a railway wheel. The division of the fatigue process into an “initiation” and a “propagation” phase is somewhat questionable since both phases may include the growth of cracks but on different scales. Here, the term “initiation” refers to the first stages of fatigue damage and crack growth where the process can be analyzed by use of criteria that set out from the state of stress as derived, disregarding any effects of cracks. A further discussion is given in section 2.1.

The reverse of the present approach is a fracture mechanics analysis. Here, the growth of an assumed initial crack until final fracture of the component is studied. Fracture mechanics criteria are discussed, but not employed, in the appended papers and they are further discussed in section 2.2.

An important issue, which is somewhat in the grey zone between where continuum criteria and fracture mechanics criteria are
applicable, is the influence of material defects. This issue is studied in papers A and B and will be further discussed in section 5 below.

This introduction starts in section 1 with a brief overview of fatigue failures of railway wheels. In section 2, some fundamentals of fatigue analysis are discussed and methods employed are put in perspective. The issues discussed are contrasted to some complicating features of rolling contact fatigue in section 3. A model for the fatigue of railway wheels is derived and presented in papers C, D and E. This model is described in section 4. In sections 5, 6 and 7, material defects, material anisotropy, random load and random material strength are discussed. Naturally, in order to verify a numerical model, full-scale testing in terms of operational data would be of interest. Some possibilities and problems involved in this are discussed in section 8. Areas for future work are proposed in section 9. Finally, a short review of the contents of the appended papers A to F is given in section 10.

1 Preliminaries – fatigue failures of railway wheels

Railway wheels may fail in a number of different ways, see eg [1], [2], [3], [4] and [5]. Even though wear is the most frequent cause behind wheel rep profilings, this is a fairly harmless (although costly) form of deterioration. Fatigue failures, on the other hand, are more violent in nature and may lead to the break-off of a large part of the wheel. Consequences of such failures include damage to rails and sleepers and to train suspensions and bearings. In rare cases, even derailments may result. These failures may be very costly in terms of economics and human injuries as well as in reliability of train operations.

In papers A and B, different types of fatigue failures are discussed. In short, these can be divided into two categories: surface-induced and subsurface-induced fatigue failures.

Surface-induced failures are normally initiated due to gross plastic deformation of the wheel material close to the running surface. The plastic flow is mainly a consequence of high frictional loading and/or low yield material strength. The cracks normally grow some millimeters into the wheel before deviating to the surface and breaking off a piece of the wheel tread.

Subsurface failures occur below the running surface. Normally, the initiation occurs at a macroscopic material defect, see section 5, but cracks can also initiate in a ‘virtually defect-free’ material, see further the discussion on the subject in paper B. After initiation, the cracks usually propagate at a depth of 10 to 30 mm below the wheel tread. If and when a crack finally branches towards the wheel surface, a large piece of the wheel tread will break loose. Consequently, these subsurface-induced failures are potentially more dangerous than surface-induced failures. Subsurface-induced failures are the type of failures that this thesis focusses on.
2 Preliminaries – fatigue analysis

A very brief introduction to fatigue design will be given. The purpose is to highlight some complicating factors in a rolling contact fatigue analysis as compared to a ‘classic’ fatigue analysis. For in-depth information, see texts on basic fatigue, eg [6], [7] and [8].

2.1 Continuum approaches

Uniaxial loading vs multiaxial loading

In an undamaged material, subjected to a uniaxial and cyclically varying state of stress, the fatigue life can be quantified by the relationship between the applied stress magnitude and the material fatigue strength. It turns out that the best correlation is obtained by quantifying the state of stress by its amplitude. A plot of applied stress amplitude vs fatigue life is called a Wöhler curve (or an S-N–curve) and is used in standard fatigue design, see figure 1.

In order to quantify the fatigue impact, the concept of fatigue damage is normally employed. For one stress cycle, the fatigue damage can be defined as , where is the fatigue life in terms of the number of stress cycles for the current stress amplitude.

The issue of evaluating fatigue impact from a sequence of load cycles of different amplitudes is far from trivial. The most commonly used rule for damage accumulation is the Palmgren–Miner linear damage accumulation formula [9]. Here the total (scalar) damage is taken as a sum of partial damage contributions independently of the order of the load cycles within the sequence. Fatigue failure is then assumed to occur at a material point when the total accumulated amount of damage attains the value 1 (unity). In many cases this is, however, not in agreement with results from experiments. Several suggestions on how to adjust the Palmgren-Miner rule according to

![figure 1](image-url)
the load spectrum have been made. A more extensive discussion on
this issue is found in [10] and in paper C.

Further, for many materials, there exists a stress amplitude, \( \sigma_{FL} \),
below which no fatigue failures will occur. This fatigue limit is shown
as a horizontal tail of the S-N-curve in figure 1. A better estimation of
the fatigue limit is obtained if it is appreciated that also the mid value
of the stress during a stress cycle will have an influence on which
stress amplitudes can be sustained without fatigue failures. Such a
combination results in a Haigh diagram, see figure 2.

![Haigh diagram showing admissible combinations of stress amplitudes
\( \sigma_a \) and mid values \( \sigma_m \) if fatigue is to be avoided. The subscripts
stand for \( Y \) = yield, FL = Fatigue Limit, FLP = Fatigue Limit in Pulsating
tension and UTS = Ultimate Tensile Strength. In order to obtain design
values of the fatigue limit, reductions are made with respect to several
parameters (factors \( \lambda, \kappa \) and \( \delta \)) that will decrease the fatigue
strength. The right-hand diagrams show a schematic sketch of such a
reduction.](image)

In a multiaxial state of stress, the fatigue behaviour is more
complicated. The state of stress is now defined, not by one, but by six
components (in the general case). In order to quantify the fatigue
impact under such a state of stress, an equivalent scalar stress
measure is normally adopted. Such an equivalent stress can be based
on an energy measure (compare with the von Mises effective stress).
Fatigue criteria of this type are the Sines criterion and the Crossland
criterion. Alternatively, a measure based on the stress history
projected on a critical plane (compare with the Tresca effective
stress) can be used. An example of such a criterion is the Dang Van
criterion, which has been adopted in the present thesis and which will
now be described.

The Dang Van criterion defines an equivalent stress as a combination
of the fluctuation of the shear stress from its mid value during a stress
cycle and the hydrostatic stress (taken positive in tension). It states that fatigue damage will occur if the inequality

\[
\sigma_{\text{EQ, DV}}(t) = \tau_a(t) + a_{\text{DV}}\sigma_h(t) > \sigma_e
\]  

(1)

is fulfilled during some part(s) of a stress cycle. Here, \(\sigma_e\) is the equivalent fatigue limit and \(a_{\text{DV}}\) a material parameter. This relationship is graphically represented in figure 3.

The use of the fluctuation of the shear stress from its mid value during a stress cycle, \(\tau_a\), in equation (1), will eliminate the effect of a superimposed static shear stress. This is in accordance with empirical data. In his original works, see eg [11] and [12], Dang Van explained this lack of influence by analyzing the fatigue process on two different scales. The ‘macroscopic stresses’, ie the stresses evaluated in a continuum analysis, will give rise to ‘microscopic stresses’ on a scale of the order of some few grain sizes\(^1\). It is then postulated that the fatigue limit corresponds to the upper bound for an elastic shakedown on the microscale. During the shakedown process, a stabilized microscopic residual stress will form and the material will undergo plastic hardening. In [11] a procedure is outlined where this stabilized microscopic residual stress, \(\rho^m_{ij}\), is defined as the residual microscopic stress tensor, \(\rho_{ij}\), fulfilling the min-max problem

\[
\text{dev}(\rho^m_{ij}) = \min \left[ \max \left( J_2 \{ S_{ij}(t) - \text{dev}(\rho_{ij}) \} \right) \right]
\]  

(2)

Here \(J_2\) is the second invariant of the deviatoric stress tensor and \(S_{ij}\) the deviatoric part of the macroscopic stress tensor.

Due to the microscopic residual stress, the ‘local’ deviatoric stress on the microscale will be

\[
s_{ij}(t) = S_{ij}(t) + \text{dev}(\rho^m_{ij})
\]  

(3)

\(^1\) This scale is frequently referred to as the mesoscale.
The local shear stress on the critical plane (note that the hydrostatic stress is an invariant) is obtained as the local Tresca shear stress

$$\tau_{\text{Tresca}}(t) = \frac{s_2(t) - s_3(t)}{2}$$  (4)

Here $s_2$ is the largest and $s_3$ the smallest eigenvalue of $s_{ij}$. The local shear stress, $\tau_{\text{Tresca}}(t)$, is now employed as $\tau(t)$ in equation (1).

The issue of 'local' stresses on the grain size level has been further investigated by Papadopoulos [13], who derived a fatigue initiation criterion similar to the Dang Van criterion from an analysis of plastic deformations on this scale.

The issue of reducing the mid value of the deviatoric stress tensor during a load cycle can also be approached from a geometric point of view. For the case of in-phase loading with fixed principal directions, this becomes quite simple, see [8]. Let us first make some definitions:

- The loading is in-phase if all stress components (and thus all deviatoric stress components) can be expressed as $\sigma_{ij} = a_{ij} + c_{ij}f(t)$, where $a_{ij}$ and $c_{ij}$ are constants and where $f(t)$ is a time-dependent multiplier
- The principal directions are fixed if the Tresca shear stress corresponds to the same plane (through the material point studied) at all instants of time

If the principal directions are fixed, we can choose to express the deviatoric stress tensor in a coordinate system with axes corresponding to the principal directions. We can call this the $x'y'z'$ system,

$$\sigma_{ij}(t) = \begin{bmatrix} \sigma_{xx}(t) & \sigma_{xy}(t) & \sigma_{xz}(t) \\ \sigma_{xy}(t) & \sigma_{yy}(t) & \sigma_{yz}(t) \\ \sigma_{xz}(t) & \sigma_{yz}(t) & \sigma_{zz}(t) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_1^d(t) \\ \sigma_2^d(t) \\ \sigma_3^d(t) \end{bmatrix}$$  (5)

$x y z$ coordinate system  $x'y'z'$ coordinate system

Since all stress components have a common time dependency, we can rewrite equation (5) as

$$\sigma_{ij}^d(t) = \begin{bmatrix} \sigma_1^d(t) \\ \sigma_2^d(t) \\ \sigma_3^d(t) \end{bmatrix} = \begin{bmatrix} a_{11}^d & 0 & 0 \\ 0 & a_{22}^d & 0 \\ 0 & 0 & a_{33}^d \end{bmatrix} + \begin{bmatrix} c_{11}^d & 0 & 0 \\ 0 & c_{22}^d & 0 \\ 0 & 0 & c_{33}^d \end{bmatrix}f(t)$$  (6)

Now, if we plot $\sigma_1^d$, $\sigma_2^d$ and $\sigma_3^d$ as functions of time, the result will be a straight line in 3D-space (in the $x'y'z'$ system). The intuitive mid value of the deviatoric stress tensor would be the midpoint of this line. Thus, for this case, the mid value of the deviatoric stress tensor is obtained simply by taking the mid value of each component.

It could also be noted that if the stress components varied out of phase, the result would be a curve in space. If the principal stress...
directions were not fixed, the result would be a straight line, but in a rotating coordinate system.

In order to evaluate the mid value of the deviatoric stress tensor for a general loading, we need to analyze the time-dependent deviatoric stress tensor in six dimensions, corresponding to the six independent components of the stress tensor. To facilitate the task, this tensor can be mapped onto a representative vector of a five-dimensional Euclidean space, see [14]. The tip of this vector will, during a stress cycle, follow a closed curve in the five-dimensional space. The mid value of the deviatoric state of stress (in the material point considered), \( \sigma_{ij, \text{mid}}^d \), is then defined as the centre of the smallest circumscribed hyper-sphere to this curve\(^2\) (when transformed back to the six-dimensional stress space).

The time-dependent ‘amplitude’ of the deviatoric stress tensor can now be defined as

\[
\sigma_{ij, a}(t) = \sigma_{ij}^d(t) - \sigma_{ij, \text{mid}}^d
\]  

Finally, the pertinent time-dependent ‘amplitude’ of the Tresca shear stress can be evaluated as

\[
\tau_{\text{Tresca}, a}(t) = \frac{\sigma_{1a}^d(t) - \sigma_{2a}^d(t)}{2}
\]

and introduced as \( \tau_a(t) \) in the Dang Van criterion as defined by the inequality (1).

A slightly different approach has been employed in paper E and is further described in [16]. Here, the mid value of the shear stress vector (with three components) during a stress cycle is defined as the centre of the smallest circumscribed circle to the path of the tip of the shear stress vector during a stress cycle. Consequently, this definition considers the state of stress projected on a specific shear plane and the reduction of the mid shear stress has to be carried out for all shear planes studied.

The Dang Van criterion was originally developed for design against fatigue initiation. This means that as long as the inequality (1) is not fulfilled, fatigue will not be initiated. The criterion will, however, not give any information on the fatigue life to be expected if the inequality (1) is fulfilled. In order to estimate the fatigue life, an extension to the Dang Van criterion is proposed and used in papers C, D and E. The issue is further discussed in section 4.4.

**High cycle fatigue vs low cycle fatigue**

In wheel/rail contact, plastic deformations will usually occur even at fairly moderate load levels. After the first load cycle, residual stresses will then be introduced in the wheel rim and the plastically deformed material will also experience plastic hardening. Due to these two effects, a load magnitude that initially causes plastic

---

\( ^2 \) Which, for instance can be found by using an optimization routine, see [15]
deformations, may, after some load cycles, only cause a purely elastic response. This effect is called *elastic shakedown*. However, if the load magnitude is severe enough, elastic shakedown will never occur. Instead, every load cycle will introduce a small additional plastic deformation. Eventually, the deformation will exceed the material ductility and the material will rupture. This process is called *ratchetting*. A third possibility is that each load cycle will introduce plastic deformation, but that the plastic deformations in loading and unloading will cancel each other out leading to a net plastic deformation which is zero. Such a process is called *plastic shakedown*. These phenomena are more thoroughly described in references [17] and [18].

![Response of a material to cyclic loading](image)

**Figure 4** Response of a material to cyclic loading (a) purely elastic deformations, (b) elastic shakedown, (c) plastic shakedown, and (d) ratchetting. After [17].

In cases where plastic deformations are introduced, stress is no longer a good measure of the loading impact on fatigue life. Instead, equivalent strain criteria may be adopted. Also in the latter case, the criteria can be of either the ‘energy’ or the ‘critical plane’ type.

### 2.2 Fracture mechanics approaches

**Linear elastic fracture mechanics (LEFM)**

In linear elastic fracture mechanics, the local stresses in the vicinity of a crack tip are analyzed according to the theory of elasticity. If the crack tip is regarded as infinitely sharp, the calculated stress field will exhibit a singularity that can be quantified by the stress intensity factor, $K$. In order to predict crack propagation, the range of the

---

3. In elastic loading, stress and strain are linearly related. Consequently, both measures give a similar ‘resolution’ to a change of load magnitude. In the plastic region, an increase in loading will produce a large increase in strain but a small increase in stress. Thus, in this case, strain measures will provide a better ‘resolution’.
Rolling contact fatigue of railway wheels

3. Rolling contact fatigue analysis

The analysis of rolling contact fatigue differs from the ‘classic’ fatigue analysis in several aspects:

· The loading causes a multiaxial state of stress in a fixed material point where the principal stress directions rotate during the load cycle. Also, the stress components are out-of-phase, meaning that they do not reach their maxima at the same instant of time. This issue was discussed in section 2.1.

· Cracks subjected to multiaxial loading normally deviate into MODE I growth or follow a weak path in the structure, see [22]. This is not the case here, due to the large confining pressures.

stress intensity factor, \( \Delta K \), during a load cycle can be employed, eg as in Paris’ law,

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

(9)

Here, \( \frac{da}{dN} \) is the crack growth per load cycle whereas \( C \) and \( m \) are material parameters.

For more information, see texts on fatigue, eg [6], [7] and [8], or on fracture mechanics, eg [19] and [20].

Elastoplastic fracture mechanics (EPFM)

Stress intensity factors are only capable of quantifying the state of stress in the vicinity of a crack tip as long as the plastic zone at the crack tip is small enough compared to characteristic dimensions of the crack. When this is no longer the case, other measures that uniquely characterize the local state of stress at the crack tip have to be adopted.

One such measure, based on energy considerations, is the \( J \)-integral. This measure also has its limitations since it is derived under the assumption of small strains. However, it relaxes the size requirements on the plastic zone for the validity of the pertinent formulae, see [20].

When adopting the \( J \)-integral in fatigue analysis, some additional precautions should be taken. Firstly, the \( J \)-integral is an energy measure, meaning that it will always be positive. This may lead to problems in defining stress cycles, cf [21]. Secondly, the \( J \)-integral was originally intended for non-linear elastic behaviour. Care must therefore be exercised if, for instance, elastoplastic loading followed by elastic unloading needs to be studied, see [19].

The fatigue impact can also be assumed to relate to the relative displacement of material points across the crack and close to the crack tip. This measure is called crack tip displacement (CTD). One drawback with such an approach is that the CTD may be difficult to define in an unambiguous manner, see [7] and [19].
under the contact patch. Instead cracks propagate mainly in a mixed MODE II – MODE III. This type of propagation is normally not seen in cases other than rolling contact and there is a lack of universal criteria to predict the crack growth direction. There are, however, several theories on the subject, see [23] for a brief overview. Further, MODE II propagation is complicated to simulate in physical experiments, making it difficult to verify and calibrate numerical models, see [24].

- Due to the compressive loading, friction between the two opposite crack faces will play a vital role in determining the rate of crack propagation under rolling contact, cf [25]. The magnitude of the crack friction is, however, difficult to quantify.

- So far, the overwhelming majority of the fatigue research studies have been made on, and most of the predictive models have been developed for, tensile loading. In a predominantly compressive loading, the validity of such models may be questioned. For instance, it can be noted that traditional crack growth prediction methods, such as Paris’ law, predict zero crack growth under compressive loading.

- The crack face deflection will increase with crack length. It is therefore likely that there exists an ‘effective crack length’ above which a further increase in crack length will have no influence on the state of stress at the crack tip (as quantified, for instance, by a stress intensity factor). This is so because the large crack face deflections will lead to a complete locking between the crack faces for part of the crack. The issue is further discussed in papers A and B.

- In rolling contact, occasional overloads may accelerate crack growth, in contradiction to the normal behaviour in tensile loading, see [26]. This may lead to non-conservative fatigue life predictions.

- The material defects that must be considered in fatigue analysis of railway wheels are small. Further, the load magnitudes that have to be considered are high. Under such circumstances, plastic deformations in the vicinity of the defects cannot be disregarded. Consequently, models based on the theory of elasticity are generally not valid in their traditional form. This subject is further treated in section 5 and in paper A.

- Finally, railway wheels are subjected to loads with stochastically distributed magnitudes, see paper F for a further discussion on the subject.

### 3.1 Fatigue of rolling bearings

Rolling contact fatigue has been studied for a long time in the rolling bearing industry and many models for fatigue life prediction have been put forward, see [27] for an overview. In many aspects, fatigue of rolling bearings and of railway wheels are similar. However, in other aspects they differ. In order to elucidate the subject, the subsequent section provides a brief description of one criterion, the SKF criterion [28].
The SKF criterion sets out from the probability of survival, \( \Delta S_i \), of a volume element, \( \Delta V_i \), which is expressed as

\[
\ln \frac{1}{\Delta S_i} = F(N, \sigma_i - \sigma_{ai}) \Delta V_i
\]

(10)

With the nomenclature used in this thesis, \( \sigma_i \) is the equivalent stress in the material point considered and \( \sigma_{ai} \) is the equivalent fatigue limit in the same material point. In the subsequent derivation of the criterion, the shape of the function \( F \) is postulated. The probability of survival of each volume element is then weighted with a ‘stress-weighted average depth’. Finally, the probability of survival for the entire component is found through integration over the stressed volume.

The question of finding an appropriate fatigue criterion is left open in equation (10). Further, in order to take parameters such as track geometry, vehicle dynamics, wheel material etc into account, their influence on stress magnitudes or fatigue limits has to be evaluated. In rolling bearing applications, this can be done through extensive laboratory testing. In railway applications, this is not a feasible approach.

As mentioned above, this thesis deals with estimating the above influences using theoretical and numerical analyses. If this task is to be successful, could then a model similar to the SKF model be employed to derive the probability of fatigue failure for a railway wheel? In theory, the answer is yes, but there are some complicating issues in the case of railway wheels, as compared to rolling bearings:

Firstly, equation (10) is only valid for subsurface initiated fatigue. In railway wheels, fatigue may also initiate at the surface, see papers A and B. Secondly, wear is not an important issue in rolling bearing applications, whereas for railway wheels it may completely alter the state of stress in the wheel tread. A third issue is the effect of material defects, as discussed in section 5. The maximum size (and scatter) of defects is larger in railway applications. Finally, rolling bearings normally operate in a more controlled environment than a railway wheel, making fatigue design a more manageable task.

This is part of the reason behind the approach to statistical design of railway wheels proposed in paper F.

4 WLIFE – a model for fatigue analysis of railway wheel treads

This section describes the basic model for fatigue analysis of railway wheels put forward in papers C and E. This model is, in some sense, the backbone of the present work since the other papers of the thesis mainly discuss extensions and implications of the results obtainable and predictable by use of the basic model. This section is intended as a description of the different ‘modules’ of the model.
4.1 Loads

The loads acting on a railway wheel is of a stochastic nature. The magnitudes of wheel/rail forces in the load spectrum can be estimated either by numerical modelling or by in-field measurements using specially equipped wheels, see eg [29].

Approach adopted in WLIFE

Introduction of a load spectrum into the fatigue analysis model can be performed in various ways. In paper C an approach was proposed with the load spectrum discretized into a set of fixed load magnitudes where each one was assigned a probability. The fatigue impact was then calculated for each load level and the different impacts accumulated using Palmgren–Miner’s rule, see section 4.4.

In paper F, a different approach was employed. Here, the statistical distribution of the load spectra was estimated and the corresponding distribution of fatigue impact evaluated. This is further described in section 7.

4.2 Contact stresses

The loads acting on a wheel give rise to a stress field in the wheel tread. Evaluation of this stress field is not a straightforward task since the contact geometry between the wheel and the rail is, initially, unknown. The analysis can be divided into two parts, viz first an evaluation of the contact stresses between the wheel and rail, and then an evaluation of the stress field in the wheel tread resulting from the contact stresses applied.

Approach adopted in WLIFE

Under the assumption of elastic materials and simple geometries, the contact stress distribution can be analyzed using Hertzian theory [30]. This theory presumes a contact with

- elastic material behaviour
- continuous surface geometries up to the second derivative
- non-conforming surfaces resulting in a small contact area compared to body dimensions and relative radii of surface curvature
- no friction
- smooth material surfaces on the microscale

Here, the spatial distribution of the contact stresses will be elliptical with magnitudes given by analytical expressions. Further information can be found in [31].

As for the frictional stresses, these have, here, been taken as proportional to the normal stresses, i.e. \( \sigma_t^h = \mu_t \sigma^v \) and \( \sigma_t^h = \mu_t \sigma^h \) in the transverse wheel axis direction and the longitudinal rail direction, respectively. Here, \( \sigma^v \) is the vertical and \( \sigma^h \) the horizontal stress in the same point of contact. Further, \( \mu_t \) and \( \mu_t \) are adhesional

[4] The expressions will contain elliptic integrals, a fact which normally calls for a numerical evaluation.
coefficients in the transverse and longitudinal direction, respectively. Such a contact stress distribution corresponds to full slip of the wheel.

**Other approaches**
In cases where the contact geometry is not simple, there will be no simple analytical expression for the contact stress distribution. Instead, this can (still presuming elastic conditions) be found in an iterative manner. More information on such approaches is given in [32].

If elastic conditions are presumed, unrealistically high local contact stress magnitudes may be found. This calls for a ‘smoothing’ of the results in some way. Another, perhaps more physically sound, approach is to perform an elastoplastic analysis of the contact stresses. Here, numerical models such as the finite element (FE) method have to be employed. The contact stress distribution is then derived from the analysis of the stress distribution of the coupled wheel/rail system. Such an analysis must account for the non-penetration of the contacting material. This calls for the introduction of special numerical procedures such as those involving Lagrangian multipliers or penalty functions. More information can be found in textbooks on finite element modelling such as [33].

### 4.3 Stresses in the wheel rim

**Approach adopted in WLIFE**
Once the contact stress distribution is known, the subsurface stress field in the wheel tread can be calculated. If elastic conditions are presumed, there are analytical expressions available for the influence of a concentrated force acting in the normal or tangential direction on a half space. In WLIFE, these expressions are employed in that the contact stress distribution over the contact patch is replaced by a number of ‘point forces’. The total response is then evaluated by integrating the responses from all of these.

**Other approaches**
A more refined elastoplastic analysis calls for the use of approximate methods such as FE-simulations. In cases where contact stresses have been found from FE-simulations, the full problem, including both contact and subsurface stresses, is normally solved in the same process.

There are also approaches where the contact stresses evaluated using the Hertzian theory are applied to an elastoplastic wheel model. In [34] an example of such an approach for fatigue analysis of rails is presented.

### 4.4 Fatigue analysis

**Approach adopted in WLIFE**
The state of stress in the wheel tread is, as mentioned in section 3, multiaxial with out-of-phase stress components and rotating principal directions.
In order to quantify the fatigue impact, the Dang Van criterion is employed. Equation (1) and the procedure outlined in section 2.1 are used to calculate equivalent stresses, $\sigma_{EQ, DV}(t)$, for a set of selected shear planes through the potentially dangerous material points of the wheel rim to be analyzed. For each material point, the shear plane which is found to have the largest magnitude of $\max_t \sigma_{EQ, DV}(t)$ is taken as representative for the fatigue impact in this point.

When the equivalent stress in a material point exceeds a threshold value, it is postulated that fatigue damage will occur. The amount of fatigue damage that will be induced is, as mentioned in section 2.1, not quantified by the Dang Van criterion. In order to make an estimation of this amount, an approach setting out from the Wöhler curve is proposed and employed in papers C, D and E. It is then assumed that the equivalent fatigue limit, $\sigma_e$, corresponds to $10^6$ stress cycles and that an equivalent stress level corresponding to the ultimate (equivalent) strength, $\sigma_u$, corresponds to $10^1$ stress cycles. By assuming linearity on a logarithmic scale between these two extremes, the fatigue damage per cycle corresponding to an equivalent stress level $\sigma_{EQ}$ can be expressed as

$$D = \frac{5(\sigma_e - \sigma_{EQ})}{\sigma_e - \sigma_u} - 6$$  (11)

Naturally, this is a rather crude approach since the Dang Van equivalent stress only reflects the fatigue behaviour of the material for load magnitudes close to the fatigue limit. The approach will, however, reflect the exponential growth of the fatigue damage with stress magnitude.

Other approaches
There is a variety of other equivalent stress approaches that can be adopted to analyze the fatigue impact in the high-cycle regime. A review and comparison of some of these is given in, for example, (14) and (35).

In the regime of low-cycle fatigue, macroscopic plastic deformations will occur and equivalent strain approaches will offer a more viable option, see section 2.1. Also in this case, there exists a wide variety of proposed criteria in the literature. A review of some of these can be found in (35).

A fundamentally different approach is to study the growth of a fatigue crack using models based on fracture mechanics. This is a physically appealing approach which, however, introduces some intricate complications as discussed in sections 3 and 9. Examples of LEFM approaches include (36) and (37). Here, MODE I, II and III stress intensity factors for a surface crack are derived and combined into an equivalent stress intensity factor which is introduced in modified versions of Paris’ law. Neither of the models takes crack friction into consideration.
4.5 Summary and discussion

The rolling contact fatigue model WLIFE includes some restricting assumptions which, at first glance, may seem to be serious. Below is a summary of some of the critical assumptions and the errors they will induce.

Elastic material

The model presumes elastic conditions both in the stress analysis and in the fatigue analysis. Even if (global) elastic shakedown occurs in the wheel, see section 2.1, the fatigue impact will be affected by the plastic deformations that have taken place prior to shakedown. Their influence will be governed mainly by three mechanisms:

- residual stresses
  Plastic deformations will induce residual stresses in the wheel tread. Since it is postulated in the Dang Van criterion that a static shear stress will have no effect, a residual stress will not influence the first term in equation (1). Consequently, the influence is confined to a change of magnitude, \( a_{\text{DV}} \sigma_{h, \text{res}} \), in the equivalent stress. Here \( \sigma_{h, \text{res}} \) is the hydrostatic part of the induced residual stress. Since the residual stresses close to the wheel tread are compressive, they will tend to decrease the high-cycle fatigue damage. Note that, in the low-cycle fatigue regime, a high compressive residual stress may lead to increased global plastic yielding and thus would be detrimental.\(^5\)

- stress redistributions
  When a part of the wheel tread is plastically deformed, its ability to sustain further loading is reduced. Consequently, a further load increase will mainly affect surrounding zones. Thus, neglecting plastic deformations will overestimate the stress levels in the mostly stressed zones and underestimate the stress levels in some more moderately stressed regions. Thus, in reality, fatigue damage magnitudes are lower but the fatigue damage zones are more widespread than what would be predicted by an elastic model.

- initially induced fatigue damage
  During initial plastic deformation, fatigue damage will be induced in the deformed material. The model WLIFE does not account for this damage in a proper manner. In order to do so, a fatigue analysis based on a low-cycle fatigue criterion (see section 2.1) has to be employed. However, since most such criteria are based on a measure of the shear strain range, the magnitude of this effect should be somewhat reflected by the Dang Van equivalent stress since it is based on the shear stress ‘amplitude’.

Naturally, in cases where gross plastic deformations occur, the elastic approach adopted in WLIFE is not suitable.

Homogeneous and isotropic material

This is a complicated issue that may be divided into two parts:

\(^5\) However, a high compressive hydrostatic residual stress would have no effect on the plastic yielding.
influence on the global stress distribution
In the work so far, it has been assumed that the error in neglecting this effect is small.

influence on fatigue endurance
This issue is further discussed in section 5 and section 6.

Hertzian contact theory
The topics of frictionless contact and elastic material have been discussed in section 4.2 and section 4.5 respectively.

As for boundary effects, these will have an influence when the contact occurs at the wheel flange or close to the wheel side. The latter is, for example, manifested in practice by the occurrence of ‘spread rim failures’ as described in paper A. In order to quantify boundary effects properly, elastoplastic FE-analyses would be needed.

The influence of local disturbances of the contact geometry is very intricate. Small asperities in the contact zone will lead to locally increased stress magnitudes. However, due to the very high contact pressure, the asperities will also be severely deformed. An analysis of these effects in a proper manner would probably require very sophisticated material models and extensive computational efforts. A more manageable problem is the contact stress distribution for non-continuous geometries. This can be solved by iterative procedures as described in section 4.2.

4.6 Results
The model Wlife has been used in order to investigate the influence of some parameters on the fatigue impact. In summary, it was found that increased vertical and lateral load magnitudes, decreased wheel and rail radii and tensile residual stresses decreased the fatigue life. A more extensive and quantitative evaluation is presented and discussed in paper E.

5 Material imperfections
In papers E and F, the fatigue life model Wlife was employed to predict the fatigue impact for a non-worn, defect-free railway wheel. However, the predictions of fatigue life were found to be too high to correlate well with in-field data, see paper D. Consequently, other causes of fatigue failure had to be searched for among factors that either increase the stress magnitudes in the wheel tread or decrease the material resistance. As for increased stress magnitudes, the main causes are to be found in a poor dynamic behaviour (misaligned bogies, non-round wheels, track irregularities etc) and/or a poor contact geometry (due to wear of the wheel, as discussed in section 4.2). Decreased fatigue resistance of the wheel material, on the other hand, is essentially caused by material imperfections.
In papers C, D, E and F, the influence of material imperfections is neglected. This is, however, not equivalent to assuming a defect-free material (no engineering materials are). Rather, it reflects an assumption that the material defects influence neither the global stress distribution, nor the fatigue resistance (as compared to the experimentally found fatigue strength of the material).

In reality, there exists a variety of different types of material imperfections in railway wheels. If these defects are large enough, they will decrease the fatigue resistance of the wheel. The amount of decrease will be dependent on the type and size of the defects and on their. A review of some different approaches to estimating the influence of material defects on the fatigue strength can be found in paper A and in [38]. In the following, a brief discussion on the effect of a small material defect will be given, followed by a discussion on the predictive criterion employed in papers A and B.

5.1 Preliminaries – fatigue limits and crack growth arrests

If we consider a material imperfection as a small crack, its influence on the fatigue resistance in uniaxial loading can be graphically represented by a so-called Kitagawa–Takahashi diagram, see figure 5. Here, the solid line shows the threshold stress magnitude for fatigue failure of the component. It is seen that if a material defect is small enough, it will have no influence on the fatigue resistance (with respect to experimentally found fatigue limits for unnotched and polished test samples). Further, for geometrically long defects, there exists a threshold stress intensity range, \( \Delta K_{th} \), which is independent of the defect size. Combinations of stress amplitudes and defect sizes between the dashed and the solid line will initially make the cracks propagate. They will, however, come to an arrest at crack sizes corresponding to the intersection with the solid line.

For railway wheels, the case is complicated by the compressive and multiaxial loading. It is not certain, or even likely, that crack arrestment mechanisms act in the same manner as for uniaxial tensile loading. Further, the load is stochastic. This implies that a crack that has come to an arrest may start to grow again under the influence of a subsequent peak load. Moreover, such overloads may be detrimental to fatigue resistance, see section 3. Consequently, fatigue cracks that would have caused no failures under constant amplitude loading might do so under the influence of overloads. The issue of fatigue crack growth arrest is further discussed in paper A and in textbooks on fatigue, such as [7].

\[ a \] 6. A stress intensity independent of the defect size (or rather the crack length, \( a \)) is equivalent to a stress amplitude proportional to \( \sqrt{a} \) for uniaxial loading. Hence, the slope 1/2 in the logarithmic diagram in figure 5.
5.2 Effects of type, shape and size of defects

Based on studies in the literature, e.g. [39] and [40], it seems as if pores are one of the most dangerous types of material imperfections in railway wheels. The shape of the defect will, according to the theory of elasticity, have a major influence. However, it is not certain that this influence is equally severe in reality for the loaded component, since a defect will then be embedded in a smoothly shaped volume of plastically deformed material.

The defect size has no influence according to the theory of elasticity. In reality, however, the size of the defect will have an effect and in order to estimate its magnitude, a model based on a criterion originally proposed by Murakami, see [38] and [41], has been adopted in papers A and B. In the derivation of the original Murakami criterion, fatigue tests in rotating bending and in tension–compression of a number of materials were carried out. The threshold stress intensity range (in [MNm$^{-3/2}$]) for a surface crack was then empirically found to be

$$\Delta K_{th} = 3.3 \times 10^{-3} (H_v + 120) (A_{area})^{1/3}$$

Here, $H_v$ is the Vickers hardness and $A_{area}$ (in [$\mu$m]) is the square root of the defect area as projected onto the plane perpendicular to the maximum tensile stress.

For these surface cracks, it was also found (through analytical methods) that the maximum value of the mode I stress intensity factor along the crack front is approximately

7. *I.e.* according to the concept of elastic stress concentration and as long as boundary effects are not important. According to linear elastic fracture mechanics, the crack size will (if boundary effects are disregarded) have an influence proportional to $\sqrt{a}$, where $a$ is the crack length.
Here, \( \sigma_0 \) (in MPa) is the maximum principal stress. Assuming that a defect does not close during a compressive stage of the loading, we can estimate the fatigue limit by setting

\[
\Delta K_{th} = (K_{l, max} - K_{l, min}) = 2K_{l, max}
\]

Combining equations (12), (13) and (14) yields the fatigue limit (meaning cracks forming at the defect will not propagate to global fracture) as

\[
\sigma_f = 1.43 \frac{(H_v + 120)}{(\sqrt[6]{\text{area}})}
\]

For an interior crack subjected to uniaxial loading, the maximum value of the stress intensity factor along the crack front is approximately

\[
K_{l, max} \equiv 0.50 \sigma_0 (\pi \sqrt{\text{area}})^{1/2}
\]

This gives a relationship, with respect to an equal value of \( K_{l, max} \), between a surface and an interior crack as

\[
\sqrt{\text{area}}_i = 1.69 \sqrt{\text{area}}_s
\]

Finally, introducing equation (17) into equation (15) yields the fatigue limit for a specimen containing an interior defect as

\[
\sigma_f \equiv 1.56 \frac{(H_v + 120)}{(\sqrt[6]{\text{area}})}
\]

Equation (18) was modified and employed in paper A to estimate a reduced fatigue limit for a railway wheel containing a defect as

\[
\frac{\sigma_{w}}{\sigma_{e}} = \left( \frac{d}{d_0} \right)^{-1/6}
\]

Here, \( \sigma_{w} \) is the fatigue limit of the material containing a spherical defect of diameter \( d \). Further, \( \sigma_{e} \) is the fatigue limit in the absence of large defects and \( d_0 \) the diameter of a spherical defect (pore) corresponding to fatigue initiation at the stress level \( \sigma_{e} \). Reasons for, and drawbacks of, the criterion (19) are discussed in papers A and B.

### 6 Anisotropy

To support the fatigue design of railway wheels, test samples are today taken from specified positions in the wheel rim. The material data obtained from the testing of these samples are then considered as representative of the entire wheel.
Rolling contact fatigue of railway wheels

Paper B accounts for a material testing of wheels which shows that their material is anisotropic and that the properties vary both with position in the wheel and with the direction of testing.

A model for anisotropy-aware fatigue design is proposed in paper B. This model has been implemented in the computer code WILFE.

7 Statistical effects

Railway wheels are subjected to stochastic loads following a certain statistical distribution. Knowing this distribution, could it be possible to predict the statistical distribution of the resulting fatigue impact?

Paper F demonstrates an approach to this problem. The prediction is made using Monte Carlo simulation. In order to keep the computational demands within reasonable limits, a neural network model is trained and employed to mimic the numerical fatigue model, WILFE, as described in section 4.

However, the statistical influence is not limited to a distribution of load magnitudes. In a thorough design, statistical properties of contact geometries, their points of application on the wheel and fatigue strength of the material should also be accounted for. Naturally, both the complexity of determining individual statistical distributions of these properties (not to mention their cross-correlations) and the statistical simulation itself will grow with the number of parameters considered.

8 Operational data

All predictive models should be verified against operational or experimental data. In the case of railway wheels, this turns out to be fairly complicated. As for experimental data, it is hardly feasible to perform reliable full-scale tests, and scaled tests will always introduce scale effects. Such tests may, nevertheless, still provide important information.

Another approach would be to study failures that have occurred with operational wheels in order to verify trends. Such an attempt is presented in paper D. The task proved to be more difficult than expected. Apart from the fact that the statistics studied contained a lot of statistical noise and erroneous information, the sole task of determining which data that are relevant and quantify these is far from trivial. An example:

- Clearly it is not only the magnitude of loading but also the transversal position of the point of application of the load on the wheel tread that influences the fatigue impact. This point will be determined by the local contact geometry of the rail and wheel,
the dynamic behaviour of the bogie etc. Even more complicated is the task of correlating transversal contact positions on the wheel tread with the corresponding (vertical, transversal and radial) load magnitudes.

The issue is further discussed in section 7 and in paper A.

9 Future work

The most appealing approach to fatigue design of a railway wheel, from a physical point of view, would perhaps be the following:

- An initial crack (with initial size, shape, orientation and location) is assumed
- A critical size of this crack, corresponding to final fracture, is determined by means of fracture mechanics
- The time of propagation of the crack, during regular train operations, from initial to critical size is determined through fracture mechanics
- Inspection intervals for the wheels are defined in order to prevent the crack from growing to fracture of the wheel

Such fatigue design principles are employed eg in the aerospace industry. Unfortunately, there are several complicating factors and unsolved issues in adopting the approach to railway wheels:

- Given an initial crack size, it is not trivial how the most dangerous shape, orientation and location of this crack should be found
- Determination of critical crack size is complicated. It even seems as if final fracture is fairly uncorrelated to the crack size, see papers A and B. This may be due to the existence of an ‘effective crack length’ as discussed in section 3
- Analysis of crack growth rates is complicated due to crack friction etc, see section 3

In order to overcome these problems, there is a need for substantial research efforts.

Another important research area concerns the influence of assumed contact conditions: How much error will the assumption of Hertzian contact conditions impose? How much influence will a change in local contact geometry due to wear have on fatigue impact? This issue is being currently studied in a parallel project.

In the author’s opinion, the most important issue concerns, however, the influence of defects. As mentioned, the theoretical and experimental bases for equation (19) are derived under the assumption of uniaxial tensile loading. If an extrapolation to the conditions of rolling contact of railway wheels is possible, and which demands for adjustments this would then raise, remains yet to be clarified. As an example, the defect size in equation (18) is defined
as the projection onto the plane perpendicular to the largest principal stress. Such a definition is clearly not appropriate in the case of rolling contact fatigue of railway wheels.

Currently, research is being carried out in this area. With a knowledge of the influence of defects on fatigue resistance, the fatigue initiation model, as described in papers C, D and E, together with a stochastic design model as described in paper F, would form a most powerful design tool as follows (based on a description in paper B):

1. The fatigue limit of the ‘defect-free’ material, $\sigma_e$, is experimentally evaluated.
2. A design fatigue limit, $\sigma_r$, reduced due to the influence of defects, is evaluated using (a modified version of) equation (19). The reduction should be made with respect to the largest expected defect size. In a thorough analysis, this can be considered a stochastic variable, resulting in a stochastic fatigue limit.
3. A force spectrum is derived from a dynamic wheel/rail analysis and/or from in-field measurements on wheels operating under statistically similar conditions.
4. A fatigue analysis is now carried out using the fatigue initiation model as described in papers C, D, E and F (and, in cases where anisotropy is to be taken into account, the appendix of paper B). This analysis will give the probability of fatigue initiation under idealized conditions with respect to wheel/rail contact geometry, track roughness etc.

It should be noted that the influence of hunting (ie, transversal movements of the wheel on the rail) has not been considered in paper F. Such motions should normally decrease the fatigue impact (as long as the contact zone is confined to the running track on the tread). If needed, this effect can fairly easily be included in the analysis.

There are, however, some other implications of non-idealized conditions that are more intricate to predict. Let us consider two examples:

- A hollow wheel profile may, locally, lead to very high contact stresses. It will also affect the vehicle dynamics, which will induce further changes in contact conditions at the wheel/rail interface. These may, in turn, cause an increased wear, further affecting the vehicle dynamics etc. Thus, a thorough analysis of the influence of hollow wear becomes very complicated and has to take several interacting mechanisms into account. Consequently, the task of specifying maintenance tolerances for hollow wear of railway wheels is today mainly based on empirical knowledge, cf [3].
- The fatigue impact of a track irregularity, such as a weld or joint, is more easily analyzed. If it can be assumed that the irregularity does not influence the overall vehicle dynamics, the influence will be limited to an induced impact force during the wheel passage. The pertinent fatigue damage is derivable from fatigue analysis, eg using wlife. Further, if the distribution of track
irregularities is known (as to frequency of occurrence, magnitude of induced forces etc), the total fatigue damage during an operational life of the wheel can be estimated.

10 Brief review of papers contributed

**PAPER A**  Effects of imperfections on fatigue initiation in railway wheels

Origins of wheel failures are discussed. Three different types of failure are investigated and their causes are highlighted. The issue of how to account for defects in a fatigue design is treated in some depth. Strategies for avoiding fatigue failures of railway wheels are discussed.

**PAPER B**  Anisotropy and fatigue of railway wheels

The strength of railway wheels is discussed from a more fundamental perspective. The issue of material anisotropy is treated. Physical tests are carried out to study the material strength. An improved wheel material is proposed and tested. A defect-aware and anisotropy-aware fatigue design procedure is outlined.

**PAPER C**  A fatigue life model for general rolling contact with application to wheel/rail damage

In papers C, D and E, the rolling contact fatigue model Wライフ (see section 4 above) is presented. Paper C contains the overall description. In this paper, however, only a simplified approach to calculating the shear stress 'amplitude' is employed. This approach is valid in the special case of pure rolling contact, making the title of the paper somewhat misleading.

**PAPER D**  Rolling contact fatigue of railway wheels – computer modelling and in-field data

Fatigue damage magnitudes estimated by the model described in paper C seemed to be very small. Further, this model predicts no influence from the hydrostatic pressures existing in the wheel rim during rolling*. This is the background of the modified fatigue criterion employed in paper D. The modified criterion utilizes the full magnitude of the Tresca shear stress, $\tau_{\text{Tresca}}(t)$, rather than the 'amplitude', $\tau_{a,\text{Tresca}}(t)$. This does, however, violate the empirical fact that a superposed static shear stress will have no influence on the fatigue resistance in terms of the fatigue limit. Thus, the criterion should only be applicable for repeated compressive loading.

In a more general form, the criterion was expressed as

$$\sigma_{\text{EQ}}(t) = (1 - \gamma)\tau_{a,\text{Tresca}}(t) + \gamma\tau_{\text{Tresca}} + a\sigma_a(t) > \sigma_e \quad (20)$$

* The model would, however, predict a beneficial influence from the hydrostatic pressure due to residual stresses.
Here, $\alpha$ is a material parameter governing the influence of the hydrostatic stress and $\sigma_e$ the equivalent fatigue limit. The criterion states that fatigue will be initiated if the inequality is fulfilled. A choice of $\gamma = 0$ would now yield equation (1). In paper D, the value $\gamma = 1$ was chosen to represent the other extreme when the full Tresca shear stress is employed.

The present approach is something of a ‘shot from the hip’. However, it turned out that the criterion did estimate parametric influences well in coincidence with the much more sophisticated model later presented in paper E. This fact justifies to some extent the fairly crude fatigue design criteria that are normally used for railway wheels today. Further, fatigue lives predicted by the modified criterion coincided better with measured fatigue lives than those by the original criterion. The reason for this could perhaps partly be attributed to a tendency of the Dang Van criterion to underestimate fatigue impact in compressive loading (as indicated in Table 1 of paper D). However, the major reason seems to be the influence of material defects as described in section 5.

Paper D also contains a brief analysis of in-field statistics of wheel reprofilings and a comparison of these to predicted magnitudes (Figure 4 and Table 2, respectively). This is the only comparison in the thesis between predicted and operational fatigue lives. This lack of verification of the numerical model originates in the difficulties in obtaining reliable and sufficient data from in-field operations as discussed in section 8 and in papers A and B.

**Paper E** Rolling contact fatigue of railway wheels — a parametric study

In paper E, $WLIFE$ is finalized with a proper treatment of the shear stress as described in section 2.1 above. A number of parametric studies are presented in the paper, which also discusses under which circumstances simplified criteria as described in papers C and D can be justified.

**Paper F** Multiaxial fatigue — a probabilistic analysis of initiation in cases of defined stress cycles

Paper F presents a model for numerical derivation of statistical distributions of fatigue impact, given a statistical load distribution. This is performed using neural network together with Monte Carlo simulations.

It should be noted that the estimated fatigue limit employed in paper F later proved to be an underestimation, see paper B. Thus, the observed fatigue of railway wheels seems to require either extreme stress magnitudes or a lowered material strength, as discussed in section 5 above.
Appended papers


B. ANDERS EKBERG & PETER SOTKOVSKII, Anisotropy and fatigue of railway wheels, submitted for international publication


E. ANDERS EKBERG, Rolling contact fatigue of railway wheels – a parametric study, Wear, vol 211, pp 280–288, 1997

F. ANDERS EKBERG, REINE LINDOVIST & MARTIN OLOFSSON, Multiaxial fatigue – a probabilistic analysis of initiation in cases of defined stress cycles, Fatigue ’99 – Proc 7th International Fatigue Congress, Beijing, China, pp 923–928, 1999

References

1. MAGEL E & KALOUSEK J, Identifying and interpreting railway wheel defects, Proc International Heavy Haul Conference – “Running heavy, running fast into the 21st century”, Montreal, Quebec, pp 5.7–5.21, 1996


15. LINDQVIST R, On high cycle fatigue analysis of as-cast nodular cast iron components, Thesis for the Degree of Licentiate of Engineering, Department of Naval Architecture and Ocean Engineering, Chalmers University of Technology, 1999


20. NILSSON F, Fracture mechanics – from theory to applications, Department of Solid Mechanics, Royal Institute of Technology, Stockholm, 159 pp, 1997


REFERENCES

35. Socie D F & Marquis G B, Multiaxial fatigue, Society of Automotive Engineers Inc, Warrendale PA, USA, 484 pp, 2000