

## Finite Element Solution of Rolling Contact and Analysis of Squats Formation

Deng, Xiangyun

**DOI**

[10.4233/uuid:670197d8-9e35-4349-9459-8366c5ef5902](https://doi.org/10.4233/uuid:670197d8-9e35-4349-9459-8366c5ef5902)

**Publication date**

2019

**Document Version**

Final published version

**Citation (APA)**

Deng, X. (2019). *Finite Element Solution of Rolling Contact and Analysis of Squats Formation*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:670197d8-9e35-4349-9459-8366c5ef5902>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Finite Element Solution of Rolling Contact and Analysis of Squats Formation



Xiangyun DENG

# **Finite Element Solution of Rolling Contact and Analysis of Squats Formation**

**Xiangyun DENG**



# **Finite Element Solution of Rolling Contact and Analysis of Squats Formation**

## **Dissertation**

for the purpose of obtaining the degree of doctor  
at Delft University of Technology  
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen  
chair of the Board for Doctorates  
to be defended publicly on  
Wednesday 10 July 2019 at 15:00 o'clock

by

**Xiangyun DENG**

Master of engineering mechanics,  
University of Science and Technology of China, Hefei, China  
Born in Xuanwei, Yunnan, China

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,  
Prof.dr.ir. R.P.B.J. Dollevoet  
Prof.dr.ir. Z. Li

Chairperson  
Delft University of Technology, promotor  
Delft University of Technology, promotor

Independent members:

Prof.dr.ir. E. van der Heide  
Prof.dr. M. Veljkovic  
Prof.dr.ir. J. Sietsma  
Dr. Yu Zhou  
Dr. A.A. Nunez Vicencio

University of Twente  
Delft University of Technology  
Delft University of Technology  
Tongji University  
Delft University of Technology

Reserve member:

Prof.dr.ir. J.G. Rots

Delft University of Technology

This dissertation was financially supported by:



ISBN: 978-94-6366-187-4

Printed by: Gildeprint – Enschede

Copyright © 2019 by Xiangyun Deng (xyd2010@gmail.com)

An electronic version of this dissertation is available at

<http://repository.tudelft.nl>

# SUMMARY

Squats are a form of surface initiated rolling contact fatigue (RCF) damage in rails. They are associated with crack initiation and propagation and local geometry deviation of rail top. Squats not only may lead to rail failure but also accelerate track deterioration, resulting in different abrupt and sometimes catastrophic consequences. In spite of previous contributions on the study of squats, a number of gaps still exist in the literature. The development process and also the dynamic wheel-rail contact mechanics and root causes behind the physical phenomena of squats are not yet fully understood. Additionally, an explicit finite element (FE) method has been used to study squats. However, accuracy and performance of this method have not been completely verified.

This thesis provides analyses of the performance of the explicit FE method for wheel-rail contact and gives new insights into the causes and development process of squats based on five-year continual field monitoring observations. One goal in this thesis is to study the accuracy of the explicit FE method for frictional rolling contact in a wider-range of conditions. The other goal is to reveal the physical phenomena of squats, such as cracking mechanisms and development process. To achieve these goals, traditional half-space-based methods are studied and employed for the verification of the explicit FE method. Additionally, the five-year field monitoring observations combined with numerical simulations are employed to obtain a better understanding of the underlying mechanisms of squats. This thesis is divided into four main parts.

The first part studies the applicability of half-space-based methods in non-conforming contact problems and their accuracy. A validated implicit FE method is compared with half-space-based methods in terms of their solutions. The static normal contact problem is considered, with a focus on the Hertz theory and the Kalker's variational theory. Different combinations of significant dimensions of the contact bodies in terms of height, width and length are investigated. Various contact patch eccentricities and contact body shapes are also considered. The results show that the half-space-based methods deliver a difference within 9% from the FE method even when the significant dimensions are as small as 1.1x the characteristic size. This indicates that the half-space-based methods yield high-accuracy solution for non-conforming contact problems.

In the second part, spin-rolling contact is analyzed using the explicit FE method for frictional rolling contact. A quasi-static solution is obtained with dynamic relaxation. Both the normal solutions and tangential solutions are analyzed and verified by comparing with those of the half-space-based methods, i.e., the Hertz theory and the Kalker's variational theory. The comparison shows that the explicit FE method is very accurate for spin-rolling contact problems. The elastic-plastic solution and the subsurface stress distribution are also investigated. It is found that plasticity has a strong effect on the size and shape of the contact

patch, as well as contact stress. Moreover, results show that spin can increase surface shear stress and micro-slip.

The third part of the thesis studies a five-year continual field monitoring on a large number of squats induced by corrugation. This field monitoring captured various stages of the life cycle of corrugation-induced squats, from small black depressions without cracks to mature two-lung shaped squats accompanied by Y-shaped and I-shaped cracks. The associated geometry evolution, crack initiation and propagation behaviors are analyzed. Accordingly, the mechanisms for crack initiation, crack propagation and geometry evolution are proposed. The proposed mechanisms are able to properly explain the observed phenomena of corrugation-induced squats.

Finally, in the fourth part of this thesis, weld-induced squats are studied based on the five-year continual field monitoring and numerical simulations. According to the observations, a hypothesis of the formation and development process of squats at welds is proposed. In this thesis, a detailed analysis is provided for the pre-cracking process. To quantitatively validate the hypothesis, a three-dimensional (3D) FE model is built to simulate the dynamic vehicle-track interaction with detailed consideration of the local wheel-rail frictional rolling contact at a thermite weld. Smooth rail surface and surface with irregularities and varying yield stress based on measured hardness are the input of the analysis. Dynamic contact forces, plastic deformation and wear are calculated. The numerical results verified the hypothesis that the squats are developed from initial V-shaped irregularities due to plastic deformation in the heat affected zones. Then the surface irregularities excite dynamic longitudinal contact force, which in turn produces W-shaped surface patterns through further differential plastic deformation. The W-shaped patterns eventually cause the formation of squats at welds.

In conclusion, this thesis extends the capability of the explicit FE method for frictional rolling contact problems. The five-year field monitoring study and the numerical simulations of dynamic wheel-rail interaction at surface defects contribute to a better understanding of the formation mechanism and development process of squats. Furthermore, the findings of this thesis provide crucial knowledge for a more scientific and cost-effective rail inspection, as well as for the design and manufacturing of squat-resistant rail materials. These will eventually lead to reductions of the maintenance costs and high improvements of railway safety.

# SAMENVATTING

Squats zijn een vorm van vermoeiingsschade door wiel-rail contact ontstaat (Rolling Contact Fatigue/RCF) aan het oppervlak van de rails. Ze worden geassocieerd met scheurinitiatie en -propagatie en lokale geometriefwijkingen van de spoorstaafkop. Squats kunnen niet alleen tot het falen van spoorstaven leiden, maar ook verslechtering van de bovenbouw versnellen, wat resulteert in verschillende abrupte en soms catastrofale gevolgen. Ondanks eerdere bijdragen aan het bestuderen van squats, bestaan er nog steeds een aantal hiaten in de literatuur. Het ontwikkelingsproces en ook de dynamische wiel-rail contactmechanica en hoofdoorzaken achter de fysieke verschijnselen van squats zijn nog niet volledig begrepen. Bovendien is een expliciete eindige elementen (FE) methode gebruikt om squats te bestuderen. De nauwkeurigheid en prestaties van deze methode zijn echter niet volledig geverifieerd.

Dit proefschrift biedt analyses van de prestaties van de expliciete FE-methode voor kortstondig wiel-railcontact en geeft nieuwe inzichten in de oorzaken en het ontwikkelingsproces van squats op basis van vijf jaar durende continue monitoring op locatie. Eén van de doelen van dit proefschrift is het bestuderen van de nauwkeurigheid van de expliciete FE-methode voor rollend contact met wrijving in een breed scala van omstandigheden. Het andere doel is om de fysieke verschijnselen rondom squats aan het licht te brengen, zoals scheurvormen en scheurgroei. Om deze doelen te bereiken worden traditionele, op halfruimte gebaseerde, methoden bestudeerd en gebruikt voor de verificatie van de expliciete FE-methode. Daarnaast wordt de vijf jaar durende monitoring op locatie gecombineerd met numerieke simulaties om een beter inzicht te krijgen in de onderliggende mechanismen van squats. Dit proefschrift is verdeeld in vier hoofdonderdelen.

Het eerste deel bestudeert de toepasbaarheid van op halfruimte gebaseerde methoden bij niet-conforme contactproblemen en hun nauwkeurigheid. Een gevalideerde impliciete FE-methode wordt vergeleken met op halfruimte gebaseerde methoden op basis van hun oplossingen. Het statische normale contactprobleem wordt beschouwd, met een focus op de Hertz-theorie en de variatietheorie van Kalker. Verschillende combinaties van relevante afmetingen van de contactlichamen worden onderzocht op het gebied van hoogte, breedte en lengte. Verschillende excentriciteiten van contactvlakken en vormen van contactlichamen worden ook beschouwd. De resultaten tonen aan dat de op halfruimte gebaseerde methoden een verschil van binnen de 9% met de FE-methode opleveren, zelfs wanneer de significante afmetingen slechts 1,1x de karakteristieke afmeting zijn. Dit geeft aan dat de op halfruimte gebaseerde methoden een zeer nauwkeurige oplossing bieden voor niet-conforme contactproblemen.

In het tweede deel wordt het spin/rol-contact geanalyseerd met behulp van de expliciete FE-methode voor rollend contact met wrijving. Een quasi-statische oplossing wordt verkregen met dynamische relaxatie. Zowel de normaal-oplossingen als de tangentiële oplossingen

worden geanalyseerd en geverifieerd door deze te vergelijken met die van de op halfruimte gebaseerde methoden, d.w.z., de Hertz-theorie en de variatietheorie van Kalker. Uit de vergelijking blijkt dat de expliciete FE-methode zeer nauwkeurig is voor spin/rol-contactproblemen. De elastisch-plastische oplossing en de spanningsverdeling onder het oppervlak worden ook onderzocht. Er is aangetoond dat plasticiteit een sterk effect heeft op de grootte en vorm van het contactvlak, evenals op contactspanning. Bovendien laten de resultaten zien dat spin de schuifspanning aan het oppervlak en micro-slip kan verhogen.

Het derde deel van het proefschrift bestudeert een vijf jaar durende monitoring op locatie van een groot aantal squats veroorzaakt door golfslijtage. Deze monitoring op locatie heeft verschillende stadia van de levenscyclus van door golfslijtage ontstane squats vastgelegd, van kleine zwarte depressies zonder scheuren tot volgroeide twee-longvormige squats met Y-vormige en I-vormige scheuren. Het bijbehorende geometrie-verloop, scheurinitiatie en scheurgroei worden geanalyseerd. Aan de hand hiervan worden de mechanismen voor scheurinitiatie, scheurgroei en geometrie-ontwikkeling voorgesteld. De voorgestelde mechanismen kunnen de waargenomen verschijnselen van door golfslijtage ontstane squats goed verklaren.

Ten slotte worden in het vierde deel van dit proefschrift door lassen veroorzaakte squats bestudeerd op basis van de vijf jaar durende monitoring op locatie en numerieke simulaties. Aan de hand van de waarnemingen wordt een hypothese van het vormings- en ontwikkelingsproces van squats bij lassen voorgesteld. In dit proefschrift wordt een gedetailleerde analyse gegeven voor het pre-scheurproces. Om de hypothese kwantitatief te valideren, is een driedimensionaal (3D) FE-model gebouwd om de dynamische voertuigspoorinteractie te simuleren met gedetailleerde beschouwing van het lokale wiel-rail wrijvingscontact bij een thermietlas. Glad railoppervlak en oppervlak met onregelmatigheden en variërende vloeispanningen op basis van gemeten hardheid zijn de input van de analyse. Dynamische contactkrachten, plastische vervorming en slijtage worden berekend. De numerieke resultaten bevestigden de hypothese dat de squats zijn gevormd uit aanvankelijke V-vormige onregelmatigheden als gevolg van plastische vervorming in de door warmte beïnvloede zones. Vervolgens wekken de onregelmatigheden in het oppervlak een dynamische longitudinale contactkracht op, die op zijn beurt W-vormige oppervlaktepatronen produceert door verdere differentiële plastische vervorming. De W-vormige patronen veroorzaken uiteindelijk de vorming van squats bij lassen.

Concluderend, dit proefschrift vergroot de mogelijkheden van de expliciete FE-methode voor problemen met rollend contact en wrijving. De vijf jaar durende veldmonitoringstudie en de numerieke simulaties van dynamische wiel-rail interactie bij oppervlaktedefecten dragen bij aan een beter begrip van het ontstaansmechanisme en het ontwikkelingsproces van squats. Bovendien bieden de bevindingen van dit proefschrift cruciale kennis voor een meer wetenschappelijke en kosteneffectieve spoorweginspectie, evenals voor het ontwerpen en produceren van scheurbestendige railmaterialen. Deze zullen uiteindelijk leiden tot verlaging van de onderhoudskosten en flinke verbetering van de spoorveiligheid.

# CONTENTS

|  |            |
|--|------------|
| <b>Summary</b>   | <b>v</b>   |
| <b>Samenvatting</b>  | <b>vii</b> |
| <b>1 Introduction</b>  | <b>1</b>   |
| 1.1. Motivation .....  | 2          |
| 1.2. Overview of rail squats.....  | 2          |
| 1.3. Consequences of squats .....  | 3          |
| 1.4. Wheel-rail rolling contact .....  | 4          |
| 1.4.1. Solution methods for wheel-rail contact .....   | 4          |
| 1.4.2. High-frequency dynamic interaction at rail defects .....                                      | 5          |
| 1.5. Geometry evolution of squats .....  | 6          |
| 1.5.1. Plastic deformation .....   | 6          |
| 1.5.2. Wear .....  | 6          |
| 1.6. RCF crack initiation and propagation .....  | 7          |
| 1.6.1. Crack initiation .....  | 7          |
| 1.6.2. Crack propagation.....  | 8          |
| 1.7. The gaps in the existing studies of rolling contact and rail squats .....                       | 8          |
| 1.7.1. Calculation of wheel-rail contact.....  | 8          |
| 1.7.2. Understanding of rail squats.....   | 9          |
| 1.8. Problem statement.....  | 9          |
| 1.9. Outline of the thesis .....   | 10         |
| References .....   | 13         |
| <b>2 Applicability of half-space-based methods to non-conforming elastic normal contact problems</b> | <b>19</b>  |
| 2.1. Introduction .....  | 20         |
| 2.2. Assessment strategy on solution methods for contact problems .....                              | 21         |
| 2.2.1. FE contact model .....  | 22         |
| 2.2.1.1. FE model description.....   | 22         |
| 2.2.1.2. Design of the significant dimensions of the lower body .....                                | 23         |

|   |           |
|---|-----------|
| 2.2.2. The Kalker's variational theory .....                              | 24        |
| 2.2.3. Assessment strategy .....  | 25        |
| 2.3. Simulation results .....   | 25        |
| 2.3.1. Normal solution when $b_0/a_0 = 0.75$ .....                        | 26        |
| 2.3.2. Influence of the contact patch eccentricity $b_0/a_0$ .....        | 28        |
| 2.3.3. Influence of the contact body shape .....                          | 28        |
| 2.4. Discussions .....  | 29        |
| 2.5. Conclusions .....  | 32        |
| References .....  | 33        |
| <b>3 Lagrangian explicit FE modeling for spin-rolling contact</b> .....   | <b>35</b> |
| 3.1. Introduction .....   | 36        |
| 3.2. Theoretical basis of the explicit FE model for rolling contact ..... | 38        |
| 3.2.1. Problem formulation of rolling contact .....                       | 38        |
| 3.2.2. Lagrangian explicit FE scheme .....                                | 39        |
| 3.2.3. Contact algorithm .....  | 39        |
| 3.3. Description of the FE model .....                                    | 40        |
| 3.3.1. Model description .....  | 40        |
| 3.3.2. Mesh .....   | 41        |
| 3.3.3. Boundary conditions .....  | 41        |
| 3.3.4. Initial equilibrium and dynamic relaxation .....                   | 41        |
| 3.3.5. Tangential loading conditions .....                                | 42        |
| 3.3.6. Material properties .....  | 42        |
| 3.3.7. Calculation of the physical quantities in the contact patch .....  | 43        |
| 3.4. Analysis of elastic spin-rolling contact and verification .....      | 44        |
| 3.4.1. Pressure distribution .....  | 44        |
| 3.4.2. Tangential solutions .....   | 44        |
| 3.4.2.1. Surface shear stress distribution .....                          | 44        |
| 3.4.2.2. Areas of slip, stick, and micro-slip .....                       | 46        |
| 3.5. Elasto-plastic solution .....  | 48        |
| 3.5.1. Solutions on the contact surface .....                             | 48        |
| 3.5.1.1. Von Mises stress of the surface layer .....                      | 48        |
| 3.5.1.2. Pressure .....   | 49        |
| 3.5.1.3. Surface shear stress and micro-slip .....                        | 50        |
| 3.5.1.4. The influence of tangential load on the solution .....           | 51        |
| 3.5.2. Stress history .....   | 52        |
| 3.6. Discussions .....  | 53        |

|  |           |
|--|-----------|
| 3.7. Conclusions.....  | 54        |
| References .....   | 56        |
| <b>4 Investigation of the formation of corrugation-induced rail squats based on extensive field monitoring</b>                         | <b>59</b> |
| 4.1. Introduction .....  | 60        |
| 4.2. Description of the field monitoring .....   | 62        |
| 4.3. Field observations and analysis .....   | 65        |
| 4.3.1. Squat initiation and cracking processes observed over time.....   | 65        |
| 4.3.1.1. Critical crack initiation size and point in the lateral direction and U-shaped cracks.....                                    | 65        |
| 4.3.1.2. U-shaped crack initiation position in the rolling direction over the complete squat development process .....                 | 67        |
| 4.3.2. Surface and subsurface cracks at squats of different severities.....  | 68        |
| 4.3.3. Cracking angle.....   | 69        |
| 4.3.4. Crack shape in 3D and secondary cracks.....   | 70        |
| 4.3.4.1. The primary U-shaped crack in 3D .....  | 70        |
| 4.3.4.2. Secondary I-shaped crack .....  | 71        |
| 4.3.4.3. Y-shaped crack and the two-lung shape of squats.....  | 72        |
| 4.3.5. Squat development process.....  | 72        |
| 4.4. Elaboration on the squat development process.....   | 73        |
| 4.4.1. Formation of small black depressions with reduced tensile strength .....  | 73        |
| 4.4.2. Crack initiation by mode I fracture .....   | 73        |
| 4.4.3. Crack propagation into the U shape.....   | 74        |
| 4.4.4. Influence of rail inclination on cracking and tertiary cracks .....   | 75        |
| 4.5. Discussion .....  | 78        |
| 4.5.1. Evolution of the squat geometry and comparison with other types of squats .....   | 78        |
| 4.5.2. Squat development after rail grinding.....  | 79        |
| 4.6. Conclusions.....  | 80        |
| References .....   | 82        |
| <b>5 Pre-cracking development of weld-induced squats due to plastic deformation: five-year field monitoring and numerical analysis</b> | <b>85</b> |
| 5.1. Introduction .....  | 86        |
| 5.2. Field monitoring observations and measurements.....   | 87        |
| 5.2.1. Squats with cracks at welds.....  | 88        |
| 5.2.1.1. Squats at thermite welds .....  | 88        |

|   |            |
|---|------------|
| 5.2.1.2. Squats at flash butt welds .....   | 92         |
| 5.2.1.3. Summary.....   | 94         |
| 5.2.2. Welds without cracking.....  | 94         |
| 5.2.3. Hardness distribution at welds.....  | 96         |
| 5.2.4. Summary of the observations and measurements.....  | 97         |
| 5.3. Hypothesis of squat formation at welds .....   | 97         |
| 5.4. Numerical model.....   | 99         |
| 5.4.1. FE model.....  | 99         |
| 5.4.2. Material properties at the weld .....  | 100        |
| 5.4.3. Surface deformation and wear .....   | 101        |
| 5.4.4. Simulation steps.....  | 101        |
| 5.5. Simulation of rail profile evolution considering deformation and wear .....                      | 102        |
| 5.5.1. Step1: Formation of V-dips from smooth rail with varying yield stress.....                     | 102        |
| 5.5.1.1. Calculated deformation and wear .....  | 102        |
| 5.5.1.2. Deriving V-dip irregularities from deformation .....   | 103        |
| 5.5.1.3. Comparing the simulation with field measurement.....   | 103        |
| 5.5.2. Step 2: Formation of W-shapes from V-dips with varying hardness .....                          | 104        |
| 5.5.2.1. Generating rail surface with V-dips.....   | 104        |
| 5.5.2.2. Formation of W-shaped patterns and comparison with field<br>measurement.....                 | 105        |
| 5.6. Discussions .....  | 107        |
| 5.6.1. W-shape is determined by the wavelength of the local dynamic system.....                       | 107        |
| 5.6.2. The presence of single or duo squats is determined by length of welds.....                     | 108        |
| 5.6.3. Occurrence or not of squats is determined by both hardness and dynamic<br>contact forces ..... | 109        |
| 5.6.4. Summary .....  | 109        |
| 5.7. Conclusions.....   | 109        |
| References .....  | 112        |
| <b>6 Conclusions, applications and recommendations</b> .....  | <b>115</b> |
| 6.1. Conclusions.....   | 116        |
| 6.2. Applications of research findings .....  | 119        |
| 6.2.1. Applications from the study on applicability of half-spaced based methods<br>.....             | 119        |
| 6.2.2. Applications of the explicit FE method for rolling contact .....                               | 119        |
| 6.2.3. Practical applications of the research findings on squat formation.....                        | 120        |
| 6.3. Recommendations on future research .....   | 120        |

---

|   |            |
|---|------------|
| 6.3.1. Extension of the explicit FE method for rolling contact..... | 120        |
| 6.3.2. Study of rail squats .....                                   | 120        |
| 6.3.3. Future study on RCF damage .....                             | 121        |
| References .....  | 122        |
| <b>Acknowledgement</b>  | <b>123</b> |
| <b>About the author</b>   | <b>125</b> |
| <b>List of publications</b>   | <b>126</b> |



# 1

## INTRODUCTION

## 1.1. Motivation

Rail squats are a main form of rolling contact fatigue (RCF) damage in rails, but their formation mechanism is still a controversial topic, among other reasons, due to the lack of sufficient field monitoring data and validated numerical investigations. In the literature, to understand the formation mechanism of squats, both stress and strain states are analyzed with an explicit finite element (FE) method. However, the accuracy of the explicit FE method on rolling contact has not been fully verified. This thesis verifies the explicit FE method for solving rolling contact problems in a wider-range of conditions, and then uses the method with a five-year continual field monitoring to study the formation mechanism of rail squats.

## 1.2. Overview of rail squats

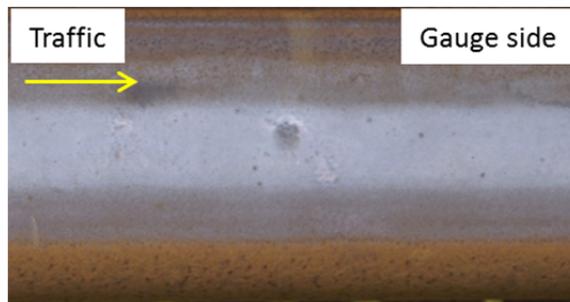
Rail squats were reported in Japan as black spots dated back to the early 1950s [1]. They became known in the UK in the 1970s [2] and were named as squats [3]. Afterward, squats were reported in some other European countries [4]. They were also recently recognized as an important rail track issue in Australia [5]. Squats are generally observed on the crown of the railhead in tangent and shallow curved tracks [6]. They are also observed in switch & crossing noses [7]. Also, squats may occur on all types of track, such as ballast and slab track, lines with passengers, freight or mixed traffic, and high speed, conventional and metro railway lines [8].

A typical mature squat, as shown in Fig. 1.1, has three characteristics [6, 9]. The first is the localized dark depression in the shape of two-lung-like, or the permanent deformation indented by someone sitting. The second is the widening of the running band, which is a result of plastic deformation because of the impact of wheels on the rail. The third is that mature squats are often associated with the V, U, Y, or circular shaped cracks.



Fig. 1.1. A typical two-lung-like squat where the dark depression, V-shaped crack and the widening of the running band are observed (From Fig. 12 in [9]).

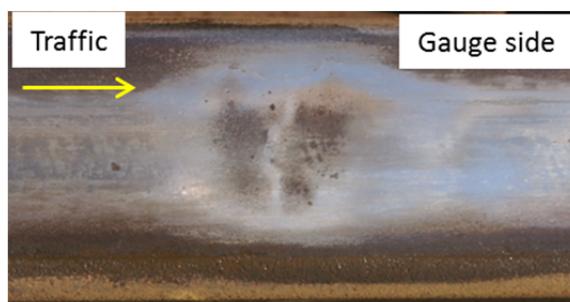
For maintenance convenience, squats are roughly classified into three categories according to their severity, i.e., class A, class B and class C, corresponding to light squat, moderate squat and severe squat, respectively [10] (see Fig. 1.2). Different maintenance measures are prescribed for different classes of squats, including grinding, welding and rail replacement [8].



(a)



(b)



(c)

Fig. 1.2. (a) Class A (the same as Fig. 4.4(a)), (b) Class B (the same as Fig. 4.4(e)) and (c) Class C (from Fig. 6 in [9]) squats.

### 1.3. Consequences of squats

Severe squats may lead to catastrophic consequences, if they are not considered in the maintenance plans. Cracks that initiated in the rail surface will grow into the rail material and could lead to rail breakage [6]. Moreover, the surface depressions excite high-frequency impact forces, which are thought to accelerate the deterioration of track and vehicle

components [5, 11]. The high-frequency impact forces also cause high-level of wheel-rail impact noise, which is a nuisance for the people living close to the railway tracks [12, 13]. Therefore, to effectively control squats and prevent their consequences, an in-depth study of squat formation mechanism is required. The study should include squat geometry evolution, crack initiation and propagation under a wide-range of wheel-rail rolling contact conditions.

## 1.4. Wheel-rail rolling contact

The knowledge about wheel-rail contact and the solution methods are essential to understand the physics/mechanics behind squat damage. Since squat formation is related to the high-frequency dynamic forces excited by the defects and local deformation due to continuous use of the railway tracks [6], two interrelated research fields are required to be considered, i.e., local contact behaviors and global dynamic vehicle-track interaction at rail defects [14]. Contact mechanics are employed to compute the local contact patches and surface stress transmitted at the wheel-rail interface, while multibody modeling is employed to assess the global dynamic vehicle-track interaction forces with consideration of vehicle and track structure components. One integrated method is required to accurately calculate the interrelation between local contact and global high-frequency dynamic vehicle-track interaction.

### 1.4.1. Solution methods for wheel-rail contact

The interface between wheel and rail is subjected to both normal and tangential loads. The normal load produces a distribution of normal surface stress or contact pressure. The tangential load is transmitted from different creepage including longitudinal, lateral and spin creepage, as well as the traction and braking of wheels [15]. The tangential load produces a surface shear stress distribution and micro-slip [16]. They are concentrated on the interface in a small contact patch. The position and shape of the contact patch and the corresponding stress distributions depend on many factors such as the profiles of wheel and rail, the curvature of track and the relative position of the wheel and rail. Due to these factors, any of the following four types of wheel-rail contact condition could take place [17]: wheel tread-rail head contact, wheel flange-gauge face contact, flange root-gauge corner/shoulder contact and multiple-point contact. Different contact conditions generate different creepage and stress distributions. Traditionally, different approaches are required for obtaining solutions under the different contact conditions.

The study of contact problems was initiated by Hertz [18], who solved the normal problem in frictionless contact. This solution involves circle and elliptical contact, which can be only applied to normal wheel tread-rail head contact. The earliest solution for frictional rolling contact was provided by Carter's [19], who presented a two-dimensional (2D) solution for the contact between two cylinders. The 2D solution was extended to the three-dimensional (3D) case by Johnson et al. [20, 21]. They considered both longitudinal and lateral creepage without spin, or pure spin only. Wheel-rail contact is a 3D problem, where the contact geometry is very complex and different creepage is coupled. The Kalker's variational theory based on virtual work is a method concerning wheel-rail rolling contact. It is capable of dealing with arbitrary creepage and spin. The actual geometry of wheel and rail can be considered. The Kalker's theory has been implemented using the Boundary Element Method (BEM) [22]. These works are based on a half-space assumption. To calculate the contact at gauge corner of rails, Li [23] extended the Kalker's theory by replacing the half-space with quasi-quarter

spaces for a more accurate stress-displacement relationship with consideration of varying spin and conformity in the contact patch. Like the approaches mentioned earlier, this approach is still limited to static elastic problems. Thus, these methods are unqualified for the study of rail squat formation where the dynamic effects and complex material properties need to be included.

FE methods are capable of overcoming the previously discussed limitations. They offer the possibilities to treat contact problems in a more general way. Then, the half-space assumption can be dropped. More complicated contact geometry and realistic material properties, such as elastic-plastic material property can be taken into account [24]. The FE methods with implicit solution schemes were initially applied to normal static contact problems, e.g., [25, 26]. They were subsequently employed to wheel-rail contact in statics, e.g., [27-29]. These approaches can only obtain normal contact solutions. The FE method for rolling contact was first provided by Padovan [30, 31] and Oden and Lin [32] in steady state. The approach of Padovan used a moving lagrange observer approach to handle the steady and transient response of rolling, but it restricts to kinematically linear problems. Oden and Lin used an arbitrary Lagrangian Eulerian (ALE) formulation to describe the rolling contact. Their approach was implemented for 2D deformable-rigid contact with a full slip in the entire contact patch. Nackenhorst [33, 34] subsequently extended the ALE formulation for the treatment of 3D stationary rolling and demonstrated the new formulation considering tire-road contact. However, the application of the ALE formulation to the wheel-rail system has not been successful because of the poor convergence in frictional rolling contact when only small deformations occur [33]. Zhao and Li [35] presented solution of frictional rolling contact using an explicit FE approach based on Lagrangian formulation that is considered to be more efficient and reliable for transient problems. An explicit time integration scheme is used to avoid the regularization problem encountered by the implicit scheme due to the no-slip condition in the stick area; the classical penalty method was employed for the contact algorithm; the accuracy of the method for rolling without large spin was verified. As in general, contact conditions might require to consider large spin, this thesis investigates spin-rolling contact. Additionally, the high-frequency vehicle-track dynamic interaction and dynamic effects should be taken into account when studying contact conditions at squats. In the literature, explicit FE approach has been reported of being able to couple contact mechanics and high-frequency structural dynamics in railway systems [35-37]. The importance of high-frequency dynamics at rail defects is discussed in the next section.

### 1.4.2. High-frequency dynamic interaction at rail defects

When a train travels on a railway track, the train and track system is commonly subjected to vibrations and dynamic contact loads induced by many excitation sources/components. These components include substructure, ballast, sleepers, sleepers supports, fastening system, rail, surface defects, rail joints, switches and crossings, vehicle suspensions and car bodies [38]. In modeling of vehicle-track dynamic interaction, the dynamic responses of these components should be considered, and the modeling methods should be capable of calculating the induced dynamic effects.

Dynamic vehicle-track interaction is usually modeled by the so-called mass-spring-beam based approaches [39], which use the coupling of multi-body models for the vehicles and beam models for the track through simplified wheel-rail contact, e.g., [40-44]. These approaches are usually only suitable for the dynamic interaction in the low- and mid-frequency range [45]. Moreover, local wear, contact stress and strain cannot be obtained

directly from these approaches. Instead, some other contact algorithms (e.g., FASTSIM and WEAR) are required, with the results of the mass-spring-beam based approaches as an input, e.g., [46-48].

Initiation of squats is related to the high-frequency components up to approximately 2 kHz [6]. The explicit dynamic FE method is one of the most effective ways to model the dynamic vehicle-track interaction while considering the local frictional rolling contact. With the method, high-frequency vibrations can be modeled and captured [35]. Using this method, an FE dynamic vehicle-track interaction model has been developed and applied to study the dynamic interaction at rail surface irregularities in [6, 36, 49]. Also, this FE model has been experimentally validated by Molodova et al. [50] using axle box acceleration (ABA) to detect squats. In this thesis, this model is further developed to study squat formation, in which the influence of dynamic effects and the resulting geometry evolution of squats are involved.

## 1.5. Geometry evolution of squats

Squats develop from surface irregularities. Geometry evolution is one of the causes of the development. The geometry evolves due to the differential plastic deformation and wear induced by the excited dynamic forces at the irregularities [6]. Therefore, the estimation of the plastic deformation and wear is important to understand the formation of squats. The widely used methods for calculating the plastic deformation and wear are discussed in this section.

### 1.5.1. Plastic deformation

Plastic deformation occurs when a material is subjected to the stresses that exceed its yield stress. The von Mises criterion and the Tresca criterion are commonly used to determine whether in a material plastic deformation occurs or not [51]. The von Mises criterion is based on the determination of the distortion energy in a given material. The Tresca criterion is based on the maximum shear stress in a given material. In practical applications, the von Mises criterion provides better correspondence with experimental data than the Tresca criterion [52]. In the literature, the von Mises criterion has been applied to the study of squats (e.g., [53]) and other types of rail damages (e.g., [54] and [46]). This thesis, therefore, uses the von Mises criterion to judge the plastic deformation of squats.

### 1.5.2. Wear

In frictional rolling contact, wear arises due to the relative motion of material points within the contact patch [55]. The rolling contact is accompanied by micro-slip, i.e., the relative velocity of the material points in a contact patch at the interface [35]. Two models, i.e., the Archard model [56] and the frictional work model [57, 58] have mainly been applied to evaluate wear in railway systems.

In the Archard model [56], the wear rate depends on the sliding distance, normal force and hardness of the material. The model calculates the removed volume  $V_{wear}$  of material with normal contact force, friction force, sliding distance and the material properties [56]. The expression is:

$$V_{wear} = \frac{KNs}{H}, \quad (1.1)$$

where  $N$  is the normal force,  $s$  is the sliding distance and  $H$  is the material hardness.  $K$  is wear coefficient, which is a constant for a particular material and can be evaluated by experimental testing. This wear model has been calibrated for wheel and rail steels [59, 60].

The frictional work model is based on an energy approach proceed from a proportional relationship of wear volume  $V_w$  and friction work  $W_f$  [57, 58]. In this model, friction work and wear volume at an arbitrary point of the rail or wheel surface are linked, as follows

$$V_w = \frac{W_f}{k}, \quad (1.2)$$

where  $k$  is a material constant, which can be determined by means of a twin disc machine.

In this thesis, due to the capability of considering material hardness with the explicit FE method, the wear model in equation (1.1) will be applied to study the development of surface irregularities at welds.

## 1.6. RCF crack initiation and propagation

In addition to the geometry evolution of surface irregularities, squat formation is associated with RCF cracks [3]. Thus, studying RCF cracks is crucial for a better understanding of squat formation. The development process of a fatigue crack is in general divided into two periods, i.e. crack initiation period and crack propagation period [61]. They are discussed in the following subsections.

### 1.6.1. Crack initiation

In wheel-rail rolling contact, surface-initiated RCF is generally resulted from plastic deformation in the surface [62, 63]. It was suggested that plastic flow produces voids and microcracks in the surface; and then, a fatigue crack is initiated by the growth or the coalescence of microcracks [64]. Microcracks may propagate by either of two ways, i.e., low cycle fatigue (or cyclic plastic strains) and ratcheting (or ductile shear fracture) [62].

The way of crack initiation depends on both the contact load characteristics and material behavior, since any of the four types of material behaviors, i.e., elastic, elastic shakedown, plastic shakedown/cyclic plasticity and ratcheting, could occur under rolling contact [65]. When the contact load exceeds the elastic shakedown limit but less than the plastic shakedown limit, the cycle of plasticity or the hysteresis loop is closed and the fatigue failure occurs by plastic shakedown or cyclic plasticity. This type of fatigue failure is characterized as low cycle fatigue. If the contact load exceeds the plastic shakedown limit, the material accumulates progressive unidirectional shear plastic deformation within a number of rolling cycles and the material fails due to ratcheting [66]. The material fails when the total accumulated strain reaches a critical value. This stage can be considered as the initiation period of the crack [67]. Cyclic plasticity and ratcheting can be considered as independent and competitive mechanisms [63].

The knowledge of fatigue mechanism is essential for a number of practical applications, including increasing fatigue strength of material in manufacturing, understanding the various influential factors of fatigue and predicting fatigue life [61]. For instance, among others, in terms of prediction of fatigue life, critical plane based models are widely applied for low cycle fatigue [68-70], while plastic strain based model is usually applied for the study of ratcheting [63].

Regarding the crack initiation of squats, the mechanism is still a controversial topic. The initiation is mostly thought to be caused by ratcheting, or white etching layer, or their combination [5, 71, 72]. However, the metallurgical observation in [73] showed that directional accumulated plastic deformation was not always observed along the running band at squats; instead, cyclic plasticity was also observed in some regions. Moreover, the observation of the plastic deformation in an early squat when crack just initiated was not found in the literature. This means that it is difficult to determine which material behavior causes crack initiation. Therefore, the mechanism of crack initiation is still required to be investigated.

### 1.6.2. Crack propagation

After initiation, a crack turns to a growth period when it reaches a critical length and a number of additional mechanisms start driving the growth of the crack [61, 74]. There is still no universal definition of the transition from crack initiation period to the growth period. One definition is that crack growth period starts when the critical length is sufficiently large to drive its propagation by fracture mechanism [75]. The growth of crack could be driven by one or several crack growth mechanisms. In rolling contact, it is thought that fluid entrapment is indispensable to generate tensile stress and promote crack growth [76-78], because the liquids could lower the friction of the crack face and may produce hydraulic pressurization which opens the crack by tensile mode (Mode I) mechanism [77]. Purely shear mode crack is unlikely to occur when the friction coefficient of the crack face is too large.

Regarding the crack propagation of squats, linear elastic fracture mechanics has been predominantly used to study crack growth path, growth rate and the influential factors by numerical analyses [77, 79-82]. These numerical analyses mainly focused on simplified crack geometry, contact geometry and loading condition. Crack face friction and influence of trapped lubricants were also considered. These studies provided insights into the influencing parameters of crack propagation. However, a satisfactory study on the formation mechanism of the realistic typical 3D V/U-shaped cracks at squats is still missing.

## 1.7. The gaps in the existing studies of rolling contact and rail squats

To efficiently mitigate the squat problems, reduce the cost of maintenance and ensure the safety of railway transportation, it is clearly required to understand in details the contact mechanics of wheel-rail, as well as the root causes and development processes of rail squats. Although an increasing number of works on rolling contact (e.g., [22, 24]) and rail squats (e.g., [2, 6, 9, 73, 77, 79-82]) have been conducted, they have various limitations in the calculation of wheel-rail contact and in the understanding of the physical phenomena of squats.

### 1.7.1. Calculation of wheel-rail contact

The analysis of wheel-rail contact stress is crucial to the study on the physical phenomenon of rail damage. Although different FE methods and traditional half-space-based methods are popular for various contact problems, their accuracies have not been fully verified. Two detailed gaps exist, as follows:

First, half-space-based methods have been widely applied to steady state elastic contact problems owing to their high computational efficiency. The half-space-based methods are also often applied to verify the accuracy of FE methods. It is generally believed that to guarantee

acceptable accuracy in these half-space-based methods, the significant dimensions (i.e., the height, width, length and the principal radii of curvature) of each contact body should be much larger than the characteristic size (twice length of one semi-axis) of the contact patch. However, this requirement has not been quantitatively justified. Moreover, in practice, the significant dimensions of the contact bodies may be close to the characteristic size of the contact patch, while the accuracy of these half-space-based methods for such cases has not been justified.

Second, the accuracy of the explicit FE method has not been verified in some complex rolling contact problems, such as frictional rolling with large spin. Large spin often arises at the contact between flange throat of the wheel and the gauge corner of the rail, and they significantly affect the contact behaviors. Due to the lack of studies in these topics, the reliability of a lot of existing works on wheel-rail contact with the FE methods is open to doubt when large spin is considered.

### 1.7.2. Understanding of rail squats

The gaps in understanding rail squats include the development process in terms of geometry evolution, crack initiation and propagation, as well as the causes behind. Some details are provided as follows.

First, the existing works (e.g., [77, 79-82]) on squat cracking concerned mature squats with large cracks. Satisfactory initiation mechanisms and early propagation of squat crack were not provided.

Second, although Li et al. [6, 9] gave insights into the mechanism of squat initiation in relation to high-frequency dynamic wheel-rail contact force, their studies mainly focused on the primary and root causes of the two-lung shape of squats. A full life cycle of squats through an efficient experimental study or long-term field monitoring has not been reported in the literature. A better understanding of the important physical phenomena in the development of squats, such as crack initiation and propagation processes, is still needed, as well as the causes of cracking.

Third, the mechanism proposed in [6] mainly focused on the cause in term of the dynamic wheel-rail contact force, whereas the cause with respect to the resulting differential plastic deformation and wear has not been provided. Moreover, the study in [6] focused on the squats caused by indentations. The formation of squats by other sources, such as corrugation, rail welds, wheel burns and insulated joints, has not been investigated.

Lastly, although squats induced by various sources have common characteristics in terms of appearance and shape of crack, whether they have a common formation mechanism to a certain degree requires a systematic study. To better understand rail squats, a unified formation theory for different types of squats still lacks.

## 1.8. Problem statement

In light of the motivation and the gaps described in the previous sections, the main research question of this thesis is set as:

What are the formation processes of rail squats in terms of the geometry evolution, crack initiation and propagation, as well as their root causes?

To answer this question, a set of sub-questions are formulated, as follows:

- 1) What is the applicability of half-space-based methods for contact problems with respect to the significant dimensions?
- 2) How well does the explicit FE method work in frictional rolling contact with large spin? What are the advantages of the explicit FE method in rolling contact with spin?
- 3) What are the fatigue crack initiation location and cracking behaviors at rail squats? How does the geometry of squats evolve? What are the causes of crack initiation and propagation, as well as the evolution of geometry?
- 4) What cause the rail surface irregularities that develop into squats? How do the rail surface irregularities develop before crack initiation? How do the mechanisms produce squats in term of not only dynamic forces but also differential plastic deformation and wear?

## 1.9. Outline of the thesis

To answer the questions above, this thesis first studies the applicability of half-space-based methods by using an FE method, then discusses the application of the explicit FE method to frictional rolling contact with large spin. Moreover, this thesis presents a five-year continual field monitoring study on squats and the comparison between field observation and numerical simulation. This thesis is divided into six chapters, as illustrated in Fig. 1.3.

Chapter 1 introduces the background of the research questions. The state of the art on rail squats is discussed. Not only the RCF and wear mechanisms but also the corresponding evaluation methods are reviewed. Furthermore, the research questions in geometry evolution and cracking associated with squat formation are summarized.

Chapter 2 applies the FE method to study the applicability of half-space-based methods for contact problems by comparing the solutions obtained from the Hertz theory and the Kalker's variational theory. Various combinations of significant dimensions in terms of height, width and length are considered; different contact patch eccentricities and contact body shapes are studied.

Chapter 3 verifies the accuracy of the explicit FE method for spin-rolling contact problem in quasi-static state against the Kalker's variational theory and the Hertz theory. The elastic-plastic problem in spin-rolling contact is further dealt with. The spin effects in rolling contact are also discussed.

Chapter 4 presents the observations and analysis of a five-year continual field monitoring of a large number of rail squats. Squats induced by corrugation are particularly considered. A development process in terms of crack initiation and propagation, as well as the geometry evolution, is proposed for this type of squats. The corresponding root causes in terms of tensile stress and impact of wheels are provided.

Chapter 5 studies the formation of squats caused by welds through the five-year field monitoring and numerical simulation. According to the field observations and measurements, a hypothesis of the formation process of weld-induced squats is proposed. The hypothesis includes three steps, i.e., Steps 1 and 2 for pre-cracking process, and Step 3 for post-cracking process. Thereafter, a 3D FE dynamic vehicle-track interaction model is built up to verify the first two steps. Varying yield stress based on the measured hardness at rail welds is used as inputs in the FE simulations. The causes for the formation of squats with respect to not only dynamic contact forces but also differential plastic deformation and wear are considered. Then

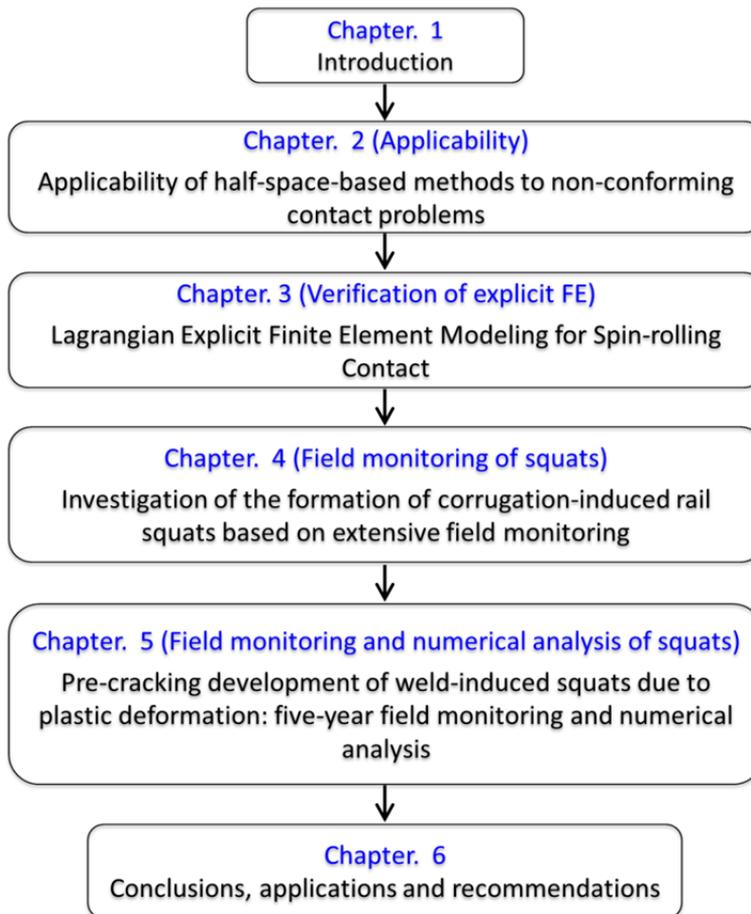


Fig. 1.3. Outline of this thesis.

the numerical results are compared with the field observations for the verification of the hypothesis.

Chapter 5 studies the formation of squats caused by welds through the five-year field monitoring and numerical simulation. According to the field observations and measurements, a hypothesis of the formation process of weld-induced squats is proposed. The hypothesis includes three steps, i.e., Steps 1 and 2 for pre-cracking process, and Step 3 for post-cracking process. Thereafter, a 3D FE dynamic vehicle-track interaction model is built up to verify the first two steps. Varying yield stress based on the measured hardness at rail welds is used as inputs in the FE simulations. The causes for the formation of squats with respect to not only dynamic contact forces but also differential plastic deformation and wear are considered. Then the numerical results are compared with the field observations for the verification of the hypothesis.

Chapter 6 summarizes the work and achievements of this thesis. Applications of the research in practice are also suggested. Outlooks and recommendations are given for future

development and potential application of the explicit FE methods. Further research on other types of RCF is also discussed.

## References

- [1] R. Nakamura, S. Owaku, N. Enomoto, The rail shelly crack in Japan, Railway Technical Research Institute, Quarterly Reports, 6 (1965).
- [2] P. Clayton, M. Allery, Metallurgical aspects of surface damage problems in rails, Canadian Metallurgical Quarterly, 21 (1982) 31-46.
- [3] S.L. Grassie, Squats and squat-type defects in rails: the understanding to date, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 226 (2011) 235-242.
- [4] J.J. Kalker, D.F. Cannon, O. Orringer, Rail quality and maintenance for modern railway operation, Springer Science & Business Media, 2013.
- [5] A. Al-Juboori, D. Wexler, H. Li, H. Zhu, C. Lu, A. McCusker, J. McLeod, S. Pannil, Z. Wang, Squat formation and the occurrence of two distinct classes of white etching layer on the surface of rail steel, International Journal of Fatigue, 104 (2017) 52-60.
- [6] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats—Correlation analysis and numerical modeling, Wear, 265 (2008) 1349-1355.
- [7] Z. Wei, A. Núñez, Z. Li, R. Dollevoet, Evaluating degradation at railway crossings using axle box acceleration measurements, Sensors, 17 (2017) 2236.
- [8] Z. Li, Squats on railway rails, Wheel–Rail Interface Handbook, Woodhead Publishing, 2009, pp. 409-436.
- [9] Z. Li, R. Dollevoet, M. Molodova, X. Zhao, Squat growth—Some observations and the validation of numerical predictions, Wear, 271 (2011) 148-157.
- [10] J. Smulders, Management and research tackle rolling contact fatigue, Railway Gazette International, 158 (2003) 439-442.
- [11] X. Zhao, Z. Li, R. Dollevoet, The vertical and the longitudinal dynamic responses of the vehicle-track system to squat-type short wavelength irregularity, Vehicle System Dynamics, 51 (2013) 1918-1937.
- [12] Z. Yang, Z. Li, R. Dollevoet, An Explicit Integration Finite Element Method for Impact Noise Generation at a Squat, Noise and Vibration Mitigation for Rail Transportation Systems, Springer, 2015, pp. 63-70.
- [13] S. Kaewunruen, M. Ishida, S. Marich, Dynamic Wheel–Rail Interaction Over Rail Squat Defects, Acoustics Australia, 43 (2015) 97-107.
- [14] Y. Berthier, J.B. Ayasse, C. Dagorn, L. Baillet, H. Chollet, M. Busquet, From railway dynamics to wheel/rail contact mechanics, an approach for the modelling of the wheel/rail contact: elasto-plastic response of the railhead, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 220 (2006) 189-200.
- [15] J.J. Kalker, Wheel-rail rolling contact theory, Wear, 144 (1991) 243-261.
- [16] K.L. Johnson, Contact mechanics, Cambridge university press, Cambridge, 1987.
- [17] R.P.B.J. Dollevoet, Design of an Anti Head Check profile based on stress relief, University of Twente, 2010.

- [18] H. Hertz, Über die berührung fester elastische Körper und über die Harte, *Journal für die reine und angewandte Mathematik*, 92 (1882) 156-171.
- [19] F.W. Carter, On the Action of a Locomotive Driving Wheel, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 112 (1926) 151-157.
- [20] P. Vermeulen, K. Johnson, Contact of nonspherical elastic bodies transmitting tangential forces, *Journal of Applied Mechanics*, 31 (1964) 338.
- [21] K.L. Johnson, The effect of spin upon the rolling motion of an elastic sphere on a plane, *Journal of Applied Mechanics*, 25 (1958) 332-338.
- [22] J.J. Kalker, *Three-dimensional elastic bodies in rolling contact*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [23] Z. Li, J.J. Kalker, Simulation of severe wheel-rail wear, *Proceedings of the Sixth International Conference on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems*, Southampton, (1998) 393-402.
- [24] X. Zhao, Z. Li, A three-dimensional finite element solution of frictional wheel-rail rolling contact in elasto-plasticity, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, (2014) 1350650114543717.
- [25] S.K. Chan, I.S. Tuba, A finite element method for contact problems of solid bodies—part I. Theory and validation, *Int J Mech Sci*, 13 (1971) 615-625.
- [26] P. Wriggers, T. Vu Van, E. Stein, Finite element formulation of large deformation impact-contact problems with friction, *Computers & Structures*, 37 (1990) 319-331.
- [27] M. Wiest, E. Kassa, W. Daves, J.C.O. Nielsen, H. Ossberger, Assessment of methods for calculating contact pressure in wheel-rail/switch contact, *Wear*, 265 (2008) 1439-1445.
- [28] U. Olofsson, T. Telliskivi, Contact mechanics analysis of measured wheel-rail profiles using the finite element method, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 215 (2001) 65-72.
- [29] F. Dörner, T. Bettinger, C. Schindler, Sensitivity analysis of the wheel-rail normal contact on the basis of continuous factors, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 231 (2017) 1022-1034.
- [30] J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure—I. Theory, *Computers & structures*, 27 (1987) 249-257.
- [31] N. Yukio, J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure—III. Impact/contact simulations, *Computers & structures*, 27 (1987) 275-286.
- [32] J. Oden, T. Lin, On the general rolling contact problem for finite deformations of a viscoelastic cylinder, *Computer Methods in Applied Mechanics and Engineering*, 57 (1986) 297-367.
- [33] U. Nackenhorst, The ALE-formulation of bodies in rolling contact: theoretical foundations and finite element approach, *Computer Methods in Applied Mechanics and Engineering*, 193 (2004) 4299-4322.

- [34] M. Ziefle, U. Nackenhorst, Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact, *Computational Mechanics*, 42 (2008) 337-356.
- [35] X. Zhao, Z. Li, The solution of frictional wheel-rail rolling contact with a 3D transient finite element model: Validation and error analysis, *Wear*, 271 (2011) 444-452.
- [36] Z. Li, X. Zhao, R. Dollevoet, An approach to determine a critical size for rolling contact fatigue initiating from rail surface defects, *International Journal of Rail Transportation*, 5 (2017) 16-37.
- [37] Z. Yang, X. Deng, Z. Li, Numerical modeling of dynamic frictional rolling contact with an explicit finite element method, *Tribology International*, 129 (2019) 214-231.
- [38] J.J. Zhu, A.K.W. Ahmed, S. Rakheja, A. Khajepour, Development of a vehicle-track model assembly and numerical method for simulation of wheel-rail dynamic interaction due to unsupported sleepers, *Vehicle System Dynamics*, 48 (2010) 1535-1552.
- [39] X. Zhao, *Dynamic wheel/rail rolling contact at singular defects with application to squats*, Delft: Delft University of Technology, 2012.
- [40] K.L. Knothe, S.L. Grassie, Modelling of Railway Track and Vehicle/Track Interaction at High Frequencies, *Vehicle System Dynamics*, 22 (1993) 209-262.
- [41] J.C. Nielsen, A. Igeland, Vertical dynamic interaction between train and track influence of wheel and track imperfections, *Journal of sound and vibration*, 187 (1995) 825-839.
- [42] W. Zhai, Z. Cai, Dynamic interaction between a lumped mass vehicle and a discretely supported continuous rail track, *Computers & structures*, 63 (1997) 987-997.
- [43] K. Popp, I. Kaiser, H. Kruse, System dynamics of railway vehicles and track, *Archive of Applied Mechanics*, 72 (2003) 949-961.
- [44] X. Chen, X. Deng, L. Xu, A Three-Dimensional Dynamic Model for Railway Vehicle-Track Interactions, *Journal of Computational and Nonlinear Dynamics*, 13 (2018) 071006.
- [45] Q. Wu, Y. Sun, M. Spiryagin, C. Cole, Parallel Co-Simulation Method for Railway Vehicle-Track Dynamics, *Journal of Computational and Nonlinear Dynamics*, 13 (2018) 041004.
- [46] A. Johansson, B. Pålsson, M. Ekh, J.C.O. Nielsen, M.K.A. Ander, J. Brouzoulis, E. Kassa, Simulation of wheel-rail contact and damage in switches & crossings, *Wear*, 271 (2011) 472-481.
- [47] N. Burgelman, Z. Li, R. Dollevoet, A new rolling contact method applied to conformal contact and the train-turnout interaction, *Wear*, 321 (2014) 94-105.
- [48] G. Tao, D. Ren, L. Wang, Z. Wen, X. Jin, Online prediction model for wheel wear considering track flexibility, *Multibody Syst Dyn*, 44 (2018) 313-334.
- [49] X. Zhao, Z. Li, J. Liu, Wheel-rail impact and the dynamic forces at discrete supports of rails in the presence of singular rail surface defects, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226 (2011) 124-139.
- [50] M. Molodova, Z. Li, A. Núñez, R. Dollevoet, Validation of a finite element model for axle box acceleration at squats in the high frequency range, *Computers & Structures*, 141 (2014) 84-93.

- [51] W. Soboyejo, Mechanical properties of engineered materials, CRC press, 2002.
- [52] R. Asaro, V. Lubarda, Mechanics of solids and materials, Cambridge University Press, 2006.
- [53] Z. Li, X. Zhao, R. Dollevoet, M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, *Vehicle System Dynamics*, 46 (2008) 237-246.
- [54] Y. Zhuang, K. Wang, Shakedown solutions for pavement structures with von Mises criterion subjected to Hertz loads, *Road Materials and Pavement Design*, 19 (2018) 710-726.
- [55] Y. Kimura, M. Sekizawa, A. Nitandai, Wear and fatigue in rolling contact, *Wear*, 253 (2002) 9-16.
- [56] J. Archard, Contact and rubbing of flat surfaces, *Journal of applied physics*, 24 (1953) 981-988.
- [57] L. Deters, S. Engel, M. Proksch, Friction and wear at tractive rolling of cylinders, *ZAMM - Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 80 (2000) 37-40.
- [58] L. Deters, M. Proksch, Friction and wear testing of rail and wheel material, *Wear*, 258 (2005) 981-991.
- [59] T. Jendel, Prediction of wheel profile wear—comparisons with field measurements, *Wear*, 253 (2002) 89-99.
- [60] J. Perez, J. Sinclair, J. Tunna, A review of wheel wear and rolling contact fatigue, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 221 (2007) 271-289.
- [61] J. Schijve, J. Schijve, J. Schijve, *Fatigue of structures and materials*, Springer, 2001.
- [62] A.F. Bower, K.L. Johnson, Plastic flow and shakedown of the rail surface in repeated wheel-rail contact, *Wear*, 144 (1991) 1-18.
- [63] A. Kapoor, A re-evaluation of the life to rupture of ductile Metals by cyclic plastic strain, *Fatigue & Fracture of Engineering Materials & Structures*, 17 (1994) 201-219.
- [64] N. P. Suh, The delamination theory of wear, *Wear*, 25 (1973) 111-124.
- [65] A.F. Bower, K.L. Johnson, The influence of strain hardening on cumulative plastic deformation in rolling and sliding line contact, *Journal of the Mechanics and Physics of Solids*, 37 (1989) 471-493.
- [66] W.R. Tyfour, J.H. Beynon, A. Kapoor, The steady state wear behaviour of pearlitic rail steel under dry rolling-sliding contact conditions, *Wear*, 180 (1995) 79-89.
- [67] J.W. Ringsberg, M. Loo-Morrey, B.L. Josefson, A. Kapoor, J.H. Beynon, Prediction of fatigue crack initiation for rolling contact fatigue, *International Journal of Fatigue*, 22 (2000) 205-215.
- [68] M. Brown, K. Miller, A theory for fatigue failure under multiaxial stress-strain conditions, *Proceedings of the Institution of Mechanical Engineers*, 187 (1973) 745-755.
- [69] D. Socie, Multiaxial Fatigue Damage Models, *Journal of Engineering Materials and Technology*, 109 (1987) 293-298.

- [70] Y. Jiang, H. Sehitoglu, A model for rolling contact failure, *Wear*, 224 (1999) 38-49.
- [71] S. Pal, W.J.T. Daniel, M. Farjoo, Early stages of rail squat formation and the role of a white etching layer, *International Journal of Fatigue*, 52 (2013) 144-156.
- [72] S. Pal, C. Valente, W. Daniel, M. Farjoo, Metallurgical and physical understanding of rail squat initiation and propagation, *Wear*, 284-285 (2012) 30-42.
- [73] S. Simon, A. Saulot, C. Dayot, X. Quost, Y. Berthier, Tribological characterization of rail squat defects, *Wear*, 297 (2013) 926-942.
- [74] D. Fletcher, F. Franklin, A. Kapoor, Rail surface fatigue and wear, *Wheel-Rail Interface Handbook*, (2009) 280.
- [75] H. Mughrabi, Microstructural mechanisms of cyclic deformation, fatigue crack initiation and early crack growth, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 373 (2015) 20140132.
- [76] S. Way, Pitting due to rolling contact, *ASME J. Applied Mechanics*, 2 (1935) A49.
- [77] A. Bower, The influence of crack face friction and trapped fluid on surface initiated rolling contact fatigue cracks, *Journal of Tribology*, 110 (1988) 704-711.
- [78] Y. Murakami, M. Kaneta, H. Yatsuzuka, Analysis of surface crack propagation in lubricated rolling contact, *ASLE transactions*, 28 (1985) 60-68.
- [79] S. Bogdański, M. Olzak, J. Stupnicki, Numerical stress analysis of rail rolling contact fatigue cracks, *Wear*, (1996).
- [80] S. Bogdański, Quasi-static and dynamic liquid solid interaction in 3D squat-type cracks, *Wear*, 314 (2014) 20-27.
- [81] M. Farjoo, W. Daniel, P.A. Meehan, Modelling a squat form crack on a rail laid on an elastic foundation, *Engineering Fracture Mechanics*, 85 (2012) 47-58.
- [82] M. Farjoo, S. Pal, W. Daniel, P.A. Meehan, Stress intensity factors around a 3D squat form crack and prediction of crack growth direction considering water entrapment and elastic foundation, *Engineering Fracture Mechanics*, 94 (2012) 37-55.



# 2

## APPLICABILITY OF HALF-SPACE-BASED METHODS TO NON-CONFORMING ELASTIC NORMAL CONTACT PROBLEMS

*The half-space assumption has been employed in many solution methods for non-conforming contact problems in elasticity such as the Hertz theory and the Kalker's variational theory. It is generally believed that to guarantee acceptable accuracy in these half-space-based methods, the characteristic size (twice length of one semi-axis) of the contact patch should be much smaller than the significant dimensions (i.e., the height, width length and the principal radii of curvature) of each body in contact. In engineering practice, the 3x rule is often employed, which requires that the significant dimensions be at least three times as large as the characteristic size. However, this requirement has not been justified. This chapter investigates the applicability of half-space-based methods by comparing the solutions obtained using two half-space-based methods (the Hertz theory and the Kalker's theory) with those of the FE method which is not limited to the half-space assumption. Different combinations of significant dimensions in terms of height, width and length are studied. Various contact patch eccentricities and contact body shapes are considered. It is found that the half-space-based methods yield high-accuracy calculation for non-conforming contact problems. Even when the significant dimensions are as small as 1.1x the characteristic size (twice length of one semi-axis), the differences between the solutions of the half-space-based methods and the FE method are within 9%. The findings of this chapter indicate that the typically assumed 3x restriction (the significant dimensions be at least three times as large as the characteristic size) can be greatly relaxed. Since a clear estimation of the deviation of the results of half-space-based methods from those of the FE method is provided, the applicability of half-space-based methods in mechanical engineering can be much better understood.*

---

This chapter is based on the publication below:

X. Deng, Z. Qian, Z. Li, R. Dollevoet. Applicability of half-space-based methods to non-conforming elastic normal contact problems. International Journal of Mechanical Sciences 2017;126:229-34.

## 2.1. Introduction

Contact problems exist in many mechanical systems such as wheel-rail interfaces, bearings, gears, mechanical linkages, and metal forming processes [1, 2]. These problems are related to the analysis of the friction, wear or fatigue failure of the components in contact and require accurate calculations of their stresses and deformations. The study of contact problems was initiated by Hertz [3], who solved the normal problem in which no friction was considered. A classical 2D solution for the frictional rolling contact problem was provided by Carter [4]. In the 1950s, Johnson [5] presented solutions for the frictional rolling of spheres. Concerning wheel-rail rolling contact, the most popular method is the Kalker's variational theory based on virtual work which is implemented using the Boundary Element Method [6]. This theory enables the treatment of arbitrary creepage and spin. All these theories are based on the half-space assumption. It is considered that this assumption requires that the characteristic size of the contact patch be much smaller than the significant dimensions of the bodies in contact. In the case of an elliptical contact, the characteristic size can be twice the length of one semi-axis. The significant dimensions of the contact bodies can be their width, length or height and/or the principal radii of curvature in the vicinity of the contact patch.

In reality, bodies in contact always have finite dimensions. In many cases, the significant dimensions of the body in contact are close to the characteristic size of the contact patch. Considering switch and crossing panels of railways as an example, the width and the radius of curvature at the crossing nose may be close to the characteristic size of the contact patch [7], as shown in Fig. 2.1(a). Other examples include asperity contacts (see Fig. 2.1(b)) [8, 9], bearing contact [10], spherical indentation contact [11] and gears [12]. In these contacts, one or more dimensions may be close to the characteristic size. Half-space-based methods have been widely used to solve these contact problems because such methods are efficient and easy to use. However, the applicability of these methods is questionable when the significant dimensions of the contact bodies are close to the characteristic size of the contact patch. In engineering practice the 3x rule is usually employed [13], which requires that the significant dimensions be at least three times as large as the characteristic size of the contact patch. However, to the authors' knowledge, the literature lacks sufficient justification of such a requirement or any analysis of its accuracy. This chapter proposes quantitative criteria for assessing the applicability of the classical half-space-based methods to non-conforming elastic normal contacts by considering the influence of the significant dimensions in the vicinity of the contact patch.

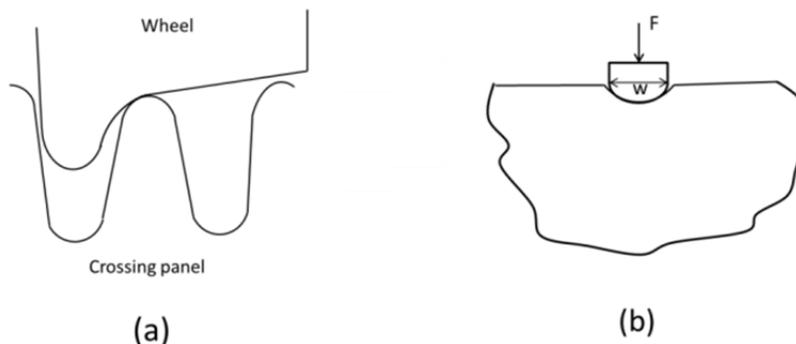


Fig. 2.1. (a) Wheel-rail contact in the case of a crossing panel [7]; (b) Single asperity contact [8].

In recent years, there have been many studies of various contact mechanical problems in which the real boundary conditions were considered. The FE method is commonly used in such studies. One advantage of the FE method is that the half-space assumption is not required. Its computational accuracy was verified several decades ago for static elastic contact problems by Chan and Tuba [14], and recently for frictional rolling contact problems by Zhao and Li [15]. Yan and Fischer [16] compared the contact pressure solution obtained using the FE method with that of the Hertz theory for the case of a standard rail, a crane rail and a switch. They found that the two solutions agreed well as long as the materials were assumed to be linear elastic in nature. Wiest et al. [7] investigated the contact pressure distribution between a wheel and a rail crossing nose using the FE method. They found that the solutions of the FE method and two half-space-based methods were in good agreement even though the radius of curvature at the crossing nose near the contact point was close to the size of the contact patch. In these works, the significant dimension was assumed to be the principal radius of curvature of the contact body, whereas other significant dimensions, such as height, width and length, were not addressed.

In this chapter, the FE method is employed to investigate the accuracy and applicability of half-space-based methods. The static normal contact between non-conforming geometric bodies is studied by comparing the solutions obtained using half-space-based methods with those of the FE method. A number of combinations of different significant dimensions of the contact bodies in terms of height, width and length with various eccentricities of the contact patch (the ratios of the semi-axes of the contact patch) and different shapes of the contact body are considered. The corresponding deviations of the half-space-based methods from the FE method are analyzed, and the critical significant dimensions are determined.

## 2.2. Assessment strategy on solution methods for contact problems

In this chapter, an FE contact model is built using the software package ANSYS. The Hertz theory and the Kalker's variational theory are chosen for comparison, as they are representative half-space-based methods. The deviations of the results of the half-space-based methods from those of the FE method are analyzed for each significant dimension. Notably, the results computed using the half-space-based methods are independent of the dimensions.

## 2.2.1. FE contact model

### 2.2.1.1. FE model description

In the FE model, two bodies with cylindrical geometries in static normal contact are considered. Schematic diagrams of the model are shown in Fig. 2.2. A Cartesian coordinate system  $Oxyz$  is created as shown in the figure, of which the origin  $O$  is located at the point of initial contact. The upper body is represented by a cylinder. The radius  $r_A$  is 460 mm, and the width  $w_A$  is 135 mm. The lower body is modeled as a partial cylinder whose cross section is perpendicular to the  $x$  axis. The radius of curvature  $r$  is 300 mm. Furthermore, cylindrical bodies in contact with other principal radii of curvature and contact bodies with arbitrary geometries are also considered to study the effects of the contact patch eccentricity and of the contact bodies shapes. The height  $h$ , the width  $w$  and the length  $l$  are selected in accordance with the designs to be discussed in Section 2.2. The lower body is connected to a base whose bottom is fixed.

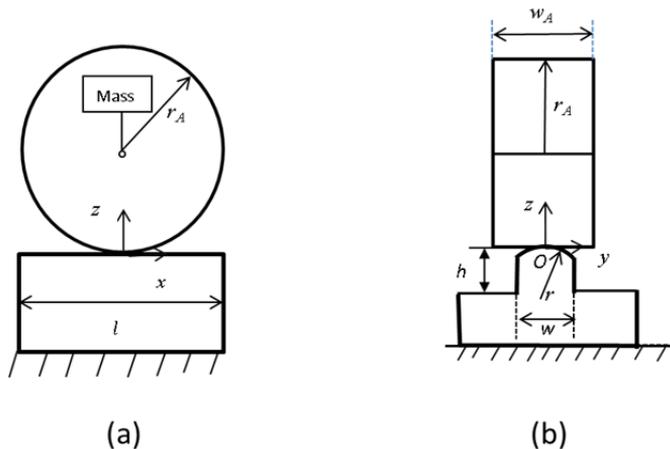


Fig. 2.2. Schematic diagrams of the FE model: (a) Front view; (b) Side view.

The contact bodies are meshed using 8-node solid elements. A non-uniform mesh is used in the potential contact zone to achieve high solution accuracy at a reasonable computational expense, as shown in Fig. 2.3. A fine mesh of 0.31 mm is assigned to both the lower body and the upper body in the potential contact zone. A relatively coarse mesh is used for the remaining portions of the bodies.

In the model, elastic material properties are assumed. The Young's modulus is 210 GPa, and the Poisson's ratio is 0.3. A total static vertical load of 86 kN is applied to the upper body.

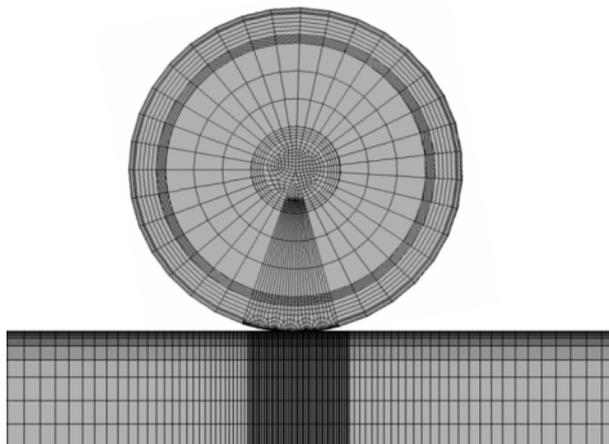


Fig. 2.3. Mesh of the FE model for static normal contact.

Various contact algorithms are available in ANSYS for studying the static contacts. In this chapter, the penalty method is used [17] to enforce the contact constraints. In studies of normal contact behavior, frictional contact is not considered, and therefore, no tangential load is applied. The model used in this chapter is static; thus, it is different from explicit dynamic models such as those in [15]. A static model is used in this chapter because the intent is to study the influence of the half-space assumption in half-space-based methods, and dynamic effects are not considered.

### 2.2.1.2. Design of the significant dimensions of the lower body

To quantitatively study the accuracy of half-space-based methods and the critical dimensions that most influence their accuracy, solutions for contacts between upper and lower bodies with various significant dimensions are analyzed. The significant dimensions of the lower body are designed based on the size of the contact patch as calculated using the Hertz theory. That is, in the design, the Hertz theory is first applied to compute the reference contact patch, which is elliptical in shape. The parameters  $a_0$  and  $b_0$  denote the reference major and minor semi-axes of the elliptical contact patch in the  $x$  and  $y$  directions, respectively. The Hertz solution is used as the reference because it can be easily calculated. The significant dimensions of the lower body in the FE model are defined as the height, width and length in the vicinity of the contact patch. Specifically, they are expressed as ratios of each actual dimension over the corresponding dimension of the reference contact patch size ( $2b_0$  or  $2a_0$ ). Therefore, the height, width and length are represented by the normalized variables  $h/2b_0$ ,  $w/2b_0$  and  $l/2a_0$ , respectively. The purpose of this normalization is to make this study more general. The normalized width of the lower body is selected to start with 1.1, which is very close to the extreme case of 1.0. Then, the width is varied from 1.1 to 2.0 and 8.0. The width of 8.0 is considered to represent the half-space scenario, as this value indicates a width that is much larger than the characteristic size of the contact patch. The length  $l/2a_0$  is set equal to the width  $w/2b_0$ . For each width, three heights ( $h/2b_0$ ), namely 0.4, 1.0 and 4.0, are considered. These combinations of significant dimensions in terms of height, width and length are investigated for the static normal contact problem.

### 2.2.2. The Kalker's variational theory

The theoretical basis and computer implementations of the Kalker's variational theory are described in [6]. A brief introduction to the theory is provided here. The theory follows the variational principle, which states that the stress field within a solid is true when the complementary potential energy is at its minimum value. It is implemented on the basis of the Boundary Element Method. The following assumptions are made. The contact patch is flat and it is small compared with the significant dimensions and the radii of curvatures of the two contact bodies. In addition, the inertial effect is ignored.

In the Cartesian coordinate system  $Oxyz$ , the contact problem can be expressed in the form of the minimization of complementary energy  $C_{u,p}$  in surface mechanical form. It is in terms of the surface tractions, including the normal pressure  $p_z$  and the tangential traction  $p_\tau$ . The expression is [6, 18]

$$\min C_{u,p} = \int_{A_c} \left( h + \frac{1}{2} u_z \right) p_z dS + \int_{A_c} \left( W_\tau + \frac{1}{2} u_\tau - u'_\tau \right) p_\tau dS \quad (2.1)$$

where  $p_z \geq 0$ ,  $|p_\tau| \leq f p_z$  in  $A_c$ ,  $f$  is the friction coefficient,  $A_c$  is the potential contact surface area,  $dS$  is the elementary area on  $A_c$ .  $h$  is the undeformed distance between the two contact bodies,  $u_z$  is the displacement in the normal direction,  $W_\tau$  is the rigid body shift,  $u_\tau$  denotes the displacement in the tangential direction,  $u'_\tau$  denotes the tangential displacement at the previous time step.

The contact bodies  $\alpha = 1, 2$  are presumed to satisfy the half-space assumption and the material is linear elastic. The relation between the displacement  $u_i$  in the direction  $i$  at the location  $x_\alpha$  (the response point) and the surface traction  $p_j$  in the direction  $j$  at the location  $y_\alpha$  (the impulse point) can be established according to Boussinesq [19] and Cerruti [20], which reads [6]

$$u_i(x_\alpha) = \iint_{A_c} A_{ij}(y_\alpha - x_\alpha) p_j(y_\alpha) dS \quad \alpha = 1, 2; i, j = x, y, z \quad (2.2)$$

where  $A_{ij}(y_\alpha - x_\alpha)$  is the influence function of body  $\alpha$ , and it can be determined analytically.

Applying the Boundary Element Method, Equation (2.1) can be discretized into

$$\min C_{p_{JJ}}^* = \frac{1}{2} p_{Ii} A_{Iij} p_{Jj} + \{ h_j p_{Jz} + (W_{J\tau} - u'_{J\tau}) p_{J\tau} \} \quad (2.3)$$

$$\text{sub } p_{Jz} \geq 0, |p_{J\tau}| \leq f p_{Jz}$$

where the subscript  $I, J$  denotes the uniformly discretized element, the  $p_{Jj}$  in  $C_{p_{JJ}}^*$  means that the primal variables for the minimization are the independent variables  $p_{Jj}$ .

In this chapter, the normal contact problem without tangential traction is considered, thus the term with respect to the tangential traction in Equation (2.3) vanishes. The influence function is determined by

$$A_{zz}(R) = \{(1 - \nu)/(\pi G)\} R^{-1} \quad (2.4)$$

where  $\nu$  is Poisson ratio,  $G$  is the combined shear modulus of the two contact bodies which reads  $1/G = 1/2(1/G_1 + 1/G_2)$ ,  $G_1$  and  $G_2$  are the shear modulus of the two contact bodies respectively.  $R$  is the distance between the response point and the impulse point.

A contact patch takes place where the surfaces of the two contact bodies coincide. The contact pressure can be computed according to the contact condition for the normal contact problems,

$$p_{Iz} = 0, \text{ the element is outside of the contact patch} \quad (2.5)$$

$$A_{IzJz}p_{Jz} + h_I = 0, \text{ the element is in contact} \quad (2.6)$$

When the normal force  $F_z$  is prescribed, then the following relationship must be satisfied,

$$F_z = Q \sum_I p_{Iz} \quad (2.7)$$

where  $Q$  is the area of an element.

### 2.2.3. Assessment strategy

As previously mentioned, the Hertz theory and the Kalker's theory are considered as representative half-space-based methods. The FE method is used to evaluate their applicability. The results of evaluation can be extended to other half-space-based methods. In the Kalker's theory, the potential contact region is divided into a number of square elements of equal size. To ensure a fair comparison with the FE method, the element size in the Kalker's theory is set equal to that in the zone of interest in the FE model, which is 0.31 mm.

To make this study more general and to facilitate the comparison between the Hertz theory, the Kalker's theory and the FE method, the FE solution is normalized with respect to the solutions of the Hertz theory and the Kalker's theory. The normalization is performed using critical values such as the maximum contact pressures ( $p_0$  and  $p_{k0}$ ) and the semi-axes of the contact patch ( $a_0$  and  $a_k$  in the  $x$  direction,  $b_0$  and  $b_k$  in the  $y$  direction) that are obtained from the Hertz theory and the Kalker's theory, respectively. The FE solution is treated as the reference. The deviations of the solutions of the half-space-based methods from the results of the FE method are therefore investigated.

In addition, contact patch eccentricity ( $b_0/a_0$ ) is one important parameter that is related to the accuracy of half-space-based methods. This eccentricity depends on the ratio of the relative principal curvatures of the contact bodies. Therefore  $b_0/a_0$  is varied by choosing various combinations of the principal radii of the curvature of the contact bodies. Using the real sizes of a wheel and a rail generates a contact patch with an eccentricity of 0.75. The value of  $b_0/a_0$  lies in the range 0 to 1.00. Three other values in this range (0.20, 0.50 and 1.00) are also investigated to study the effects of the contact patch eccentricity.

## 2.3. Simulation results

In this chapter, frictional effects are not included. Therefore, only normal contact behaviors are evaluated, including the contact pressure distribution as well as the shape and the size of the contact patch. The influence of the significant dimension is investigated by varying a

single dimension (e.g., height) while keeping the other dimensions (e.g., width and length) constant.

### 2.3.1. Normal solution when $b_0/a_0 = 0.75$

The cylindrical contact with a contact patch eccentricity of 0.75 is investigated in detail, and the results are presented in Table 2.1. In total, nine different combinations of significant dimensions of the contact body in terms of height, width and length are modeled. The contact patch for each combination is elliptical in shape, as shown in Fig. 2.4. The parameters  $a$  and  $b$  denote the semi-axes of the contact patch in the  $x$  and  $y$  directions, respectively.

Each normalized normal solution obtained using the FE method includes the normalized size of the contact patch and the maximum normalized contact pressure. A normalized value of 1.0 implies that the corresponding half-space-based solution matches the FE solution. It is observed that the values are not 1.0, which means that some differences do exist between the FE method and the two half-space-based methods. Possible causes include the finite dimension effect accounted for in the FE method, and numerical discretization.

Regarding the computed contact patch sizes, the maximum difference between the two half-space-based solutions and the FE solution is 3%, which is observed for the minor semi-axis  $b/b_k$  when comparing the FE method with the Kalker's theory. Moreover, Table 2.1 shows that the size of the contact patch is not affected by various combinations of significant dimensions. This finding suggests that the differences between the FE method and the half-space-based methods are due to the numerical discretization only. The corresponding error is up to 3% when the contact patch size is considered. The Hertz theory is an analytical approach, thus no numerical discretization process is involved. The Kalker's theory and the FE method involve numerical discretization, but the processes are different. For instance, nodes are located at the centers of the elements in the Kalker's theory, whereas the nodes are located at the corners of the elements in the FE method.

Regarding the maximum normalized contact pressures ( $p_{max}/p_{k0}$  and  $p_{max}/p_0$ ), their values are not equal to 1.0. This suggests that the FE method and the half-space-based methods predict different pressures. Among the nine combinations of significant dimensions, the case with a height of  $h/2b_0 = 0.4$  and a width of  $w/2b_0 = 8.0$  is the closest one to the half-space assumption. The corresponding pressure is 1.010, which indicates a difference of 1% in pressure between the FE method and the half-space-based methods. This difference is caused by numerical discretization. When the height is increasing to 4.0 and the width is decreasing to 1.1, the pressure increases by 3% from 1.010 to 1.040. This increase is caused by the finite dimension effect accounted for in the FE model. In other words, the half-space assumption introduces an error up to 3% in the maximum pressure compared with the FE method.

Table 2.1. FE solutions for static normal contact for each combination of significant dimensions.

| Dimensions of lower body |            | Size of contact patch |              |                 |              |              |              | Maximum pressure |                  |
|--------------------------|------------|-----------------------|--------------|-----------------|--------------|--------------|--------------|------------------|------------------|
| Height                   | Width      | Major semi-axis       |              | Minor semi-axis |              | Area         |              | $p_{max}/p_0$    | $p_{max}/p_{k0}$ |
| $h/2b_0$                 | $w/2b_0$   | $a/a_0$               | $a/a_k$      | $b/b_0$         | $b/b_k$      | $A/A_0$      | $A/A_k$      |                  |                  |
| <b>0.4</b>               | 1.1        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.024            | 1.024            |
|                          | 2.0        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.010            | 1.010            |
|                          | <b>8.0</b> | <b>1.001</b>          | <b>1.002</b> | <b>0.995</b>    | <b>0.970</b> | <b>0.996</b> | <b>0.972</b> | <b>1.010</b>     | <b>1.010</b>     |
| 1.0                      | 1.1        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.040            | 1.040            |
|                          | 2.0        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.010            | 1.010            |
|                          | 8.0        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.010            | 1.010            |
| <b>4.0</b>               | <b>1.1</b> | <b>1.001</b>          | <b>1.002</b> | <b>0.995</b>    | <b>0.970</b> | <b>0.996</b> | <b>0.972</b> | <b>1.040</b>     | <b>1.040</b>     |
|                          | 2.0        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.010            | 1.010            |
|                          | 8.0        | 1.001                 | 1.002        | 0.995           | 0.970        | 0.996        | 0.972        | 1.010            | 1.010            |

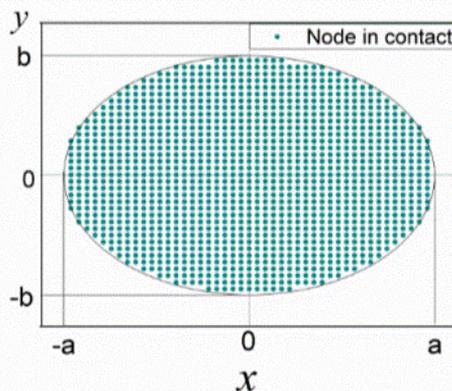


Fig. 2.4. Elliptical contact patch predicted by the FE method.

Fig. 2.5 shows the normalized contact pressure distributions along the longitudinal axis, as computed by the Hertz theory, the Kalker's theory and the FE method, for the most critical height of 4.0 and the width varying at 1.1, 2.0, and 8.0. All curves are Hertzian-type symmetric parabolas. This plot suggests that the half-space-based methods agree well with the FE method regarding the pattern of the pressure distribution.

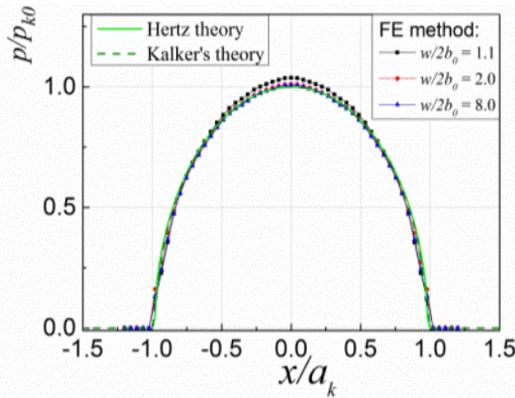


Fig. 2.5. Normalized contact pressure distributions along the longitudinal direction when  $h/2b_0 = 4.0$ .

### 2.3.2. Influence of the contact patch eccentricity $b_0/a_0$

The findings presented in the previous section are obtained based on a constant ratio of the principal radii of the cylindrical contact bodies, with the corresponding contact patch eccentricity of  $b_0/a_0 = 0.75$ . In this section, the ratio of the principal radii of cylindrical bodies is varied, with the resulting contact patch eccentricities of  $b_0/a_0 = 0.20, 0.50, 0.75$ , and  $1.00$ . The most critical combination of significant dimensions is examined, that is, a height of  $h/2b_0 = 4.0$  and a width of  $w/2b_0 = 1.1$ . The calculated contact patch sizes and maximum pressures are given in Table 2.2. The largest deviation is observed when the contact patch eccentricity  $b_0/a_0$  is  $0.20$ , and the major semi-axes  $a/a_0$  is in question. The difference is up to 9%. It consists of contributions from both the numerical discretization and the finite dimension effect. This finding suggests that the deviation of the half-space-based methods from the FE method is dependent on the contact geometry, i.e. the ratio of the principal radii of the contact bodies. However, the possible deviation is limited to 9% in all cases.

Table 2.2. FE solutions for static normal contact for various contact patch eccentricities when the contact body geometry is cylindrical.

| Eccentricity of contact patch | Dimensions of lower body |          | Sizes of contact patch |                 |         | Maximum pressure |
|-------------------------------|--------------------------|----------|------------------------|-----------------|---------|------------------|
|                               | Height                   | Width    | Major semi-axis        | Minor semi-axis | Area    |                  |
| $b_0/a_0$                     | $h/2b_0$                 | $w/2b_0$ | $a/a_0$                | $b/b_0$         | $A/A_0$ | $p_{max}/p_0$    |
| 0.20                          | 4.0                      | 1.1      | <b>1.090</b>           | 0.940           | 1.034   | 0.980            |
| 0.50                          | 4.0                      | 1.1      | 1.040                  | 0.997           | 1.037   | 1.015            |
| 0.75                          | 4.0                      | 1.1      | 1.001                  | 0.995           | 0.996   | 1.040            |
| 1.00                          | 4.0                      | 1.1      | 0.983                  | 0.963           | 0.946   | 1.063            |

### 2.3.3. Influence of the contact body shape

Two cylindrical contact bodies with arbitrary principal radii were considered in the previous section. In this section, the investigation is extended to the contact between two non-

cylindrical bodies with arbitrary principal radii. The resulting contact patches carry varying eccentricities of  $b_0/a_0 = 0.20, 0.50, 0.75$  and  $1.00$ . The most critical combination of significant dimensions, i.e. a height  $h/2b_0 = 4.0$  and a width  $w/2b_0 = 1.1$ , is studied.

The calculated contact patch sizes and maximum pressures are given in Table 2.3. Comparing Table 2.3 with Table 2.2, it is found that all corresponding results are very close to each other, although some slight differences exist. It is concluded that the deviation of the half-space-based methods from the FE method is no greater than 9% even when the contact bodies are not cylindrical.

Table 2.3. FE solutions for various contact patch eccentricities when the contact geometry is arbitrary.

| Eccentricity of contact patch | Dimensions of lower body |          | Sizes of contact patch |                 |         | Maximum pressure |
|-------------------------------|--------------------------|----------|------------------------|-----------------|---------|------------------|
|                               | Height                   | Width    | Major semi-axis        | Minor semi-axis | Area    |                  |
| $b_0/a_0$                     | $h/2b_0$                 | $w/2b_0$ | $a/a_0$                | $b/b_0$         | $A/A_0$ | $p_{max}/p_0$    |
| 0.20                          | 4.0                      | 1.1      | <b>1.090</b>           | 0.907           | 1.000   | 0.980            |
| 0.50                          | 4.0                      | 1.1      | 1.040                  | 0.997           | 1.037   | 1.020            |
| 0.75                          | 4.0                      | 1.1      | 1.009                  | 1.014           | 1.024   | 1.023            |
| 1.00                          | 4.0                      | 1.1      | 1.000                  | 0.968           | 0.968   | 1.070            |

## 2.4. Discussions

In this chapter, the contact patch sizes and the normal pressures predicted by the FE method are regarded as accurate and are used to assess the performance of half-space-based methods. However, the FE solutions may also carry some errors due to numerical discretization and therefore may deviate from reality. These numerical errors can either exaggerate or cancel out the differences between the half-space-based solutions and the FE solutions. It is important to keep this in mind when comparing the half-space-based methods with the FE method.

An idealized contact problem involves contact bodies with infinite dimension. The contact stresses in the contact bodies form a certain distribution at and near the contact location but vanish at infinity. However, in reality, all contact bodies have finite dimensions; therefore, the contact stress distribution deviates from the idealized situation because of boundaries at finite distances. When these boundaries move toward the contact location, i.e. as the significant dimensions decrease, their influence on the stresses is increasing. In half-space-based methods such as the Hertz theory and the Kalker's variational theory, the contact stresses are computed without taking into account the influences of these finite-distance boundaries. Therefore, their predictions may increasingly deviate from reality as the significant dimensions approaches the size of the contact patch. Despite this drawback, such methods have gained popularity for solving non-conforming contact problems with finite dimensions, as long as the significant dimensions of the contact bodies are at least 3x times larger than the characteristic size of the contact patch. This study indicates that this limit can be reduced to 1.1x times larger under the condition that a deviation of up to 9% introduced by the half-space assumption can be accepted (see Fig. 2.6). This conclusion is valid for most ratios of the relative principal radii of arbitrarily shaped contact bodies.

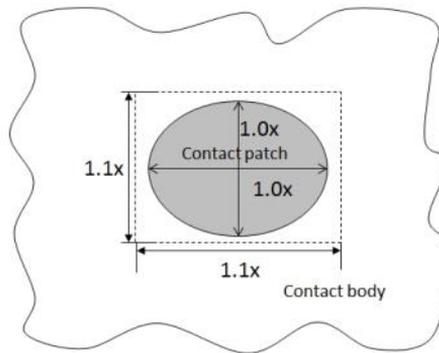


Fig. 2.6. Significant dimensions of the contact bodies and characteristic size of the contact patch (top view).

The simulations presented in the study of this chapter consider Hertzian non-conforming contacts between two arbitrary geometric bodies. The findings discussed above should also apply for the following types of contacts.

- 1) In the numerical simulations, specific chosen values were assigned to the dimensions of the contact bodies, the load and the material properties. When those values are scaled, the findings should remain valid as long as the contact problems in question remain linear elastic in nature and the deformation is small. This self-similarity rule is explained in [21, 22].
- 2) Although the simulations presented in this study consider Hertzian contact, the findings should also apply to some non-Hertzian elastic contacts, excluding conforming contacts. A typical non-Hertzian contact is the two-point contact between a wheel and a rail at the rail crossing nose [23, 24], as shown in Fig. 2.7(a). This type of contact problems results in two contact patches. For stress calculations using half-space-based methods, this problem may be decomposed into two individual contact problems at each point, as illustrated in Fig. 2.7(b). The part of the contact body in the vicinity of each contact point is considered as a local contact body. If each individual contact is of the Hertzian type and the significant dimension of each local contact body in the vicinity of the resulting contact patch satisfies the  $1.1x$  condition, then the findings of this study can be extended to the original two-point non-Hertzian contact. A more general type of non-Hertzian contact is a multi-point contact such as that shown in Fig. 2.7(c). Provided that the  $1.1x$  condition is satisfied for each individual contact problem, the decomposition technique can again be applied. Therefore the findings can be extended to a multi-point contact as well.

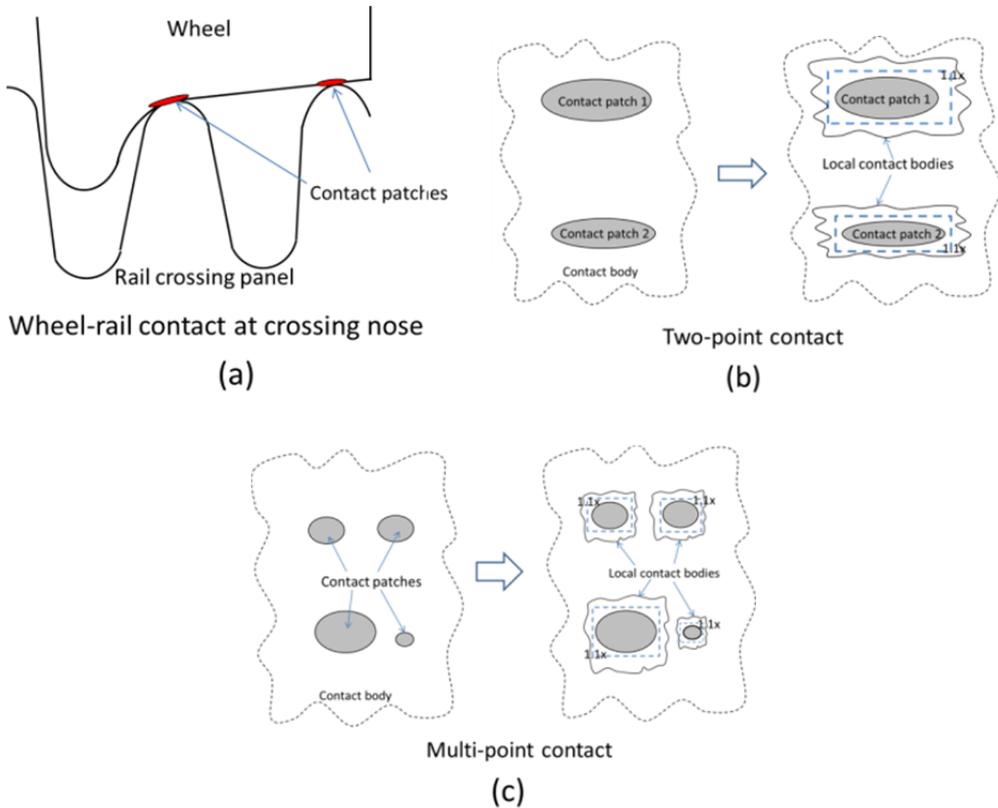


Fig. 2.7. Schematic diagrams of: (a) A wheel-rail contact at the rail crossing nose with two contact patches, a typical non-Hertzian contact problem; (b) Each contact patch satisfies the  $1.1x$  condition and thus the non-Hertzian problem of (a) may be decomposed into two Hertzian problems; (c) A general multi-point contact and the decomposition.

- 3) Although this study was implemented using static normal contact examples, the findings can also be applied to the case of a steady state rolling contact without traction. This is because the normal solution for static contact problems is equivalent to that for a steady state pure rolling contact problem. This equivalence can be seen from the identical normal solutions obtained for these two types of contact problems using the Kalker's theory. In practice, lubrication is widely applied to rolling components, such as bearings, to reduce the traction transmitted at the contact interface. Therefore these findings are also valid for most lubricated rolling contact problems. However, for a case in which the traction is too high to be ignored, the applicability of half-space-based methods for obtaining the tangential solution requires further investigation.

## 2.5. Conclusions

In this chapter, the FE method is used to assess the applicability of half-space-based methods to normal contact problems. Different combinations of significant dimensions of the contact bodies in terms of height, width, and length are studied. Various contact patch eccentricities are considered. Both cylindrical and non-cylindrical contact geometries are examined. The scenarios considered effectively cover the entire range of interest from the nearly smallest possible significant dimension (1.1x the characteristic size of the contact patch) to infinite dimension. Based on the observations from the numerical experiments and the analysis and the discussions, the following conclusions can be drawn.

- 1) Half-space-based methods can deliver reasonable accurate solutions for most non-conforming elastic contact problems, as long as the significant dimensions of the contact bodies are 1.1x larger than the characteristic size of the contact patch. The corresponding error is within 9%.
- 2) This new finding broadens the applicability of half-space-based methods, and lends greater confidence to studies of contact problems in mechanical engineering where half-space-based methods are applied. In engineering practice, this study can be a guide to assist engineers in choosing appropriate methods when solving contact problems. Should the 1.1x conditions be satisfied and an error tolerance of 9% be acceptable, the half-space-based methods are preferred, as they are relatively easier to apply and more computationally efficient. Otherwise, the FE method should be considered [24].

## References

- [1] K.L. Johnson, *Contact mechanics*, Cambridge university press, Cambridge, 1987.
- [2] F. Sadeghi, B. Jalalahmadi, T.S. Slack, N. Raje, N.K. Arakere, A review of rolling contact fatigue, *J Tribol*, 131 (2009) 041403.
- [3] H. Hertz, Über die berührung fester elastische Körper und über die Harte, *Journal für die reine und angewandte Mathematik*, 92 (1882) 156-171.
- [4] F.W. Carter, On the action of a locomotive driving wheel, *Proc R Soc London, Ser A*, 112 (1926) 151-157.
- [5] K.L. Johnson, Tangential traction and microslip in rolling contact, in: J.R. Bidwell (Ed.) *Rolling contact phenomena*, Elsevier, Amsterdam, 1962, pp. 6-28.
- [6] J.J. Kalker, *Three-dimensional elastic bodies in rolling contact*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [7] M. Wiest, E. Kassa, W. Daves, J.C.O. Nielsen, H. Ossberger, Assessment of methods for calculating contact pressure in wheel-rail/switch contact, *Wear*, 265 (2008) 1439-1445.
- [8] B. Bhushan, Contact mechanics of rough surfaces in tribology: multiple asperity contact, *Tribol Lett*, 4 (1998) 1-35.
- [9] P.K. Rajendrakumar, S.K. Biswas, Elastic contact between a cylindrical surface and a flat surface — a non-Hertzian model of multi-asperity contact, *Mech Res Commun*, 23 (1996) 367-380.
- [10] T.A. Harris, *Rolling bearing analysis*, John Wiley and sons, New York, 2001.
- [11] E. Yoffe, Modified Hertz theory for spherical indentation, *Philosophical Magazine A*, 50 (1984) 813-828.
- [12] S.-H. Wu, S.-J. Tsai, Contact stress analysis of skew conical involute gear drives in approximate line contact, *Mech Mach Theory*, 44 (2009) 1658-1676.
- [13] J. Kalker, Survey of wheel—rail rolling contact theory, *Veh Syst Dyn*, 8 (1979) 317-358.
- [14] S.K. Chan, I.S. Tuba, A finite element method for contact problems of solid bodies—part I. Theory and validation, *Int J Mech Sci*, 13 (1971) 615-625.
- [15] X. Zhao, Z. Li, The solution of frictional wheel–rail rolling contact with a 3D transient finite element model: Validation and error analysis, *Wear*, 271 (2011) 444-452.
- [16] W. Yan, F.D. Fischer, Applicability of the Hertz contact theory to rail-wheel contact problems, *Arch Appl Mech*, 70 (2000) 255-268.
- [17] P. Wriggers, Finite element algorithms for contact problems, *Archives of Computational Methods in Engineering*, 2 (1995) 1-49.
- [18] Z. Li, *Wheel-rail rolling contact and its application to wear simulation* (Ph.D. thesis), Delft University of Technology, Delft, The Netherlands, 2002.
- [19] J. Boussinesq, *Application des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques*, Gauthier-Villars, Paris, 1885.
- [20] V. Cerruti, *Accademia dei Lincei, Roma. Mem. fis. mat* (1882).
- [21] J.R. Barber, *Elasticity*, Springer, Kluwer, 1992.

- [22] F.M. Borodich, Similarity in the problem of contact between elastic bodies, *Journal of Applied Mathematics and Mechanics*, 47 (1983) 440-442.
- [23] E. Kassa, C. Andersson, J.C. Nielsen, Simulation of dynamic interaction between train and railway turnout, *Vehicle System Dynamics*, 44 (2006) 247-258.
- [24] Z. Wei, C. Shen, Z. Li, R. Dollevoet, Wheel–Rail Impact at Crossings: Relating Dynamic Frictional Contact to Degradation, *Journal of Computational and Nonlinear Dynamics*, 12 (2017) 041016-041011.

# 3

## LAGRANGIAN EXPLICIT FE MODELING FOR SPIN-ROLLING CONTACT

*Spin in frictional rolling contact can cause significant stress, which is the key to understanding and predicting the wear and fatigue behavior of contact components, such as wheels, rails and rolling bearings. The lateral creep force arising from spin influences the kinematics of a wheelset and thus of vehicles. The solution currently employed in the field of elasticity and continuum statics was developed by Kalker and uses a boundary element method (BEM). This chapter presents a new approach based on Lagrangian explicit FE method for spin-rolling contact. This approach is able to consider arbitrary geometry profiles of rails and wheels, complex material behavior and dynamic effects, etc. It is demonstrated using a 3D model of a wheel with a coned profile rolling along a quarter cylinder and can be easily adapted to wheel and rails of arbitrary profiles. The 3D FE model is configured with elastic material properties and used to obtain the normal and tangential solutions. The results are compared with those from the Hertz theory and the Kalker's variational theory. The model is then configured with elasto-plastic material properties to study the spin-rolling contact with plasticity. The continuum dynamics phenomenon is captured by the FE model, which increases the ability of the model to mimic reality. This improvement extends the applicability of the model considerably. The model can be applied to fatigue and wear analyses at gauge corners of rails, as well as to deep groove bearings where large geometric spin is present and plastic deformation may be important.*

---

This chapter is based on the following publication:

X. Deng, Z. Qian, R. Dollevoet. Lagrangian Explicit Finite Element Modeling for Spin-Rolling Contact. *Journal of Tribology*, 2015. 137(4): pp. 041401.

### 3.1. Introduction

Creep forces arise when creepage is present. Creep forces play an important role in wear, deformation and rolling contact fatigue of wheels and rails and are critical for guiding and stabilizing railway vehicles. Spin, a special component of creepage, occurs when there is an angular velocity about the local normal direction of the contact surfaces at the point of contact in question. It may arise due to the kinematics of the wheel, such as when the yaw velocity of the wheel is non-zero. Another cause is related to contact geometry, namely, a non-zero contact angle, which is the smallest angle between the local normal direction of the surfaces at the point of contact and the plane that is perpendicular to the rolling axis of the wheel, which is shown in Fig. 3.1. Spin resulting from this phenomenon is called geometric spin. This type of spin is constant when the contact patch is flat. It is often treated as constant when the contact patch is nearly flat, e.g., when it is between the wheel tread and the rail head. When the contact is between the flange throat and the gauge corner, especially when the profiles are worn so that the contact is conformal, the local normal direction in the contact patch varies significantly, and thus, the spin is not constant. With gauge corner contact, the contact angle is sufficiently large for geometric spin to dominate.

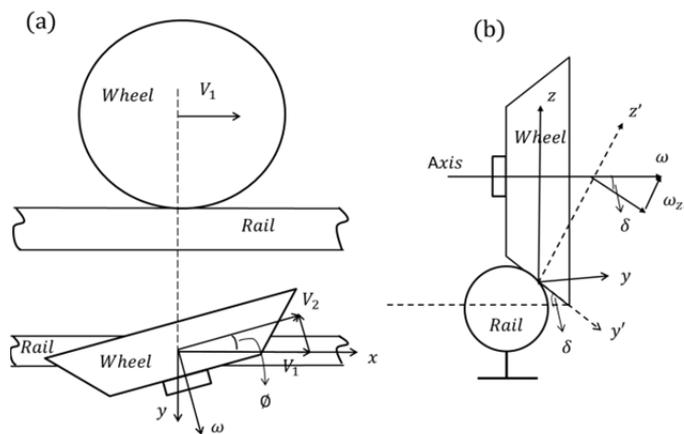


Fig. 3.1. (a) Spin due to the yaw velocity of the wheel:  $\phi$  is the yaw angle; (b) spin due to geometric factors:  $\delta$  is contact angle, and the wheel is rolling with an angular velocity  $\omega$  around the axis. A component of  $\omega_{zr}$  is produced due to the coned geometry of the wheel. (After Johnson [1]).

The importance of spin was first studied by Johnson in his analytical solution [2]. Johnson found that spin brings a twist to the contact interface. Hence, it gives rise to a lateral creep force and micro-slip. The relationship between spin and creep force has aroused the interest of many researchers due to its significance in wheel steering and vehicle dynamics. Kalker [3] developed a variational theory implemented with a BEM that treats the contact of an arbitrary spin magnitude with a flat contact patch. The works of Johnson and Kalker are based on a half-space approximation. To address wear and head checking at gauge corners, Li et al. [4-6] extended Kalker's work by replacing half-space with quasi-quarter spaces for a more accurate stress-displacement relationship and by considering varying spin in the contact patch. These

approaches are restricted to elastic static problems and therefore are not able to deal with non-linear material properties, more complicated contact geometries and dynamic rolling contact problems. Regarding the non-linear material in rolling contact, although some other studies published by Carbone and Putignano [7], and Chaise and Nelias [8] have considered viscoelastic material and elastic-plastic materials using BEM-based approaches for rolling contact, respectively, only the normal solution is included, that is, no friction and spin were considered in these studies.

The FE method offers more possibilities for the treatment of rolling contact. The half-space assumption can be omitted, and a more complicated contact geometry can be simulated. The FE method was first introduced into the solutions of steady-state rolling by Padovan [9, 10] and Oden and Lin [11]. Yukio and Padovan proposed the moving Lagrange observer approach and applied it to impact contact between deformable tire and rigid obstacles in two dimensions [10]. The disadvantages of this approach arise from the additive decomposition of motion with restrictions to kinematically linear problems. An arbitrary Lagrangian Eulerian (ALE) formulation was introduced in [11]. The method was implemented for 2D deformable-rigid contact and had difficulty in producing a no-slip velocity, and thus, the entire contact patch was in slip. Nackenhorst subsequently extended the ALE formulation for the treatment of 3D stationary rolling [12, 13] and demonstrated the new formulation considering tire-road contact. The ALE formulation is well established for the numerical analysis of rolling automotive tire behavior because it is a problem of large deformation and the contact can be considered with satisfactory accuracy as deformable-rigid. However, applications of the ALE formulation to the wheel-rail system have not been successful because of the poor convergence in frictional rolling contact when only small deformations occur [12]. Therefore, further research on reliable and efficient numerical algorithms is necessary in this field.

The explicit FE approach based on a Lagrangian formulation has been already proven to be accurate, efficient and reliable when applied to a modeling rolling contact [14, 15]. In this method, an explicit time integration scheme is used to avoid the regularization problem encountered by the implicit scheme due to the no-slip condition in the stick area [11, 12]. The classical penalty method is used for the contact algorithm. The primary advantage of the explicit technique is its stability in obtaining the solution and its low computational cost. Complex material constitutive relations can be easily implemented in this algorithm. Moreover, vibration modes and waves up to the necessary high frequency can readily be included in the solution [16-18], and the computational cost is not excessive.

Further developed from [15], this chapter presents the solution for spin-rolling contact. Because the explicit procedure tends to intrinsically include transient phenomena, this chapter first discusses how to attain quasi-static state rolling, as the spin-induced lateral contact force tends to induce additional vibrations. The accuracy of the quasi-static elastic solution is verified against the Hertz theory and the Kalker's theory since these two methods can solve elastic static contact problems with a high-level of accuracy, as studied in Chapter 2. The stresses and deformations are analyzed in detail when considering plasticity. In addition, the effects of spin and plastic deformation are discussed. The FE approach employed in this chapter is critical for modeling rolling contact with a complex geometry and for when non-linear material models are required for the analysis of wear, plastic deformation and rolling contact fatigue in the presence of geometric spin creepage.

## 3.2. Theoretical basis of the explicit FE model for rolling contact

### 3.2.1. Problem formulation of rolling contact

The continuum formulation for contact-impact problem [19] is adapted for rolling contact in this chapter. In this formulation, consider the motions of two deformable bodies  $B^1$  and  $B^2$  with boundaries  $\partial B^1$  and  $\partial B^2$ , respectively, in an undeformed configuration at time zero. At some later time, the two bodies become  $b^1$  and  $b^2$  with boundaries  $\partial b^1$  and  $\partial b^2$ , respectively. The deformed configurations cannot penetrate, i.e.,

$$b^1 \cap b^2 = \emptyset. \quad (3.1)$$

Two bodies come in contact if

$$\partial b^1 \cap \partial b^2 \neq \emptyset, \quad (3.2)$$

where  $\emptyset$  is an empty set.

The common contact surface is defined by

$$\partial c = \partial b^1 \cap \partial b^2. \quad (3.3)$$

For a given point  $X \in \partial c$  on the common surface, the normal vector must be collinear. Let  $\mathbf{n}^1$  and  $\mathbf{n}^2$  be the unit outward normal vectors. Let  $\dot{\mathbf{u}}$  be the velocity of material point  $X$ . Then,  $\dot{\mathbf{u}}_n^\alpha$  and  $\dot{\mathbf{u}}_s^\alpha$  ( $\alpha = 1, 2$ ), which are the normal and tangential components of  $\dot{\mathbf{u}}$ , respectively, are defined by

$$\dot{\mathbf{u}}_n^\alpha = (\dot{\mathbf{u}}^\alpha \cdot \mathbf{n}^\alpha) \mathbf{n}^\alpha \quad (3.4)$$

and

$$\dot{\mathbf{u}}_s^\alpha = \dot{\mathbf{u}}^\alpha - \dot{\mathbf{u}}_n^\alpha \quad (3.5)$$

Let  $\mathbf{s}^1$  and  $\mathbf{s}^2$  be the vectors tangent to the surface in the direction of the relative velocity of the surface. Normal and tangential vectors to the surface  $\partial c$  at  $X$  are governed by

$$\mathbf{n}^1 = -\mathbf{n}^2, \mathbf{s}^1 = -\mathbf{s}^2 \quad (3.6)$$

On the contact surface, the impenetrability condition has to be satisfied:

$$\mathbf{v}_N = \dot{\mathbf{u}}^1 \cdot \mathbf{n}^1 + \dot{\mathbf{u}}^2 \cdot \mathbf{n}^2 \leq 0 \quad (3.7)$$

The impenetrability condition expresses the fact that when the two surfaces are in contact, they must remain in contact ( $\mathbf{v}_N = 0$ ) or they must separate ( $\mathbf{v}_N < 0$ ).

The relative tangential velocity is given by

$$\mathbf{v}_T = \dot{\mathbf{u}}^1 \cdot \mathbf{s}^1 - \dot{\mathbf{u}}^2 \cdot \mathbf{s}^2 \quad (3.8)$$

The traction components  $\tau_i^\alpha$  acting on the surface at  $X$  are given by

$$\tau_i^\alpha = \sigma_{ij}^\alpha n_j \quad (3.9)$$

where  $\sigma_{ij}$  is the Cauchy stress tensor (The range on the indices  $i, j = 1, 2, 3$  and the standard summation convention for tensor components is implied).

The requirement of momentum balance across the contact surfaces results in the following interface traction condition at  $X$ :

$$\boldsymbol{\tau}^1 = -\boldsymbol{\tau}^2 \quad (3.10)$$

### 3.2.2. Lagrangian explicit FE scheme

For dynamic contact problems, based on a Lagrangian formulation [20], with an application of the FE discretization procedure [21] to the governing equation of the contact system, the discretized equation of motion to be solved with explicit FE analysis at time  $t$  may be expressed in the following form:

$$\mathbf{M}^t \ddot{\mathbf{u}} = \mathbf{Q}^t - \mathbf{F}^t + \mathbf{R}_c^t \quad (3.11)$$

where  $\mathbf{M}$  is the mass matrix,  $\ddot{\mathbf{u}}$  is the acceleration vector,  $\mathbf{Q}$  is the prescribed external load vector,  $\mathbf{F}$  is the internal force vector and  $\mathbf{R}_c$  is the force vector associated with contact forces.

By integrating Equation (3.11) over time, the displacements  $\mathbf{u}$  at time  $t + 1$  are obtained with the explicit central difference time integration scheme:

$$\begin{aligned} \dot{\mathbf{u}}^{t+1/2} &= \dot{\mathbf{u}}^{t-1/2} + \Delta t^{t+1/2} \ddot{\mathbf{u}}^t \\ \mathbf{u}^{t+1} &= \mathbf{u}^t + \Delta t^{t+1} \dot{\mathbf{u}}^{t+1/2} \end{aligned} \quad (3.12)$$

where  $\Delta t^{t+1/2} = 1/2(\Delta t^t + \Delta t^{t+1})$ ,  $\Delta t^{t+1}$  is the time increment that can be determined in last step based on the characteristic dimension of the element  $L_c$  and the dilatational wave speed of the material  $C_d$ . The dilatational wave speed is given as [22]

$$C_d = \sqrt{\frac{E(1-\nu)}{(1+\nu)(1-2\nu)\rho}} \quad (3.13)$$

where  $\nu$  and  $E$  are Poisson's ratio and Young's modulus, respectively, and  $\rho$  is the mass density.

The Courant-Fredrichs-Levy stability criterion [23] must be satisfied to ensure the stability of the calculation, i.e.,

$$\Delta t \leq \frac{L_c}{C_d} \quad (3.14)$$

### 3.2.3. Contact algorithm

A surface-to-surface contact-searching scheme based on a master-slave algorithm is used [24]. The kinematic contact condition is enforced using the penalty method [25, 26]. In applying this penalty method, each slave node is checked for penetration through the master surface. If the slave node does penetrate, then an interface force is applied between the slave node and its contact point. The magnitude of this force is proportional to the extent of penetration. This force can be thought of as the addition of an interface spring. Note that the time-step can be influenced by the penalty method. When the penalty method is employed, it is equivalent to

inserting a spring between two contact bodies; a contact critical time-step is calculated based on the contact stiffness. If the contact critical time-step is smaller than the critical time-step based on the characteristic element size, the contact critical time-step is used in the computations. Therefore, the stiffness is chosen so that the necessary accuracy is achieved with the lowest computational expense.

Friction is based on a Coulomb formulation [27]. Assume  $\mathbf{F}_n$  is the normal force, and  $f$  is the frictional coefficient. Thus, the limiting friction force is  $f|\mathbf{F}_n|$ . In this approach, the initial trial frictional force is calculated as

$$\mathbf{F}_f^* = \mathbf{F}_f^t - k\Delta\mathbf{u}_s \quad (3.15)$$

where  $k$  is the interfacial stiffness,  $\Delta\mathbf{u}_s$  is the incremental tangential displacement during a time increment, and  $\mathbf{F}_f^t$  is the frictional force at time  $t$ .

If  $\mathbf{F}_f^* \leq f|\mathbf{F}_n|$ , then the contact points are in stick, which means that the relative velocity between the two contact points is zero, and the frictional force at time  $t + 1$  is

$$\mathbf{F}_f^{t+1} = \mathbf{F}_f^* \quad (3.16)$$

If  $\mathbf{F}_f^* > f|\mathbf{F}_n|$ , then contact points are in slip, and the frictional force at time  $t + 1$  is

$$\mathbf{F}_f^{t+1} = f|\mathbf{F}_n| \frac{\mathbf{F}_f^*}{|\mathbf{F}_f^*|}. \quad (3.17)$$

### 3.3. Description of the FE model

#### 3.3.1. Model description

When one contact body rolls over another, vibration and wave propagation will physically be excited in the bodies. In statics, such dynamic effects are ignored, such as in the solutions of Hertz [28], Johnson [2] and Kalker [3]. However, an explicit integration will intrinsically include such vibration and waves. To verify the explicit FE method on spin-rolling contact against the results of Hertz and Kalker, the aforementioned dynamic effects should be minimized. To this end, the rail is generalized as a quarter cylinder with its flat surfaces constrained, which is shown in Fig. 3.2 and 3.3. A coned wheel tread is considered. Another reason for this simplification is to study a general spin-rolling contact because spin can occur not only at the wheel-rail contact but also some other contact components, such as deep groove ball bearings and roller-guide.

In a simulation, the wheel rolls along the cylinder starting from point  $P$ , as illustrated in Fig. 3.3. Two Cartesian coordinate systems  $Oxyz$  and  $Oxy'z'$  are defined with the origin at  $O$  and the  $x$ -axis along the rolling direction.  $y$  is the horizontal direction, and  $y'$  is tangent to the contact surfaces at  $O$ . The angle between  $y$  and  $y'$  is the contact angle  $\delta$ . The spin automatically occurs in the rolling process of the wheel. The magnitude of the spin creepage depends on the contact angle. Contact angles of  $12.5^\circ$  and  $25^\circ$  are considered in this study. The coefficient of friction is  $f = 0.5$ . With  $f = 0.5$ , the friction angle is  $26.6^\circ$ , and thus, for a contact angle of  $25^\circ$ , the wheel is on the verge of sliding down and off the quarter cylinder under vertical force alone.

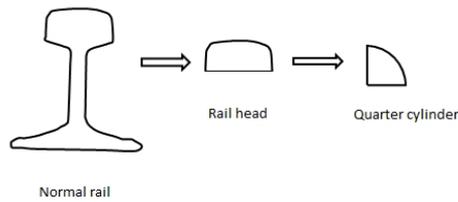


Fig. 3.2. Generalization of the rail gage corner.

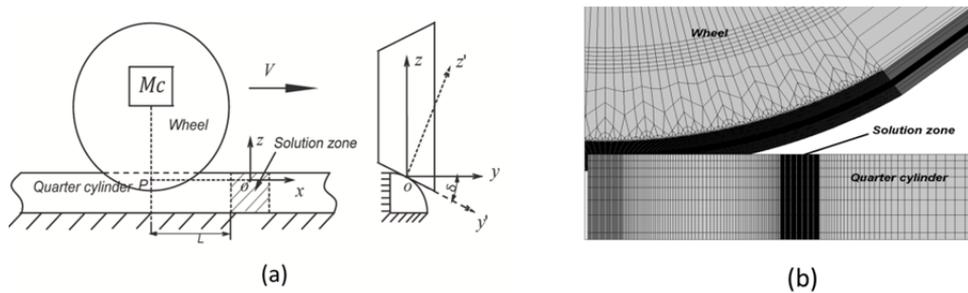


Fig. 3.3. FE model: (a) schematic diagram; (b) mesh.

### 3.3.2. Mesh

In the FE model, the wheel and cylindrical rail are meshed with 8-node solid elements. A finer mesh is applied to the solution zone to obtain a solution with sufficient accuracy at acceptable computation costs, as shown in Fig. 3.3(b), where more detailed analyses are performed, with a finer grid of  $0.34 \times 0.34$  mm over the contact surface.  $O$  is in the middle of the solution zone and is the point of contact when the two bodies are considered to be rigid.

### 3.3.3. Boundary conditions

The vertical surface of the quarter cylinder is constrained in the  $y$  direction, and its bottom is fixed, as shown in Fig. 3.3(a). The ends of the cylinder are constrained in the  $y$  and  $x$  directions. The dimensions of the cylinder are sufficiently large for the constraints on the boundaries to have a negligible influence on the contact behavior. The axis of the wheel is constrained in the  $y$  direction.

### 3.3.4. Initial equilibrium and dynamic relaxation

Before the wheel is set to roll, the wheel is brought into contact with the rail at the starting point  $P$  (Fig. 3.3). In an explicit solution procedure, if the wheel comes in contact with the cylinder too fast, an impact will occur, and the system will oscillate violently. If the contact occurs in a quasi-static manner, the process will take too long. To minimize both the dynamic effects and the time of this contact, each simulation procedure is performed with a coupling of static equilibrium analysis and explicit analysis for the rolling process. An implicit method is first used to calculate the initial static equilibrium of the contact under gravity. The displacement, strain and stress from the implicit solution are used as the initial conditions in the explicit analysis. Then, the explicit analysis begins at time zero with a stable pre-stressed

system after applying additional rolling speed and torque about the rolling axis of the wheel. In the explicit analysis, a dynamic relaxation process is used to damp out the dynamic effects due to imperfect static equilibrium by letting the wheel travel a distance  $L$ . Distance  $L$  is designed to be long enough to relax the dynamic effects to an approximate steady state before entering the solution zone. This dynamic relaxation is performed to dissipate the internal energy induced by incompatibility in the displacement fields, which is due to the imperfect static equilibrium solution. Meanwhile, in the explicit analysis, to damp out high-frequency oscillations, a material damping is also applied on the wheel and the quarter cylinder using Rayleigh damping [21]

$$C = \alpha M + \beta K \quad (3.18)$$

where  $C$ ,  $M$ , and  $K$  are the damping, mass, and stiffness matrices, respectively.  $\alpha$  and  $\beta$  are the mass- and stiffness-proportional damping constants, respectively. In this thesis, the mass-proportional damping is not included because it is only effective for low frequencies and will damp rigid body motion. After the system reaches a quasi-static state, a solution is determined at the instant when the center of the contact patch is at  $O$ .

### 3.3.5. Tangential loading conditions

When a driving or braking torque is applied to the wheel, a longitudinal tangential force  $F_L$  will develop in the rolling direction at the contact. The corresponding traction coefficient is

$$\mu = \frac{F_L}{F_N} \leq f \quad (3.19)$$

where  $F_N$  is the normal contact force and  $f$  is the friction coefficient. This is the conventional traction coefficient.

When spin is present, a lateral tangential force will arise. Together with  $F_L$ , the resultant tangential force is  $F_R$ , and a resultant traction coefficient can be defined as

$$\mu_R = \frac{F_R}{F_N} \leq f \quad (3.20)$$

The friction law must also be satisfied for the surface stresses at every point in the contact patch, i.e.,

$$\frac{\tau_R}{p_N} \leq f \quad (3.21)$$

where  $\tau_R$  is the resultant surface shear stress and  $p_N$  is the pressure.

Therefore, in the present work, the solution for spin-rolling contact is demonstrated with a combination of geometric spin and applied longitudinal creep force. Lateral creep force is not applied but instead arises as a result of the spin and the components of the weight of the wheel and the mass  $M_c$ . This lateral creep force due to spin is applied in the  $y'$  direction in Fig. 3.3.

### 3.3.6. Material properties

Linear elastic material with a Young's modulus of 210 GPa and a Poisson's ratio of 0.3 is considered for verification of the FE model against the Hertz theory and the Kalker's theory.

In Section 3.5, a bilinear plastic material model with kinematic hardening is used for the elasto-plastic analysis. This model assumes the total stress range is equal to twice the yield stress such that the Bauschinger effect is included. The material behavior is described by a bilinear total stress-total strain curve starting at the origin and with positive stress and strain values. The initial slope of the curve is taken as the elastic modulus of the material. At the specified yield stress, the curve continues along the second slope defined by the tangent modulus. A yield stress of 800 MPa and a tangential modulus of 21 GPa are used in this material model.

### 3.3.7. Calculation of the physical quantities in the contact patch

To obtain a solution in the contact patch that includes the pressure, surface shear stress, micro-slip, and stick and slip areas, the nodal force, displacement and velocity in the general coordinate system  $Oxyz$  are transformed into the local coordinate system  $Oxy'z'$  (Fig. 3.3).

The contact patch is determined with the normal nodal force on surface as follows. A node is in the contact patch if

$$F_{n_N} < \varepsilon_N \quad (3.22)$$

where  $F_{n_N}$  ( $n$  is the node number) is the nodal force in the normal direction of the contact surface. Since no tension in the contact patch is introduced, the normal nodal force is always negative in the contact zone.  $\varepsilon_N$  is the minimum inertial force of all nodes on the contact surfaces. In this chapter,  $F_{n_N}$  can be calculated using the formula below:

$$F_{n_N} = \sin\delta F_y + \cos\delta F_z \quad (3.23)$$

where  $F_y$  and  $F_z$  are the nodal forces in the  $y$  and  $z$  directions, respectively, and  $\delta$  is the contact angel, as shown in Fig. 3.3(a).

For the tangential contact force, the local  $x$  direction is the same as the global  $x$  direction. The nodal force in the  $y'$  direction is

$$F_{y'} = \cos\delta F_y - \sin\delta F_z \quad (3.24)$$

where  $F_y$  and  $F_z$  are the nodal forces in the  $y$  and  $z$  directions, respectively. The frictional force  $F_f$  can be calculated based on the nodal forces in the  $x$  and  $y'$  directions. The stick and slip areas can be determined by comparing the frictional force with the limiting frictional force  $f|F_{n_N}|$ . If the frictional force of a node

$$F_f < f|F_{n_N}| \quad (3.25)$$

the node is in the stick area. Otherwise, the node is in the slip area. Micro-slip in the slip area can be obtained according to Eq. (3.8).

According to the normal and frictional forces, the contact pressure and the surface shear stress can be obtained in the plane  $y'Ox$ , respectively.

### 3.4. Analysis of elastic spin-rolling contact and verification

The contact angle  $\delta = 12.5^\circ$  is mainly used for verification in this section. The contact angle  $\delta = 25^\circ$  is used for comparison when necessary. The physical quantities are transformed into the local coordinate system  $Oxy'z'$  (Fig. 3.3(a)).

#### 3.4.1. Pressure distribution

Two cases are investigated in this section, namely, a traction coefficient  $\mu = 0$  for pure spin and  $\mu = 0.3$  for tractive rolling when the contact angle is  $\delta = 12.5^\circ$ .

The FE results of the two cases are compared with those obtained from the Kalker's theory and the Hertz theory in Fig. 3.4. The results obtained in this study and those from the Kalker's theory and the Hertz theory correspond well with each other. In the proposed model, the normal and tangential problems of rolling contact are coupled and solved together. In the Hertz theory, the effect of the frictional tangential contact stress is not included, and the effect of the tangential contact stress on the normal solution is negligible with steel-on-steel contact in the presence of spin.

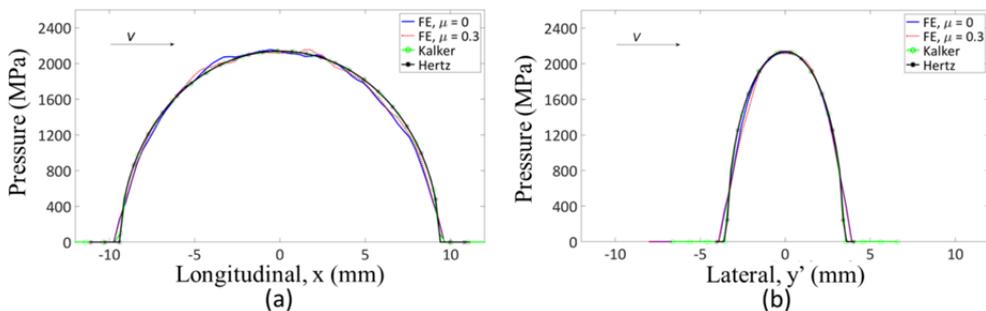


Fig. 3.4. The distributions of contact pressure ( $\delta = 12.5^\circ$ ): (a) along the longitudinal axis ( $y' = 0$ ); (b) along the lateral axis ( $x = 0$ ).

Some small fluctuations occur in the FE solutions due to the high-frequency vibration excited by the moving wheel on the cylinder. Small differences in the contact patch are produced by the different methods due to the difference in the finite size of the elements in the Kalker's theory and the FE model, the half-space approximation employed by the Hertz theory and the Kalker's theory, and possibly by the high-frequency oscillations.

#### 3.4.2. Tangential solutions

In this section, surface shear stresses, stick and slip areas and micro-slip in the contact patch are investigated.

##### 3.4.2.1. Surface shear stress distribution

Fig. 3.5 shows the surface shear stress  $\tau_s$  computed by the FE model and Kalker's theory for  $\mu = 0$  and  $0.3$  and  $\delta = 12.5^\circ$ . The arrows point in the direction of the shear stress and their lengths are proportional to their magnitudes. The surface shear stress turns around a spin pole,

where slip vanishes at this point. For  $\mu = 0$ , the spin pole is located at a point near the trailing edge but is out of the contact patch, as shown in Fig. 3.5. For  $\mu = 0.3$ , the spin pole is away from the pole for  $\mu = 0$  and moves to the left bottom side due to the influence of the longitudinal traction force, as observed by Wernitz [29].

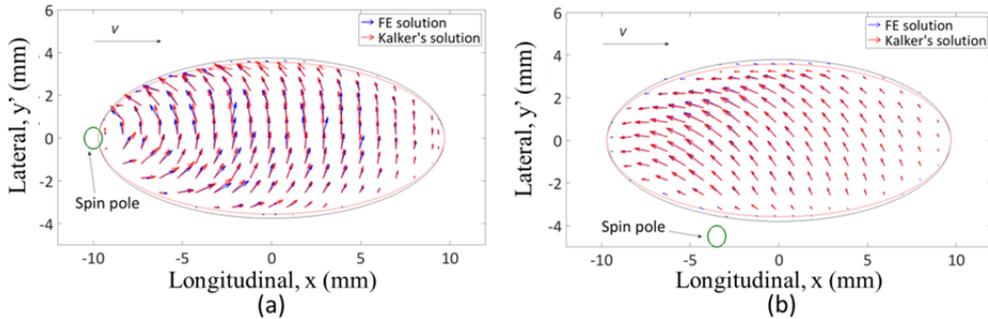


Fig. 3.5. The vector graph of the surface shear stress for different traction coefficients ( $\delta = 12.5^\circ$ ): (a)  $\mu = 0$ ; (b)  $\mu = 0.3$ .

Fig. 3.5 also shows that the direction and magnitude of the surface shear stress from the FE solution correspond well with those of Kalker's solution. Taking the case of  $\mu = 0.3$  as an example, the agreement in magnitude of the surface shear stress along the longitudinal axis is more clearly shown in Fig. 3.6(a), which illustrates the distribution of surface shear stress and limiting shear stress  $\tau_{max}$  (where  $\tau_{max} = f p_N$ ). The small difference in the surface shear stress at the border between the slip and stick areas is attributed to the different discretization and finite size of the elements in the two methods. In addition, some differences in the magnitudes of the stresses exist not only in the shear stress but also in the pressure when the contact angle is increased to  $25^\circ$ , as shown in Fig. 3.6(b). These differences are due to the larger contact angle, leading to a higher frequency vibration and increased noise. For  $f = 0.5$ , the friction angle is  $26.6^\circ$ ; thus, with a contact angle of  $25^\circ$ , the wheel is on the verge of sliding down and off the rail under the vertical force alone. For  $\mu = 0.3$ , a longitudinal force is applied in addition to the vertical load. Thus, the contact patch is in nearly complete slip because the corresponding  $\mu_R = 0.49$ , as discussed in the next paragraph. If some vibration exists, the rolling contact can fall into a stick-slip cycle, and the wheel can slide down. However, the sliding-down motion of the wheel is prevented by the constraint on the wheel axle in the  $y$  direction. The tendency to slide down, the constraint and the flexibility of the wheel and the cylindrical rail result in the high-frequency vibration. The quarter cylinder and its constraints on its boundary are designed to minimize such vibration.

The lateral creep force caused by the spin and the weight of the wheel and mass  $M_c$  is  $F_T = 0.19F_N$  for the contact angle  $\delta = 12.5^\circ$ ; thus, for  $\mu = 0$ , the resultant traction coefficient  $\mu_R$  is 0.19. For  $\mu = 0.3$ ,  $F_T = 0.18F_N$ ; then,  $\mu_R = 0.35$ . For  $\delta = 25^\circ$  and  $\mu = 0.3$ ,  $F_T = 0.32F_N$  and  $\mu_R = 0.49$ . This result represents nearly complete sliding of the wheel and demonstrates that the traction of a driven wheel with spin is the sum of the traction induced by spin and the tractive force in the rolling direction.

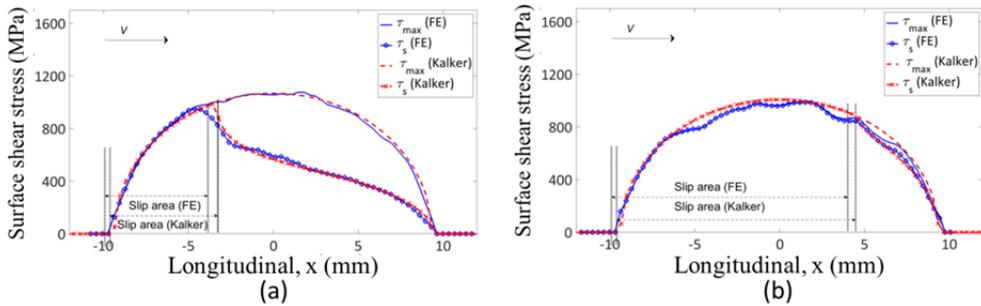


Fig. 3.6. The distribution of surface shear stress along the longitudinal axis ( $y' = 0$ ) for  $\mu = 0.3$ : (a) contact angle  $\delta = 12.5^\circ$ ; (b) contact angle  $\delta = 25^\circ$ .

3

### 3.4.2.2. Areas of slip, stick, and micro-slip

Another important feature of the frictional rolling contact is that the contact patch is divided into the slip area and the stick area. The corresponding areas of slip and stick are shown in Figs. 3.7 and 3.8 for the two cases of  $\mu = 0$  and  $\mu = 0.3$ , respectively, for  $\delta = 12.5^\circ$ . If the spin creepages were zero, areas of slip would vanish for  $\mu = 0$ . The non-vanishing slip areas in the present case are due to the geometric spin, as shown in Fig. 3.7. However, the spin is rather small, and thus, when the coefficient of friction is large and  $\mu = 0$ , the slip areas are only along the trailing edge where the magnitude of the surface stresses are rather small compared to the tolerance  $\varepsilon_R$ . These small surface stresses make the result more vulnerable to numerical errors and the effect of high-frequency vibrations. In the lower left portion of Fig. 3.7(a), the slip area is less regular compared to Kalker's solution in Fig. 3.7(b). However, for  $\mu = 0.3$ , the FE result looks better than that of Kalker's solution - the erratic nodes in slip along the leading edge of Fig. 3.8 do not exist in Fig. 3.8(a). The FE solution is also better than Kalker's solution along the leading edge of Fig. 3.7. The errors in the FE solution and Kalker's solution are trivial because the stress along the border of the contact patch where the error occurs is extremely small.

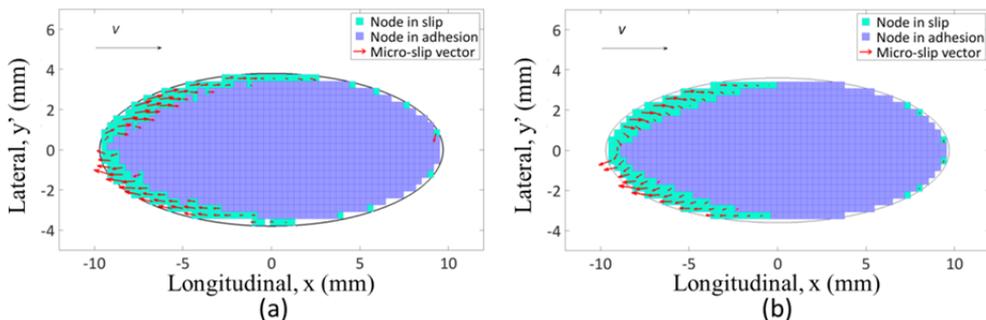


Fig. 3.7. The stick-slip area distribution and the corresponding micro-slip for  $\mu = 0$  ( $\delta = 12.5^\circ$ ): (a) FE solution; (b) Kalker's solution.

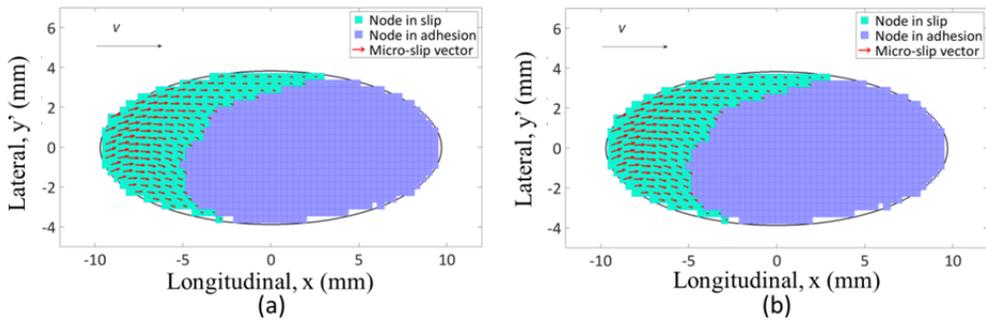


Fig. 3.8. The stick-slip area distribution and the corresponding micro-slip for  $\mu = 0.3$  ( $\delta = 12.5^\circ$ ): (a) FE solution; (b) Kalker's solution.

The stick areas obtained using the FE model are 85.6% and 64.5% of the contact patch areas for  $\mu = 0$  and  $\mu = 0.3$ , which is shown in Fig. 3.7(a) and Fig. 3.8(a), respectively, whereas the stick areas from Kalker's solution are 85.3% and 61.8% of the contact patch areas for  $\mu = 0$  and  $\mu = 0.3$ , respectively. The relative differences for these two cases are 0.3% and 3.7%, respectively, which are due to numerical errors as well as the omission of the half-space approximation in the FE solution.

The spin also changed the shape of the stick area from elliptical in the case of only longitudinal creepage without spin to an asymmetrical “pear”-like shape when spin occurs, as was observed by Johnson [30] (see Fig. 3.7 and Fig. 3.8).

The field of micro-slip for the contact angle  $\delta = 12.5^\circ$  for  $\mu = 0$  and  $\mu = 0.3$  is also shown in Fig. 3.7 and Fig. 3.8, respectively. The direction of the micro-slip is in the opposite direction of the surface shear stress. The directions and magnitudes of the micro-slip obtained with the FE solution are in good agreement with those from Kalker's theory. Taking the case of  $\mu = 0.3$  as an example, the good agreement in the magnitude of the micro-slip along the longitudinal axis is more clearly presented in Fig. 3.9. Only small differences exist at the trailing edge, where the stress is small and computational errors may dominate.

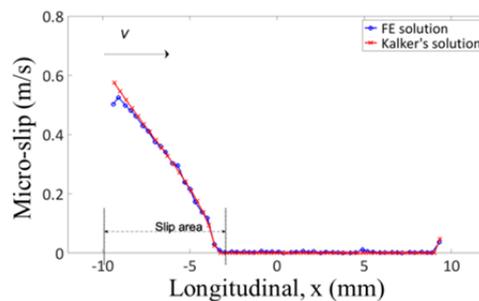


Fig. 3.9. The distribution of micro-slip (absolute values) along the longitudinal axis ( $y' = 0$ ) for  $\mu = 0.3$  ( $\delta = 12.5^\circ$ ).

### 3.5. Elasto-plastic solution

The bilinear elasto-plastic material model with kinematic hardening is applied to examine the effects of plastic deformation on spin-rolling contact. The results and a comparison with those of the elastic solution are presented and discussed below. Similar to the previous section, the contact angle  $\delta = 12.5^\circ$  was considered first. The contact angle  $\delta = 25^\circ$  was used for comparison when necessary.

#### 3.5.1. Solutions on the contact surface

In this section, the comparison of the elasto-plastic solution with the elastic solution on a contact surface was performed for  $\mu = 0.3$ .

##### 3.5.1.1. Von Mises stress of the surface layer

Fig. 3.10 shows that the maximum von Mises stress exceeds the yield stress (800 MPa) for the contact angle  $\delta = 12.5^\circ$ , and plastic deformation occurs. Compared with the results of the elastic solution, the maximum von Mises stress moves back toward the trailing edge when plastic deformation occurs. The reason for this trend is the shrinking of the slip area and the resulting backward movement of the maximal surface shear stress to be discussed below (see more details at the end of Section 3.5.1.3). The von Mises stress immediately outside of the contact patch is not zero because the elements in this region remain under stress. For elastic materials, the von Mises stress eventually dissipates when it is far from the contact patch. When the material is elasto-plastic, the von Mises stress behind the contact patch does not decay, as observed in Fig. 3.10(a) for  $x < -9.6$  mm. This remaining stress is the residual stress due to plastic deformation, and it is left behind in the rolling band after the wheel passage. This residual stress does not occur outside of the running band, and it cannot be found outside of the contact patch in the lateral direction, as shown in Fig. 3.10(b). The von Mises stress of the two materials corresponds well at places where no plastic deformation takes place.

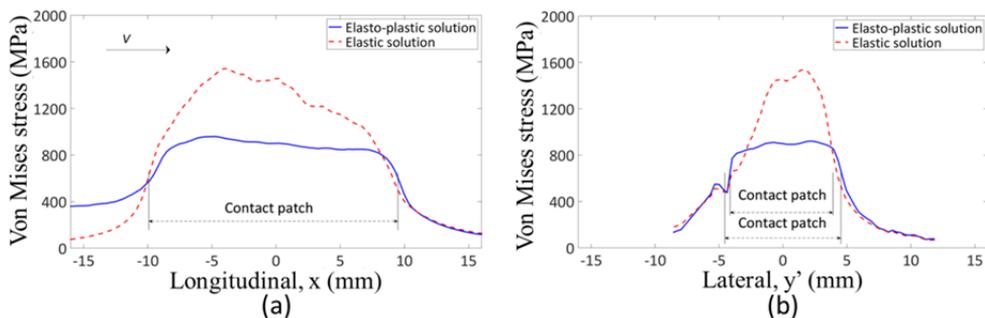


Fig. 3.10. The von Mises stress distributions for  $\mu = 0.3$  ( $\delta = 12.5^\circ$ ): (a) along the longitudinal axis ( $y' = 0$ ); (b) along the lateral axis ( $x = 0$ ).

In Fig. 3.10(a), the von Mises stress for each of the two materials is not symmetrical about the lateral axis ( $x = 0$ ). The maximum values are both on the trailing side of the contact patch. In the distribution along the lateral axis ( $x = 0$ ), the maximum values are both located along

the side where more slip occurs. The distribution of the von Mises stress is flatter in the contact patch when plastic deformation occurs. This distribution may lead to the shakedown of stress after a repeated wheel passage and is because the material is hardened and the contact geometry becomes more conformal.

### 3.5.1.2. Pressure

Fig. 3.11 compares the pressure distributions between the two cases. In the longitudinal direction (Fig. 3.11(a)), the pressure of the elasto-plastic solution coincides with the elastic solution in the frontal portion of the figure, whereas it falls below the elastic solution in the rear part. This result occurs because plastic deformation has occurred in the rear part; as a result, the contact patch is enlarged, and the pressure is reduced (see Fig. 3.11, 3.12(d), and 3.13). This leads to a loss of symmetry in the pressure distribution, as shown in Fig. 11(a) for the elasto-plastic case. The same phenomenon was observed and reported in [31].

This asymmetry in the pressure distribution also exists in the lateral direction (see Fig. 3.11(b)) due to the asymmetrical surface shear stress distribution (Fig. 3.5) and the asymmetrical von Mises stress (Fig. 3.10(b)). This asymmetry can be seen from the asymmetry of the contact patch for both the longitudinal ( $y' = 0$ ) and lateral ( $x = 0$ ) axes, as shown in Fig. 3.12(d) and more clearly in Fig. 3.13. The contact patch has changed from an ellipse to an “egg” shape. The contact patch becomes noticeably wider when plastic deformation occurs (see Fig. 3.11, 3.12 and 3.13). It is reasonable to expect that the lengths of the contact patch should be significantly different for the two materials. However, this difference is not apparent in Fig. 3.11(a) because with the given parameter set in this study, the difference is not sufficiently large to be distinguished from the numerical error resulting from the finite size of the elements. As discussed in Section 3.5.1.4 and shown in Fig. 3.14, the contact patch expands in all directions with an increasing traction coefficient. As a result, the area of the contact patch is  $116 \text{ mm}^2$  in the full elastic solution, whereas it becomes  $131 \text{ mm}^2$  in the elasto-plastic solution. The increased area of the contact patch leads to a more conformal contact geometry and reduces the maximum value of pressure from  $2,210 \text{ MPa}$  in the elastic solution to  $1,774 \text{ MPa}$  in the elasto-plastic solution, a reduction of 20%.

There is a dip in the von Mises stress for both cases at approximately  $y' = -4.5$  in Fig. 3.10(b). This dip occurs because this location is close to the spin pole (see Fig. 3.5(b)), around which the stress field changes rapidly.

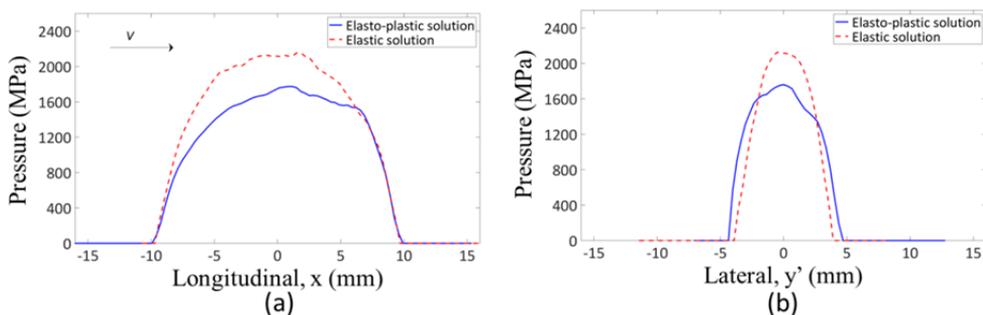


Fig. 3.11. The pressure distributions in the contact patch ( $\mu = 0.3$  and  $\delta = 12.5^\circ$ ): (a) along the longitudinal axis ( $y' = 0$ ); (b) along the lateral axis ( $x = 0$ ).

### 3.5.1.3. Surface shear stress and micro-slip

The tangential solutions for the contact surface for  $\delta = 12.5^\circ$  are shown in Fig. 3.12. First, the surface shear stress distribution in the elasto-plastic solution changes significantly in magnitude, as shown in Fig. 3.12(a) and 3.12(b). In both the longitudinal and lateral directions, the shear stresses of the elasto-plastic solution follow the same trend as that of the elastic solution. In the longitudinal direction, because the stick area becomes larger when plastic deformation occurs, it extends toward the trailing edge. As a result, point A in Fig. 3.12(a), which is located along the border between the stick area and slip area in the elastic solution, moves downward to point B along the limiting shear stress  $\tau_{max}$  (where  $\tau_{max} = fP_N$ ). This downward movement of point A, together with the reduced pressure, results in a considerably lower maximal shear stress than in the elastic case (30% lower). This reduced maximal shear stress may have a significant impact on the reduction in rolling contact fatigue and wear. In the lateral direction, the shear stress increases with an increasing  $y'$ , i.e., with an increasing slip trend (see Fig. 3.12(d)) until the limiting friction  $\tau_{max}$  is reached. The slip increases with  $y'$  because the spin pole is on the negative  $y'$  side that is outside of the contact patch (see Fig. 3.12(d) and 3.5(b)), and the tendency to slip increases with increasing distance from the spin pole.

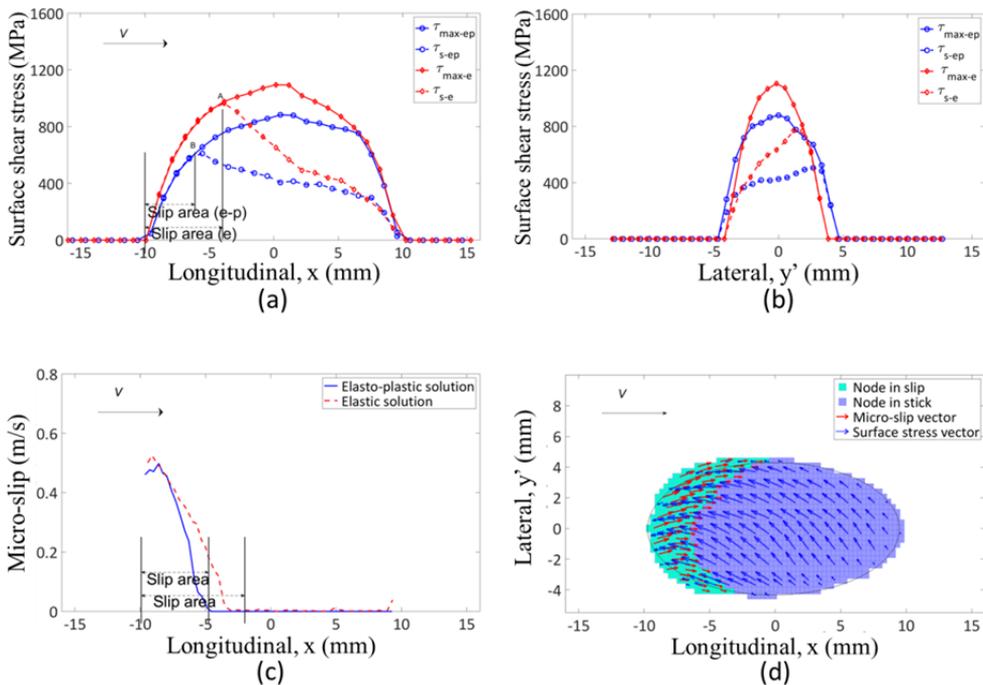


Fig. 3.12. Tangential solutions for  $\delta = 12.5^\circ$  ( $\mu = 0.3$ ): (a) shear stress distribution along the longitudinal axis ( $y' = 0$ ); (b) shear stress distribution along the lateral axis ( $x = 0$ ); (c) micro-slip distribution along the longitudinal axis ( $y' = 0$ ); (d) the stick-slip area distribution, surface shear stress and micro-slip, where the black oval indicates the location of the contact patch of an elastic solution for comparison.

The direction distribution of the surface shear stress remains largely unchanged, as shown by a comparison of Fig. 3.12(d) and 3.5, because it is predominantly determined by the geometric spin, which is the same in both cases.

The direction of micro-slip is nearly the same as in the elastic solution (see Fig. 3.12(d) and 3.8). Only the magnitude becomes smaller in the elasto-plastic solution, as shown in Fig. 3.12(c). Whereas the maximum value of the elastic solution is 0.53 m/s for 40 m/s of rolling speed, it is 0.5 m/s in the elasto-plastic case (a difference of 6%).

Compared with that of the elastic solution (see Fig. 3.8), the asymmetric shape of the stick area almost disappears in the elasto-plastic solution, which is shown in Fig. 3.12(d). When the contact angle is increased to  $25^\circ$ , the asymmetry of the stick area becomes more apparent again (see Fig. 3.13). For a contact angle of  $12.5^\circ$ , the stick area becomes larger (up to 76.8% of the contact patch area), and the shape of the stick area resembles an ellipse. The slip area shrinks because the material can now flow plastically with the surface shear stress, reducing the tendency toward relative motion between the shearing contact particles.

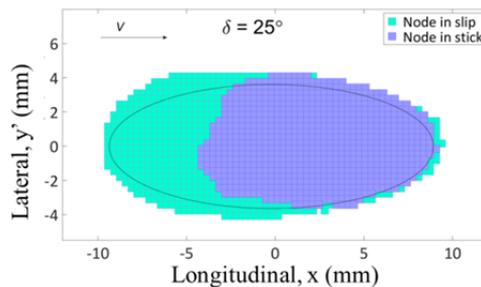


Fig. 3.13. The stick-slip area distribution for  $\delta = 25^\circ$  and  $\mu = 0.3$ , where the black oval indicates the location of the contact patch of an elastic solution for comparison.

#### 3.5.1.4. The influence of tangential load on the solution

In this section, different tractive cases are introduced by applying traction coefficients of  $\mu = 0, 0.15$  and  $0.3$  for a contact angle of  $\delta = 12.5^\circ$  to investigate the influence of tangential load on the pressure and von Mises stress, as shown in Fig. 3.14. The material is elasto-plastic in all cases.

The pressure distribution changes slightly with the variation in the traction coefficient. The maximum pressure decreases slightly with an increase in the traction coefficient, as shown in Fig. 3.14(a), due to the increased size of the contact patch with increases in the traction coefficient, resulting in plastic deformation. These changes can also be seen in the pressure distribution along the lateral axis ( $x = 0$ ) (see Fig. 3.14(b)). The results also show that the size of the contact patch varies slightly with the variation in the traction coefficient in both the longitudinal and lateral directions.

Fig. 3.14 also shows the von Mises stress distribution along the longitudinal ( $y' = 0$ ) and lateral ( $x = 0$ ) axes. The maximum von Mises stresses are 810, 900 and 980 MPa for  $\mu = 0, 0.15$  and  $0.3$ , respectively. Because the yield stress is 800 MPa, plastic deformation takes place in all three cases, and the material hardens more with a higher traction coefficient. The distribution of von Mises stress varies greatly with the traction coefficient. With an increasing

traction coefficient, the location of the maximum von Mises stress moves toward the trailing part, and the contact patch expands in both the longitudinal and lateral directions. In fact, the expansion of the contact patch caused by increasing the von Mises stress occurs in all directions.

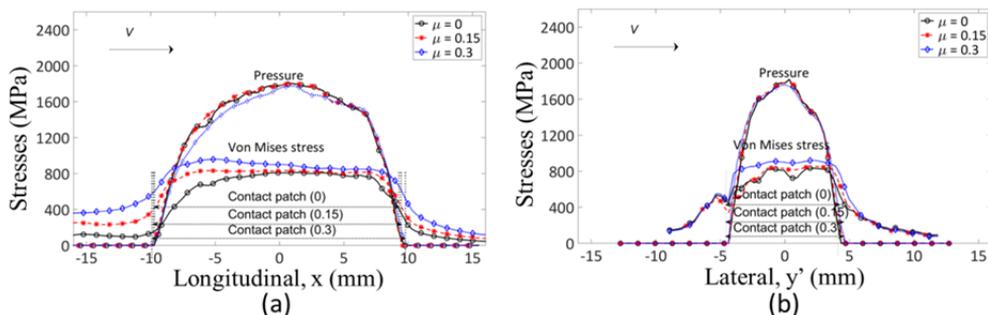


Fig. 3.14. The distributions of pressure and von Mises stress in the contact patch for different traction coefficients ( $\delta = 12.5^\circ$ ): (a) along the longitudinal axis ( $y' = 0$ ); (b) along the lateral axis ( $x = 0$ ).

Residual stress can be observed in Fig. 3.14(a) for  $x < -9.6$  mm for all three traction cases. A larger traction coefficient leads to a larger tangential contact force and thus a larger residual stress. In the lateral direction, for the same reason as discussed in section 3.5.1.3, the shear stress increases with increasing  $y'$  (see Fig. 3.14(b)). The stress fluctuates less at a higher  $\mu$  because more plastic deformation occurs with a larger tangential force, which dampens out more vibration in the bodies that were excited by the rolling wheel.

### 3.5.2. Stress history

In this section, the FE model is used to investigate the detailed stress history of elasto-plastic material for contact angles of  $12.5^\circ$  and  $25^\circ$  and for  $\mu = 0$ . The results are compared with those obtained by Bower and Johnson [32], and Kapoor [33].

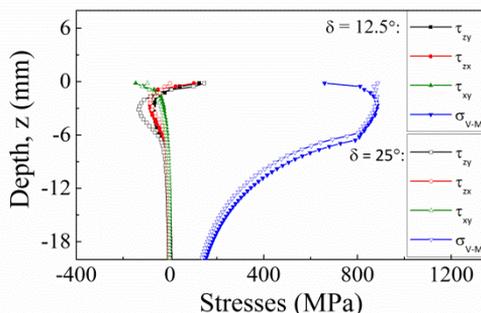


Fig. 3.15. Subsurface stress distributions with depth for contact angles of  $12.5^\circ$  and  $25^\circ$ .

Fig. 3.15 shows the orthogonal shear stress distributions along the depth at points in the middle of the contact patch at a fixed moment in time. The largest maximum orthogonal shear stresses components  $\tau_{zy}$  and  $\tau_{zx}$  for the two contact angles appear at the contacting surface, whereas the largest value of  $\tau_{xy}$  is not necessarily located at the surface. Therefore, the von Mises stress is induced by the combination of stress components, and the location of the largest von Mises stress defined as the critical point. This point is used to investigate the stress history. The critical point is located at subsurface  $z = 0.28a$  ( $a$  is the radius of the contact patch in the longitudinal direction) for  $\delta = 12.5^\circ$ , whereas it is located on the surface  $z = 0$  for  $\delta = 25^\circ$ .

Fig. 3.16(a) and (b) show the history of the stress components for the two cases at the critical point where a wheel is rolling across the surface. The critical point is located at the position  $x = 0$ . The wheel is rolling along the  $x$ -axis from the left to the right: all of the stress components are non-proportional. The direct stresses  $\sigma_{xx}$ ,  $\sigma_{yy}$  and  $\sigma_{zz}$  are completely compressive, and their values reach a maximum when the wheel is just above the critical point. Because it is a 3D scenario, the shear stresses  $\tau_{zy}$  and  $\tau_{xy}$  are close to zero under the surface for  $\delta = 12.5^\circ$ . Conversely, they have a relative large value for  $\delta = 25^\circ$ . The shear stress  $\tau_{zx}$  undergoes a complete reversal.

Similar work has been conducted by Bower and Johnson [32], and Kapoor et al. [33] in 2D under Hertzian pressure without spin and tangential loads, where the maximum orthogonal shear stress  $\tau_{zx}$  occurred at  $z = 0.5a$  in [32] and at  $z = 0.53a$  in [33]. The results presented herein are based on a 3D analysis with spin. The location of the maximum orthogonal shear stress  $\tau_{zx}$  is at  $z = 0$ . Wear and rolling contact fatigue are frequently related to large geometric spin, e.g., at the gauge corner of rails [5, 34] and in deep groove bearings [35]. Therefore, an accurate evaluation of stress and micro-slip using the presented method for 3D contact with a realistic treatment of spin can improve our understanding of fatigue damage mechanisms and provide an optimized design for wheel-rail systems and machine elements.

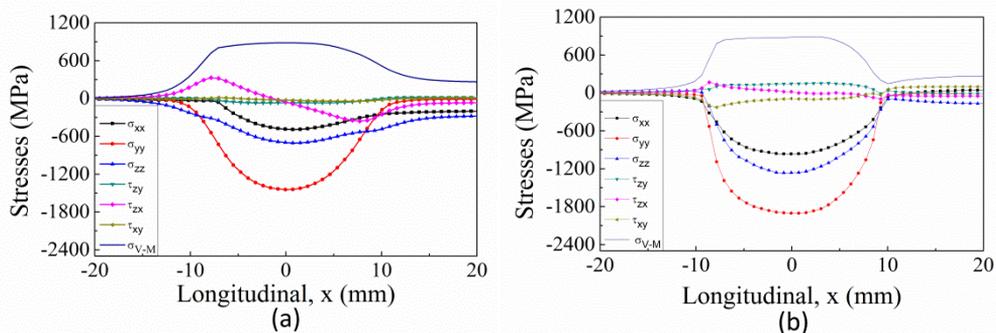


Fig. 3.16. Stress histories of the critical points under and on the surface: (a)  $\delta = 12.5^\circ$ ; (b)  $\delta = 25^\circ$ .

### 3.6. Discussions

The 3D explicit FE model presented in this chapter is compared with the Hertz theory and the Kalker's theory for rolling contact with spin. The FE model has many advantages over the classical methods. It can be used to evaluate the spin effects on the distribution of normal

pressure, the areas of stick and slip, the distribution of surface shear stress, the distribution of micro-slip, the distribution of von Mises stress and the pattern of subsurface plastic flow in elasto-plasticity. The elastic solution demonstrates that the explicit FEM yields more accurate results than the traditional solutions for rolling contact with spin. In the elasto-plastic analysis, the FEM yields more realistic results than those approaches (e.g. the Hertz theory and the Kalker's theory) which take into account only elasticity when plastic deformation occurs. This evaluation suggests that the explicit FEM can be used for rolling contact problems between wheels and rails at gauge corners and in bearings. The stress and strain histories at the contact surfaces and subsurface in the elasto-plastic material can serve as a guideline for future fatigue tests and assist in choosing the appropriate fatigue theory.

A quarter cylinder and a coned wheel are used in this chapter instead of an actual geometry of a wheel and rail to demonstrate the potential for solving spin-rolling contact problems using an explicit FEM. Realistic contact geometries can be readily treated with the presented method for not only wheel-rail contact but also rolling bearings *etc.*. In this model, to reach the quasi-static state, vibration is controlled by selecting a suitable contact angle and applying lateral constraints to the quarter cylinder.

Finally, a simple bilinear elasto-plastic model is used to demonstrate the ability of the model to incorporate non-linear material models, the need for employing such a model for a more realistic computation of stress and strain and the benefit of avoiding complexity that would be introduced by a more complicated material model. In future studies on shakedown, ratcheting and rolling contact fatigue, appropriate material models can be adopted in the FE model and multiple wheel passages loading condition should be considered. In addition, strain rate and thermal effects can be considered in the material models, as fatigue of steels may be related to the strain rate and temperature [34, 36]. Therefore, in further rolling contact fatigue investigations, the realistic material properties should be considered in the explicit FE method.

### 3.7. Conclusions

A 3D explicit FE method based on a Lagrangian formulation is presented to solve spin-rolling contact problems. An explicit time integration scheme and the penalty contact algorithm are used in the model. The accuracy of the numerical scheme has been demonstrated through a detailed analysis and a comparison with Kalker's solution. Moreover, the effect of plastic deformation is investigated by comparing the elasto-plastic solution with the elastic solution. The following conclusions can be drawn:

- 1) To date, accurate solutions for the spin-rolling problem have only been available by considering Kalker's work based on the BEM for half-space and a quasi-quarter space contact geometry in statics and linear elasticity. In this chapter, a new method based on the FEM is presented to address systems with an arbitrary geometry and continuum dynamics, in which large geometric nonlinearity and non-linear material properties, such as plasticity, can be considered. A quasi-static solution can also be obtained with dynamic relaxation.
- 2) The FE model is verified against the Hertz theory and Kalker's theory in a quasi-static state. The normal solution corresponds well with the Hertz theory, and the normal and tangential solutions correspond well with Kalker's solution.
- 3) Some fluctuations in stress and micro-slip are observed in the FE solutions. These fluctuations occur because the moving wheel excites the physical vibration of the system

and the system is insufficiently relaxed to a steady state. Such high-frequency vibrations can be dampened out to a certain degree by plastic deformation.

- 4) In the frictional rolling contact problems with large geometric spin, the effect of the tangential contact stress on the normal solution is negligible in the elastic solution. However, the effect may need to be considered in the elasto-plastic solution.
- 5) Spin changes the shape of the stick area from a symmetrical ellipse to an asymmetric shape. Spin can increase the resultant traction coefficient. In addition, the position of the spin pole varies with variations in the traction coefficient.
- 6) Plasticity has a strong effect on the size and shape of the contact patch, areas of stick and slip, and magnitude of the pressure, shear stress and von Mises stress. Plasticity can lead to the shakedown of stress after repeated wheel passages. It can also make the contact geometry more conformal. An analysis of a rolling contact with plasticity can provide more realistic insights.
- 7) When spin is considered, the stress history of the critical point is different from that of Hertz contact in two dimensions.

## References

- [1] K.L. Johnson, *Contact mechanics*, Cambridge university press, Cambridge, 1987.
- [2] K.L. Johnson, The effect of spin upon the rolling motion of an elastic sphere on a plane, *Journal of Applied Mechanics*, 25 (1958) 332-338.
- [3] J.J. Kalker, *Three-dimensional elastic bodies in rolling contact*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [4] Z. Li, J.J. Kalker, Simulation of severe wheel-rail wear, *Proceedings of the Sixth International Conference on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems*, Southampton, (1998) 393-402.
- [5] R. Dollevoet, Z. Li, O. Arias-Cuevas, A method for the prediction of head checking initiation location and orientation under operational loading conditions, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 224 (2010) 369-374.
- [6] N. Burgelman, Z. Li, R. Dollevoet, A new rolling contact method applied to conformal contact and the train-turnout interaction, *Wear*, 321 (2014) 94-105.
- [7] G. Carbone, C. Putignano, A novel methodology to predict sliding and rolling friction of viscoelastic materials: Theory and experiments, *Journal of the Mechanics and Physics of Solids*, 61 (2013) 1822-1834.
- [8] T. Chaise, D. Nélias, Contact Pressure and Residual Strain in 3D Elasto-Plastic Rolling Contact for a Circular or Elliptical Point Contact, *Journal of Tribology*, 133 (2011) 041402.041401-041402.041409.
- [9] J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure—I. Theory, *Computers & structures*, 27 (1987) 249-257.
- [10] N. Yukio, J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure—III. Impact/contact simulations, *Computers & structures*, 27 (1987) 275-286.
- [11] J. Oden, T. Lin, On the general rolling contact problem for finite deformations of a viscoelastic cylinder, *Computer Methods in Applied Mechanics and Engineering*, 57 (1986) 297-367.
- [12] U. Nackenhorst, The ALE-formulation of bodies in rolling contact: theoretical foundations and finite element approach, *Computer Methods in Applied Mechanics and Engineering*, 193 (2004) 4299-4322.
- [13] M. Ziefle, U. Nackenhorst, Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact, *Computational Mechanics*, 42 (2008) 337-356.
- [14] Z. Li, X. Zhao, R. Dollevoet, M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, *Vehicle System Dynamics*, 46 (2008) 237-246.
- [15] X. Zhao, Z. Li, The solution of frictional wheel-rail rolling contact with a 3D transient finite element model: Validation and error analysis, *Wear*, 271 (2011) 444-452.

- [16] Z. Li, R. Dollevoet, M. Molodova, X. Zhao, Squat growth—Some observations and the validation of numerical predictions, *Wear*, 271 (2011) 148-157.
- [17] M. Pletz, W. Daves, H. Ossberger, A wheel set/crossing model regarding impact, sliding and deformation—Explicit finite element approach, *Wear*, 294–295 (2012) 446-456.
- [18] X. Zhao, Z. Li, J. Liu, Wheel-rail impact and the dynamic forces at discrete supports of rails in the presence of singular rail surface defects, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226 (2011) 124-139.
- [19] T. Belytschko, W.K. Liu, B. Moran, K. Elkhodary, *Nonlinear finite elements for continua and structures*, John Wiley & Sons, Chichester, New York, 2000.
- [20] Z.-H. Zhong, *Finite element procedures for contact-impact problems*, Oxford University Press, Oxford, London, 1993.
- [21] K. Bath, *Finite element procedures in engineering analysis*, Pentice-Hall, Englewood Cliffs, NJ, 1982.
- [22] W. Johnson, *Impact strength of materials*, Edward Arnold, London, 1983.
- [23] R. Courant, K. Friedrichs, H. Lewy, On the partial difference equations of mathematical physics, *IBM Journal of Research and Development*, 11 (1967) 215-234.
- [24] J. Hallquist, G. Goudreau, D. Benson, Sliding interfaces with contact-impact in large-scale Lagrangian computations, *Computer Methods in Applied Mechanics and Engineering*, 51 (1985) 107-137.
- [25] D.J. Benson, J.O. Hallquist, A single surface contact algorithm for the post-buckling analysis of shell structures, *Computer Methods in Applied Mechanics and Engineering*, 78 (1990) 141-163.
- [26] L. Taylor, D. Flanagan, PRONTO 3D: A three-dimensional transient solid dynamics program, Sandia National Labs., Albuquerque, NM (USA), 1989.
- [27] J.O. Hallquist, LS-DYNA theory manual, Livermore software Technology corporation, (2006).
- [28] H. Hertz, Über die Berührung fester elastischer Körper, *Journal für die Reine und Angewandte Mathematik*, 92 (1882) 156-171.
- [29] W. Wernitz, Friction at Hertzian contact with combined roll and twist, *Rolling Contact Phenomena*, Bidwell, J. B., ed., Elsevier, New York, (1962) 132-1567.
- [30] K.L. Johnson, Tangential traction and microslip in rolling contact, J.B. Bidwell (Ed.), *Rolling Contact Phenomena*, Elsevier, Amsterdam, 0 (1962) 6-28.
- [31] J. Rońda, O. Mahrenholtz, R. Bogacz, M. Brzozowski, The rolling contact problem for an elastic-plastic strip and a rigid roller, *Mechanics Research Communications*, 13 (1986) 119-132.
- [32] A.F. Bower, K.L. Johnson, The influence of strain hardening on cumulative plastic deformation in rolling and sliding line contact, *Journal of the Mechanics and Physics of Solids*, 37 (1989) 471-493.
- [33] A. Kapoor, G. Morales-Espejel, A. Olver, A shakedown analysis of simple spur gears, *Tribology transactions*, 45 (2002) 103-109.

- [34] A. Ekberg, E. Kabo, Fatigue of railway wheels and rails under rolling contact and thermal loading—an overview, *Wear*, 258 (2005) 1288-1300.
- [35] L. Hua, S. Deng, X. Han, S. Huang, Effect of material defects on crack initiation under rolling contact fatigue in a bearing ring, *Tribology International*, 66 (2013) 315-323.
- [36] J. Ahlström, B. Karlsson, Fatigue behaviour of rail steel—a comparison between strain and stress controlled loading, *Wear*, 258 (2005) 1187-1193.

# 4

## INVESTIGATION OF THE FORMATION OF CORRUGATION-INDUCED RAIL SQUATS BASED ON EXTENSIVE FIELD MONITORING

*Rail squats originate from a number of sources, such as corrugations, indentations and welds. In this chapter, a five-year continual field monitoring study was performed on squats induced by corrugations. This study indicated that a small black depression formed at the corrugation under wheel-rail dynamic forces, and then, a primary crack typically initiated on the gauge side edge of the depression. Subsequently, the crack began to propagate in the rail surface in a U shape toward the gauge side in both the traffic direction and the opposite-traffic direction and into the rail toward the field side at an angle of approximately 20°. Rail inclination could influence the crack initiation location and propagation path. The geometry of the black squat depression was initially elliptical, and then, its edge followed the U-shaped cracking path as it grew. The squats turned into a kidney-like shape, typically with a U-shaped crack. Tensile stress likely led to the squat crack initiation and propagation. This cracking phenomenon and mechanism are analogous to the ring/cone crack formation of brittle materials under sphere-sliding contact. As the squats grew further, a ridge formed in the middle part of the depression, and an I-shaped crack appeared at this ridge due to the impact of the wheels. This process eventually led to two-lung-shaped mature squats, typically with a Y-shaped crack. The findings of this chapter provide insight into the formation of rail squats.*

---

This chapter is based on the following publication:

X. Deng, Z. Qian, Z. Li, R. Dollevoet. Investigation of the formation of corrugation-induced rail squats based on extensive field monitoring. *International Journal of Fatigue* 2018;112:94-105.

## 4.1. Introduction

Rail squats are one of the main types of rolling contact fatigue (RCF) defects [1]. Squats were reported as black spots dating back to the early 1960s [2]. They are typically observed in the crown of the railhead in straight tracks. A typical mature squat is characterized by a localized, dark depression in a two-lung-like shape, with cracks in the rail surface and under the surface [3, 4] (Fig. 4.1). Further development of cracks leads to rail breakage and thus threatens the safety of rail traffic. In Europe, squats are currently the most considerable RCF threat to rails, and they increase the cost of rail maintenance dramatically. A good understanding of the root causes and formation mechanism of squats contributes to the prevention of such defects from their undesired consequences and to the reduction of the cost of maintenance.

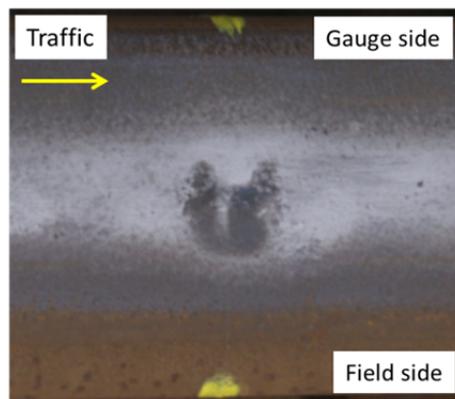


Fig. 4.1. A typical squat in a two-lung shape with cracks at the monitored site.

Several comprehensive studies of rail squats have been performed over the past few decades. Clayton and Allery [1] conducted a pioneering metallurgical analysis of mature squats and found that cracks initiate due to shear stresses and propagate into the rail. The development of these cracks coincides with the formation of black spots in the rail running surface. Pal et al. [5, 6] also performed some metallurgical observations and pointed out that the crack initiation in the rail surface was related to microstructural changes and the consequential white etching layer (WEL) and ratcheting. Grassie et al. [7] argued that the WEL is not a necessary condition for squat initiation, and squat-like defects associated with the WEL could be classified as studs. Simon et al. [8] presented a tribological characterization of squats, considering plastic deformation, surface microstructures and crack fronts. Bogdanski et al. [9, 10] investigated the fracture phenomenon and liquid entrapment mechanism for squat cracks by calculating stress intensity factors using linear elastic fracture mechanics. Farjoo et al. [11] conducted a similar study, further considering the resulting bending effects of the elastic foundation on crack propagation through stress intensity factors. Steenbergen [12] studied squat formation, in which RCF produced a leading crack and developed via trailing brittle failure cracks into the squats. These works concerned mature squats with large cracks. Satisfactory mechanisms of squat crack initiation and development were still missing.

Li et al. [3, 4, 13, 14] conducted a series of investigations on squat initiation and growth. In these investigations, the initiation sources were classified into two types of rail surface irregularities, i.e., the active type and the passive type. The active type includes short pitch corrugation and rail welds. The common characteristic of this type is that the irregularities and the resulting squats arise internally and spontaneously from the consequence of imperfect design, construction or maintenance of tracks. The irregularities at corrugation are resulted from dynamic forces excited by corrugation, and the irregularities at welds are developed from differential wear and differential plastic deformation because of material inhomogeneity. This type of irregularities initiate and grow by themselves. The passive type includes indentations and wheel burns, the latter is also called rail burns in [13]. In contrast to the active type, the passive type are generated by external objects such as bearing rollers/balls and spinning wheels (wheel slip) [4]. In addition to the afore-mentioned active and passive types, further observations have shown a third type of initiation sources, i.e., irregularities by design, such as insulation joints [15] and crossings [16]; both the gaps at the joint and the crossing cause battering of the rail by passing wheels, thus squats.

In spite of the three types of very distinctive initiation sources, mature squats bear a typical common appearance of two-lung shape with typical U (or V) and Y shaped cracks [4]. Numerical analysis and field observations [3, 4] showed that this two-lung shape is the result of the dynamic wheel-rail contact force induced by the irregularities. It was also found that the irregularities and their resulting squats can induce corrugation-like wave patterns. These patterns occurred immediately after squats. Both the appearance and wavelengths of them are similar to those of short pitch corrugation in the Dutch railway network. A numerical approach for determining the critical size for squats to initiate from the passive type irregularities was presented in [13]. The above numerical works [3, 4, 13, 14] assumed that cracks were not deep enough to significantly affect the relevant wavelengths, therefore, cracks were not considered in the models. Li et al. were mainly interested in the primary and root causes of the two-lung shape of squats when cracks are in their initiation and early stages of growth, the assumption that cracks were not deep is therefore reasonable.

A clear distinction was made between the corrugation-induced squats and the squat-induced corrugation-like wave patterns in [4]. In principle, all the afore-mentioned initiation sources but the corrugations could cause squat-induced corrugation-like wave patterns. A field survey reported in [4] showed that among all the squats 33% were caused by corrugation; among the other 67% of the squats, 61% had wave pattern following them, and the other 39% not. Some authors call the squat-induced corrugation-like wave patterns also corrugation [17].

By considering dynamic effects, recently, Andersson et al. [18] investigated the propensity of squat initiation by examining the effects of surface irregularities on local stresses using 2D numerical models. Experimental study or field observation was not presented in this work. Crack growth and geometry evolution of squats were also not included.

Cracks, however, are the most dangerous aspect of squats. As corrugation-induced squats consist a major part of the squats population, and the initiation and early growth of the cracks have not yet been investigated. Therefore, this chapter focuses on this type of squats. The objective is to reveal the formation of corrugation-induced squats and the accompanying crack initiation and propagation processes based on the observations and analyses of a five-year continual field monitoring. The focus is on primary virgin cracks, in other words, the origins of the squats and the accompanying cracks were not residual cracks of previous damages. This

study contributes to a better understanding of the root causes of rail squats and RCF-related phenomena.

## 4.2. Description of the field monitoring

A straight section of the Dutch railway network was selected for a five-year continual field monitoring of squat initiation and evolution. The monitored track was located in Steenwijk in the Netherlands, as shown in Fig. 4.2. During the period from 2007 to 2012, ten field observations were conducted at intervals of approximately 6 months. Two times of grinding were performed on the monitored track during the period of the monitoring. The first one was performed between the fifth monitoring observation (November, 2009) and the sixth observation (June, 2010) with a material removal of 1.1 mm (including natural wear by traffic because these 1.1 mm were derived from the profiles measured during the fifth and sixth observations); the second one was performed between the seventh observation (November, 2010) and the eighth observation (June, 2011) with a material removal of 0.3 mm (including natural wear). There was no grinding before the start of the monitoring. One-directional traffic ran on the track at a frequency of approximately 20 minutes per train during the day and no trains during the night. The majority of the traffic consisted of passenger trains. The maximum axle load of the trains was approximately 22 tons [19], and the annual traffic load was about

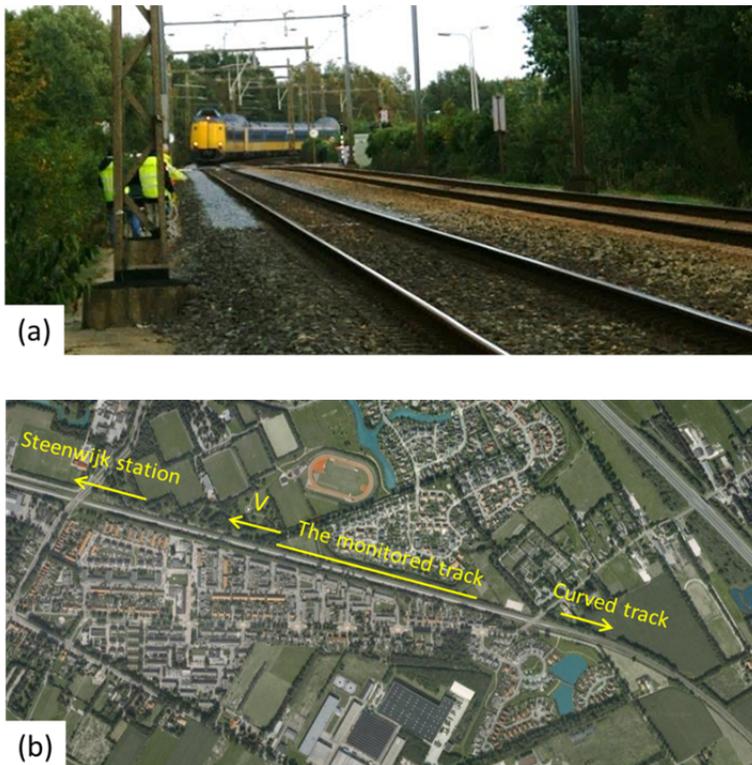


Fig. 4.2. Overview of the monitored straight track: (a) Monitored track looking in the direction of the curve against the traffic; (b) Google satellite view of the site;

3.4 MGT. When trains were running on the straight section, they had just exited a curved track. The speed of the trains was 130 km/h and some traction was needed to maintain this speed. The rails were of the R260Mn grade with UIC54E1 as the nominal profile with a weight of 54 kg/m. They were installed in the track in 1989.

The section of track was selected for monitoring because of the following reasons. First, some corrugation and resulting squats of various severities were observed at the beginning of the monitoring, as shown in Fig. 4.3(a). In view of the active nature of corrugation-induced squats, it was predictable where new squats would initiate and grow from the developing corrugation. This makes a fruitful selection of monitoring locations possible. The second reason is that there were no visible defects or residual cracks in the rail surface other than the intended corrugation and squats. The third reason is that the rails had not been ground. To prove that any grinding was not performed before the start of the monitoring, the rail profiles measured during the monitoring at a number of defect-free locations were analyzed. It was found that the average rail head loss due to the natural wear was 0.07 mm/year between the first and fifth observations, and the average rail head height loss was 0.06 mm/year between 1989 and 2007. The height loss of 0.06 mm/year would have not been possible if any corrective grinding had taken place. This was further confirmed by looking at the measured profiles, as shown in Fig. 4.3(c) and (d), where the measured profiles are aligned with the nominal profile (in black) on the field side (the left side). The measurements show good repeatability. The profile (in magenta) measured immediately after a corrective grinding at the sixth observation is as smooth as the other profiles on the gauge side. In contrast, the ground profile has a larger rail head height loss (see Fig. 4.3(d)) and a sharp turning point at the field side (indicated in Fig. 4.3(c)), apparently it is different from other profiles. If there was any earlier grinding, such a larger height loss and sharp turning point should be recognizable in the first measurement.

During the monitoring, more than 100 squats of various severities were observed. These squats initiated and developed from short pitch corrugations because they occurred in the middle of the corrugation [4]. Several examples of squats at the first and fifth observations are shown in Fig. 4.3(a) and (b). Among the ten field observations, the first five observations were performed before the first rail grinding. Dozens of selected squats were photographed during each observation. The cross-sectional profile and vertical-longitudinal profile at the squats were measured using the MiniProf and RailProf devices, respectively. Non-destructive tests were conducted to detect cracks under the rail surface. These tests included eddy current tests using an Elotest D 300, which can capture shallow cracks under the rail surface, and ultrasonic tests using a Krautkramer USM 25, which is able to detect deeper cracks [4].

By the fifth observation (shortly before the rail grinding), 30 squats were mainly traced. Among them, 9 squats were light, 17 were moderate, and 4 were severe. The analysis in this chapter is mainly based on these 30 squats, with consideration of supporting evidence from the remaining squats and from several of the other monitoring observations at this monitored site. Eleven of the 30 squats are shown in Figs. 4.1, 4.3, 4.4-4.7, 4.9, 4.10, 4.14, 4.15 and 4.17 to provide as much direct information as possible.



(a)



(b)

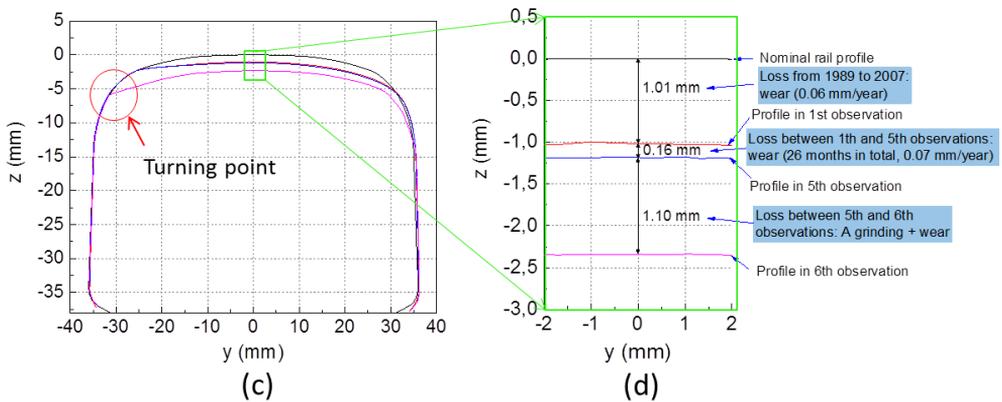


Fig. 4.3. Typical corrugation with several squats of various severities: (a) in the first observation, corrugation and some squats already existed; (b) in the fifth observation, the corrugation and squats in the first observation had become severe and new squats of different severity appeared; (c) measured rail profiles by MiniProf in the vicinity of one squat: from top to down they were the nominal (black) and those measured in the first (red), fifth (blue) and sixth (magenta) observations, respectively; (d) zoom-in of (c) in the middle of the profiles.

### 4.3. Field observations and analysis

The initiation and evolution processes of the squats are postulated based on a continual field monitoring and analysis. The following aspects are considered in this investigation: the geometry of the squats (size and shape) and the accompanying cracking phenomena. A squat development process is proposed based on the observations and analyses.

#### 4.3.1. Squat initiation and cracking processes observed over time

##### 4.3.1.1. Critical crack initiation size and point in the lateral direction and U-shaped cracks

The series of photographs in Fig. 4.4, which were obtained during the first six field observations, show the evolution of a squat. The changes in size and in the bright running band near the squat can be seen. At the time of the first observation, the squat was a small black depression in an approximately elliptical shape with a size of approximately 5 mm in the lateral direction and 6 mm in the traffic direction. The position of T1 is approximately 4 mm from the rail centerline. This black depression did not have visible surface cracks. It was growing in size until the fifth observation, shortly before the rail grinding. The squat eventually evolved into a kidney-like shape with its concave part resembling a shallow U, as seen in Fig. 4.4(d) and (e). As will be shown later, this shallow U developed into a full U that was the primary surface crack of the squats. Thus, this concave part is referred to hereafter as U-shaped. At the time of the fifth observation, the running band at the squat was widened on the gauge side, as indicated by the shining part above point T1, indicating that the depression broadened the running band.

Although it was hard to confirm the presence of surface macrocracks by visual inspection during the first five observations, eddy current tests showed that a crack began to appear as of the second observation. At this time, T1 of the depression was approximately 6 mm from the rail centerline, and the dimension of the depression was approximately 8 mm in the lateral direction and 9.5 mm in the traffic direction. These results indicate that cracks initiated when the black depression grew to a certain size. This finding was confirmed by the observations of another 5 squats of the 30 selected squats, which did not have macrocracks in the beginning, but cracks were observed at the sixth observation. This finding is in line with the critical size of squat initiation that was numerically derived in [13] by comparing the maximal von Mises stress with tensile strength, and the material was also R260Mn. After the crack had appeared, the shape of the depression began to change into a kidney-like shape, and the location of T1 remained approximately 6 mm from the rail centerline. Eddy current tests and ultrasonic tests showed that the crack continued to grow after their initiation. The phenomenon of crack growth with the fixed location of T1 relative to the rail center was observed with all the other monitored squats.

After the rail grinding to a depth of 1.1 mm, the black depression was removed, and a shallow U-shaped crack was exposed, as shown in Fig. 4.4(f). This observation shows that the crack had propagated into the rail deeper than 1.1 mm and that the depth of the black depression was less than 1.1 mm. This finding confirmed the presence of the cracks measured by the eddy current tests and ultrasonic tests. The removal of the black depression and the exposure of the residual cracks were also observed at the other monitored squats with cracks deeper than 1.1 mm.

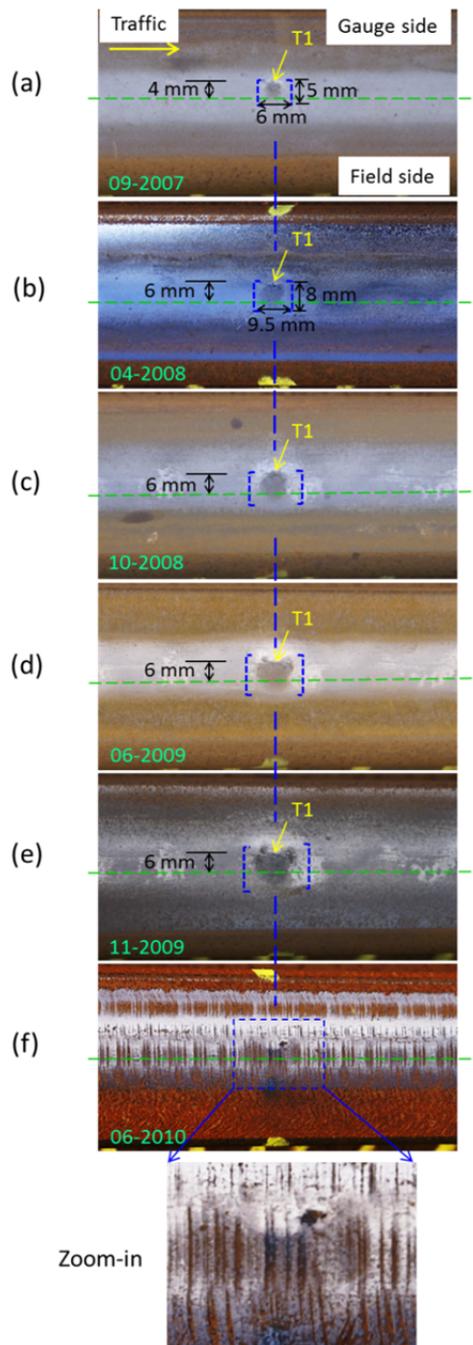


Fig. 4.4. Evolution of a squat with the critical size of 8 mm for crack initiation and subsequent cracking (the dashed straight green line represents the rail centerline): (a)-(e) Before rail grinding to a depth of 1.1 mm; (f) after the rail grinding, revealing a shallow U-shaped crack, as seen in the zoom-in.

#### 4.3.1.2. U-shaped crack initiation position in the rolling direction over the complete squat development process

Fig. 4.5 illustrates the development process with another squat. The figure consists of six photographs from the first five consecutive observations before the rail grinding and from the seventh observation after the grinding; the photo from the sixth observation was not sufficiently clear. These photographs show that T1 was the bottom point of the U shape and that its position stayed fixed not only in the lateral direction, as demonstrated with Fig. 4.4, but also in the traffic direction. The latter concluded as follows: the rail head has a constant width of 72.2 mm. With this value, the sizes of the depressions and cracks were derived from the photos for this thesis. In Fig. 4.5, there is, beside the squat, another small black spot. With the known rail head width, the distance between the small black spot and point T1 is determined, and it can be seen in Fig. 4.5 that T1 remained at a constant distance from the small black spot.

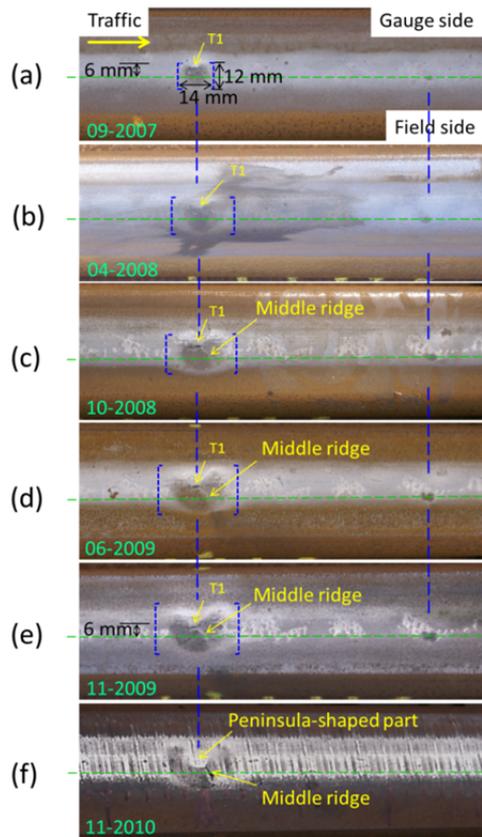


Fig. 4.5. Confirmation of the crack initiation point according to the relative positions of the black spot and point T1 (the green line represents the rail centerline): (a)-(e) Before rail grinding; (f) after rail grinding to a depth of 1.1 mm.

With T1 aligned according to the blue dashed lines, Fig. 4.5 shows that the squat grew longitudinally in both the traffic and opposite-traffic directions as well as laterally toward the gauge side and field side simultaneously. This growth pattern was observed at all 21 of the moderate and severe squats. The squat in Fig. 4.5 first grew into a kidney-like shape with the concave part of the kidney bounded by a U-shaped crack, and then it gradually evolved into a two-lung shape with a middle ridge (Fig. 4.5(c)-(f)). After the grinding, the part bounded by the full U-shaped crack took on a peninsula-like shape, as seen in Fig. 4.5(f). A more detailed discussion about the formation of the middle ridge and the peninsula-shaped part is in Sections 4.3.4.1 and 4.3.4.2. Notably, the typical two-lung shape and the peninsula-like shape were not clearly observable at the squat in Fig. 4.4 because the squat was still in a relatively early stage; the crack and geometry had not yet grown into the typical shape.

The very small black spot in Fig. 4.5(a), barely visible to the naked eye, was growing before the grinding. At the fifth observation, it turned into an obvious surface depression; its size was already comparable to that of Fig. 4.4(a), and it should have been able to develop into a squat if not removed. The initiation and growth of black spots are, in the very beginning, driven by the corrugation. At the same time, the black spots are rail surface irregularities that excite (additional) dynamic contact force; the larger the black spot, the larger the dynamic force, creating a positive feedback loop that promotes the continuous and accelerates growth of the black spots into squats. Combining these observations, a complete picture is obtained of how a mature squat is developed from the smallest visible black spot. The longitudinal position of the crack initiation point should be close to the location of the very first visible black spot.

### 4.3.2. Surface and subsurface cracks at squats of different severities

To further show the growth process of the squats from early stages to mature stages, five squats of different severities are presented in Fig. 4.6. The top five photographs were obtained at the fifth observation before the rail grinding, and first four of the bottom five were obtained at the sixth observation after the grinding and the last one was obtained from the seventh observation after the grinding. The photo of the last one from the sixth observation was not sufficiently clear. The photographs are shown from left to right with increasing severity; therefore, they can be interpreted as a representation of five different steps in a squat over its life cycle.

In (a), the black depression was still significantly smaller than the critical size of 8 mm; no crack was observed in the rail surface either before or after the rail grinding, as seen in Fig. 4.6(a). It is reasonable to assume that the macrocrack had not yet initiated at this time. In (b), the size of the depression (10 mm (in the lateral direction) x 14 mm (in the traffic direction)) was larger than the critical size of crack initiation; a small surface primary crack became visible at this stage prior to the rail grinding (see the arrow in Fig. 4.6(b)). After the grinding, the crack was removed. This observation suggested that the crack was shallower than 1.1 mm. In (c) and (d), the squats were in a kidney-like shape before the grinding, and a U-shaped crack became visible after the grinding. The length of the U-shaped crack continued to grow from the bottom along the U in both directions, i.e., in the traffic and the opposite-traffic directions, with the evolution of the squat. These observations confirm that the bottom of the U-shaped cracks was the initiation point of the cracks in or near the rail surface on the gauge side of the depression. The point of the crack initiation corresponded to position T1 in Figs. 4.4 and 4.5. After initiation, the crack propagated along the U shape, as well as into the rail toward the field side, as discussed in Sections 4.3.3 and 4.3.4. Moreover, the crack grew more

rapidly in the traffic direction than in the opposite-traffic direction; see the discussion in Section 4.4.3. In (e), the squats reached a more severe stage: they evolved into the two-lung shape with a middle ridge, and an I-shaped crack appeared at this ridge after the grinding. More details about the I-shaped crack are provided in Section 4.3.4.

From the above analysis, it can be concluded that the crack initiation point is in the rail surface or at least in the top layer of material not deeper than 1.1 mm from the surface.

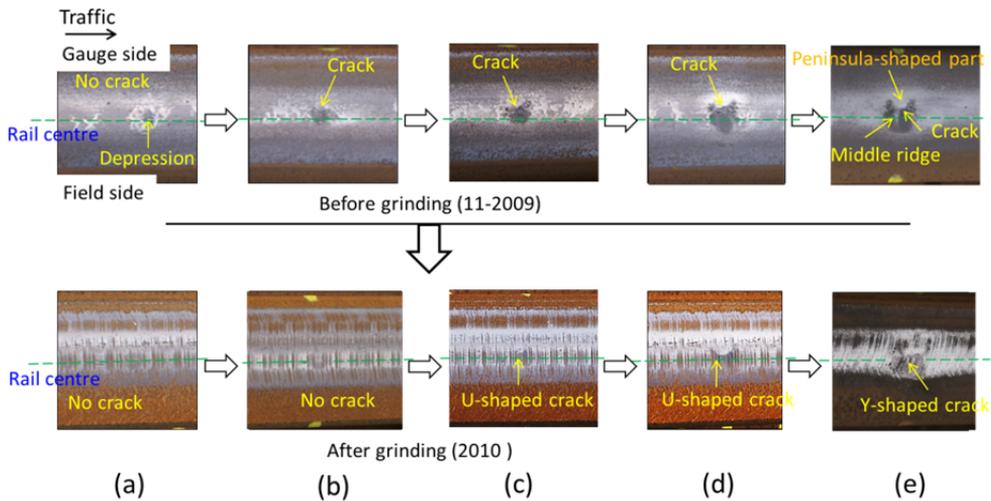


Fig. 4.6. Different squats at various severities: earlier stages to later stages.

Although the shape of the black depression at the very beginning was not very clearly visible, as shown in Fig. 4.6(a), the convex part of the kidney-shaped depression was elliptical, as seen in Fig. 4.6(b)-(d), and this elliptical shape most likely originated from the elliptical contact area between the wheel tread and the rail crown. The elliptical shape is more clearly visible in Fig. 4.4, especially in Fig. 4.4(a) and (b). Therefore, the small black depression in (a) should have been approximately elliptical in shape.

### 4.3.3. Cracking angle

A moderate squat with a U-shaped surface crack was examined to determine its cracking angle in the vertical-lateral cross-section, which passes through the bottom of the U shape, as shown in Fig. 4.7. In Fig. 4.7(a), bottom point T1 of the crack was the crack initiation point, and it was located on the gauge side at a distance of 6 mm from the rail centerline. After the rail grinding to a depth of 1.1 mm, the bottom of U-shaped crack T2 shifted 3 mm toward the field side, as shown in Fig. 4.7(b). The cracking angle  $\alpha$  was then approximately  $20^\circ$ , as illustrated in Fig. 4.7(c), indicating that the surface crack propagated into the rail at an angle of approximately  $20^\circ$  toward the field side. This finding agrees with the findings in the computed tomography investigation in [20].

For verification purposes, another 18 moderate and severe squats with cracks on the gauge side were also examined. All the crack initiation locations were approximately 6 mm from the rail centerline on the gauge side. These locations were also the bottom points of the U-shaped

cracks of the squats, and they corresponded to point T1. After the rail grinding to a depth of 1.1 mm, the bottom points moved 3 mm toward the field side. Thus, the cracking angles were again approximately  $20^\circ$  for all the examined squats, 7 of which are shown in Figs. 4.1, 4.4, 4.5, 4.6, 4.7 and 4.15. All the results were in good agreement with the cracking angle shown in Fig. 4.7.

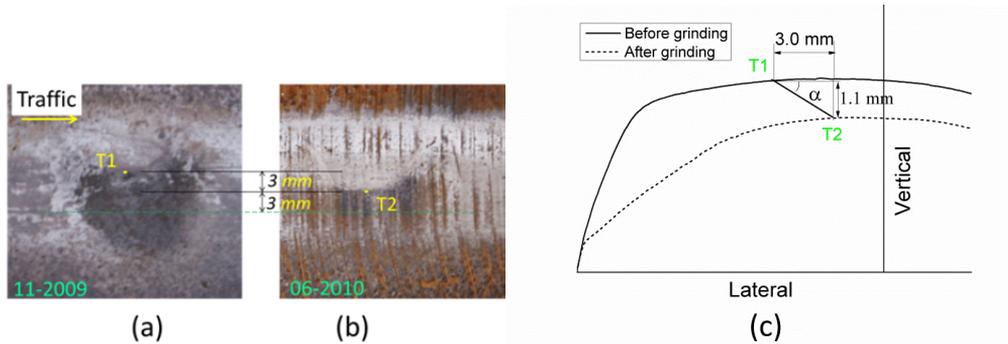


Fig. 4.7. Determination of the cracking angle relative to the rail surface in the vertical-lateral cross-section (the green line represents the rail centerline; T1 and T2 denote the bottom of the U-shaped crack): (a) Before rail grinding; (b) after rail grinding to a depth of 1.1 mm; (c) cross-sectional profiles of the rail.

#### 4.3.4. Crack shape in 3D and secondary cracks

##### 4.3.4.1. The primary U-shaped crack in 3D

The U-shaped crack initiated at the bottom point of the U and then propagated toward the gauge side in both the traffic and opposite-traffic directions simultaneously, as illustrated in Fig. 4.8. The crack also propagated into the rail at an angle of approximately  $20^\circ$ . This cracking process was observed at more than 20 monitored squats, as shown above. Based on this finding, the shape of the crack was constructed and is shown in Fig. 4.8(a). This shape was verified with 3D computed tomography [20], as seen in Fig. 4.8(b).

In Fig. 4.8, the crack face encloses a peninsula-shaped part. The intersection of the peninsula-shaped part with the rail surface is a U-shaped crack. When the U-shaped crack grows, the two arms of the U extend toward the gauge side; this makes the peninsula-shaped part longer. At the same time, the crack propagates downward into the rail, i.e., the peninsula-shaped part extends deeper into the rail. Consequently, the depressed area grows with the peninsula-shaped part simultaneously in both the surface dimension and depth.

Thus this peninsula-shaped part increasingly becomes an obstacle to the rolling wheels, and impact occurs when the wheels meet the peninsula-shaped part and fall from the peninsula-shaped part. These two impacts are clearly seen from the two shining patches at the peninsula-shaped parts and at the leading edges (see Fig. 4.8 for definitions of leading and trailing edges) of the black depressions in Figs. 4.1, 4.3, 4.4(c)-(e), 4.5(c)-(e), 4.6(d)-(e), 4.7, 4.9, 4.15 and 4.17. The two impacts correspond to peak contact forces  $B_2$  and  $C_2$  in Fig. 8(b) of [3]. Thus, the development process of the squats initiated from the corrugation and converged with the squat growth process postulated in [3] and validated in [4], although the initiation sources are different: in [3], the squats were assumed to be caused by pre-existing

rail surface defects, such as indentations and wheel burns; these squats belong to the passive type [13] that are caused by external factors. Corrugation-induced squats belong to the active type because these squats arise spontaneously due to defective tracks.

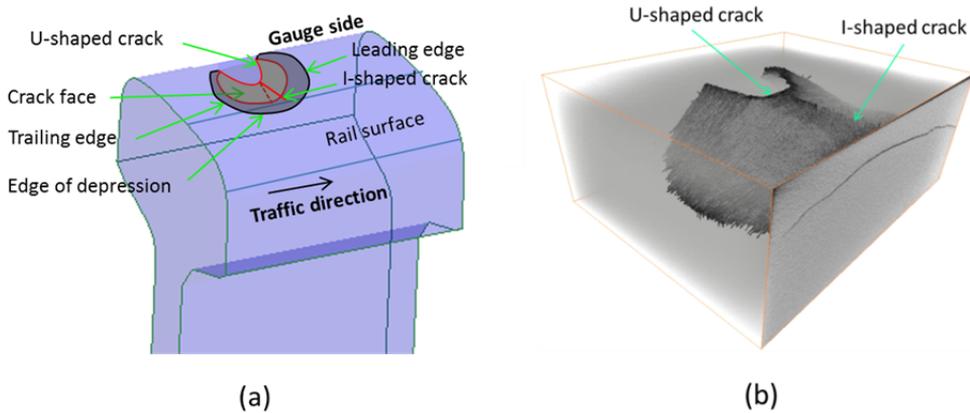


Fig. 4.8. 3D crack face in a two-lung-shaped squat: (a) Inferred from the surface-visible Y shape and the cracking angle  $\alpha$  of Fig. 4.7(c); (b) obtained from specimen 2b (squat caused by corrugation) from [20] with computed tomography, courtesy of Meysam Naeimi.

#### 4.3.4.2. Secondary I-shaped crack

The two impacts at the peninsula-shaped part punched and pushed the depressed rail material so that a “ridge” was formed in the middle of the black depression, as indicated in Fig. 4.9(a). This ridge is clearly shown on the surface in Figs. 4.1, 4.5(c)-(f), 4.6(e), 4.9, 4.10 and 4.15. Cracks can develop at the middle ridge, as seen in Figs. 4.5(f), 4.6(e), 4.8, 4.9 and 4.10, especially in severe squats. This crack is the secondary I-shaped crack. The I-shaped crack is also shown in the 3D geometry of the crack in Fig. 4.8(b).

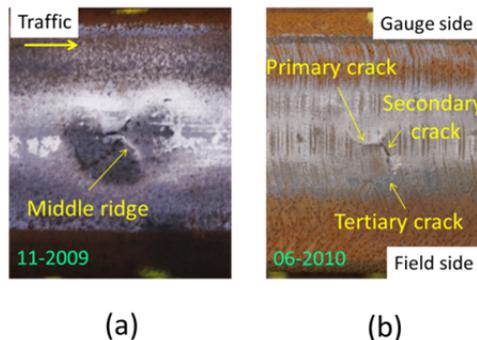


Fig. 4.9. A typical two-lung-shaped squat with a Y-shaped crack and a semi-circular tertiary crack (the same squat as in Fig. 4.5): (a) Before rail grinding; (b) after rail grinding to a depth of 1.1 mm.

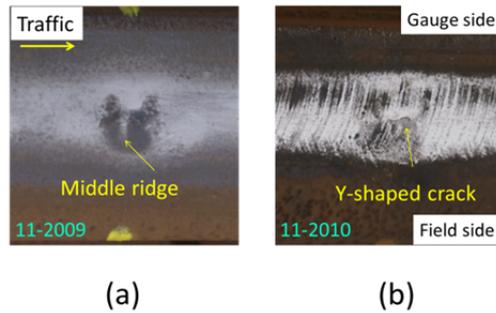


Fig. 4.10. Another typical two-lung-shaped squat with a Y-shaped crack (the same squat as in Fig. 4.6(e)): (a) Before rail grinding; (b) after rail grinding to a depth of 1.1 mm.

#### 4.3.4.3. Y-shaped crack and the two-lung shape of squats

When the U-shaped crack and I-shaped crack occur simultaneously, they together become a Y-shaped crack, as shown in Figs. 4.9 and 4.10. The Y divides the elliptical shape into the shape of two lungs, which occurs with the typical shape of severe squats.

4

#### 4.3.5. Squat development process

Based on the observations and analyses described above, a squat development process is proposed. A schematic diagram of the process is shown in Fig. 4.11 to describe the squat evolution and cracking processes. Initially, a small black depression is generated at a corrugation (Fig. 4.11(a)). The black depression results from microcracks due to plastic deformation and work hardening in the surface layer of the rail material, typically up to a depth of 1.1 mm, and it is caused by the cyclic wheel-rail impact forces excited by the corrugation. The black depression grows roughly in the shape of a complete ellipse until it reaches a critical size, and a primary macrocrack is initiated in/near the rail surface on the gauge side edge of the depression (Fig. 4.11(b)). Then, the shape of the black depression starts to change and follows the propagation of the crack, while the uncracking part remains elliptical. The surface crack propagates in a U shape toward the gauge side in both the traffic direction and the opposite-traffic direction (Fig. 4.11(c)). This process forms a U-shaped crack with the bottom point of the U being the initiation point of the primary crack. In this stage, the squat becomes a kidney-like shape with a concave part in the middle region bordered by the U-shaped crack. Moreover, the squat grows more rapidly in the traffic direction if the longitudinal contact force is tractive. It also propagates into the rail toward the field side at an angle of approximately  $20^\circ$ . When the squat grows to a certain stage, a ridge is formed in the middle of the black depression, and an I-shaped crack can occur at this ridge. The combination of the U-shaped crack and I-shaped crack forms a Y-shaped crack (Fig. 4.11(d)). In this process, the U-shaped part grows into a peninsula-like shape around the time of the formation of the I-shaped crack. Eventually, this evolution process leads to a mature two-lung-shaped squat with a Y-shaped crack. The proposed development process is derived from and is valid for corrugation-induced squats, and it is elaborated in the next section.

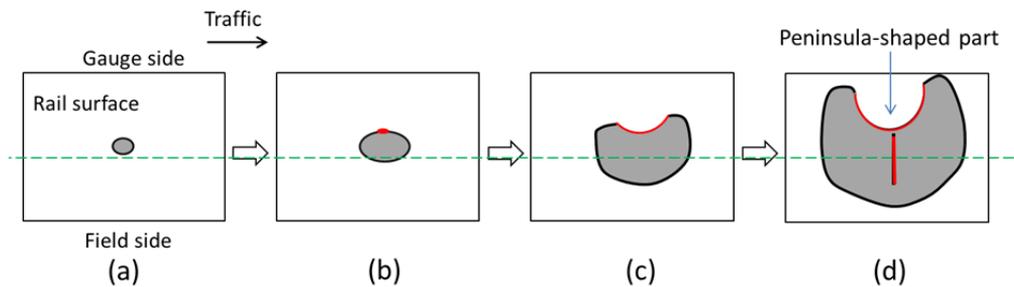


Fig. 4.11. Schematic diagram of the squat development process, including crack initiation and growth. The green line represents the rail centerline, the red line represents the cracks, and the gray area represents the squat/black depression with the black curved line representing its edge: (a) Elliptical small black depression; (b) crack initiation point, usually on the gauge side edge of the black depression; (c) U crack starts to form, causing the elliptical depression to become kidney shaped; crack propagates slightly more rapidly in the traffic direction due to traction force; (d) two-lung-shaped mature squat with a Y-shaped crack.

#### 4.4. Elaboration on the squat development process

The proposed squat development process is discussed in detail in this section. The following aspects are addressed: the formation of the small black depression, the crack initiation and propagation, and the influence of the rail inclination on the initial cracking locations.

##### 4.4.1. Formation of small black depressions with reduced tensile strength

In the running band, the rails are subjected to large cyclic wheel-rail interaction forces. These forces produce repeated plastic deformation and work hardening associated with microcracks [21]. The plastic deformation and work hardening deteriorate the RCF resistance [22], in turn increasing the brittleness of the materials. As a result, macrocracks initiate more rapidly when new crack initiation mechanism arises. The crack can initiate in the manner of the growth of a single microcrack or the coalescence of many microcracks, as noted in [23] and [24].

When corrugation is present, the plastic deformation and work hardening typically occur most significantly at the places where the dynamic contact stress is greatest. When the hardening reaches a degree of saturation, the ductility of the rail material is exhausted, and the local material becomes brittle accompanied by a network of microcracks. The network of microcracks expands in the rail surface layer and causes the surface to collapse locally to form a small elliptical black depression, such as those shown in Figs. 4.4, 4.5 and 4.6.

##### 4.4.2. Crack initiation by mode I fracture

When a small black depression grows to a certain size, it significantly changes the effective contact patch at the wheel-rail interface, as illustrated in Fig. 4.12. The contact patch can be approximated as elliptical when the rail surface is smooth, as shown in Figs. 4.12(a) and 4.10(b). However, its location and shape are different in the region when a small black depression exists. The main contact patch at the depression shifts to the gauge side, as shown in Fig. 4.12(c) and (d). The border of the main contact area at the edge of the depression exhibits a U shape with the opening of the U facing toward the gauge side. The U shape comes

from the intersection of the wheel and the rail surfaces as a wheel is loaded onto a rail. This change in the contact patch is generalized and marked in Fig. 4.12(c) and (d). This main contact area is most clearly indicated by the brighter contact area bordered by the U-shaped crack at the monitored squats, such as those seen in Figs. 4.1, 4.4(d) and (e), 4.5, 4.6(d) and (e), 4.7, 4.9, 4.10, 4.15 and 4.17.

The region at the black depression sustains the impact from the wheels; the dynamic force is large. At the same time, the depression causes stress concentration at edge T1, as seen in Fig. 4.12(c), (d) and (e). The large contact force and stress concentration lead to stress that is much larger than elsewhere on the rail surface. According to the Hertz theory [25], tensile stress is present immediately outside the contact patch, and its value is maximal at the border of the contact. In the case of corrugation with surface depression, a tensile stress distribution arises along the border of the main contact area in a U shape, as indicated by the red arrows in Fig. 4.12(c). The tensile stress is the greatest at T1. Since the material at the depression is brittle, this greatest tensile stress results in a macrocrack at T1 by mode I fracture. This crack initiation phenomenon is analogous to the ring/cone crack formation of brittle materials under sphere contact [26] and the formation of crack in RCF by asperities [27]. Note, T1 is located on the gauge side edge of the depression approximately 6 mm from the rail center for most of the squats in this chapter. It could change depending on the profiles of wheels and rails, the local track conditions and the material property.

4

#### 4.4.3. Crack propagation into the U shape

Once this primary crack is present, the tensile stress drives it to propagate along the border (the solid black line in Fig. 4.12(c) and (d)) of the main contact area in a U shape by stress concentration and mode I fracture toward the gauge side in both directions of the U-shaped crack. This fracture mode was observed in [8]. It forms the U-shaped primary cracks.

When traction force is applied, the additional shear stress adds tension to the leading side of the contact patch and subtracts from the tensile stress on the trailing side [28]. Therefore, the resulting tensile stress is larger on the leading side of the contact area than that on the trailing side, as shown in Fig. 4.12(d). This outcome causes the crack to propagate more rapidly on the leading edge side; this result agrees with most of the observations at the monitored site, such as the squats in Figs. 4.4, 4.5, 4.6(c)-(e), 4.9, 4.10 and 4.15. The mechanism of this cracking phenomenon is analogous to that of partial ring/cone crack formation of brittle materials under sphere-sliding contact [29].

In summary, the initiation of the primary crack is caused by tensile stress with mode I fracture at the T1 location, which corresponds to the bottom of the U-shaped crack. The U-shaped crack corresponds to the border of the main contact patch at the depression. This finding was observed at the monitored squats that have crack initiation on the gauge side edge of the depressions, corresponding to the situation in Fig. 4.12(e) and Fig. 4.13(a).

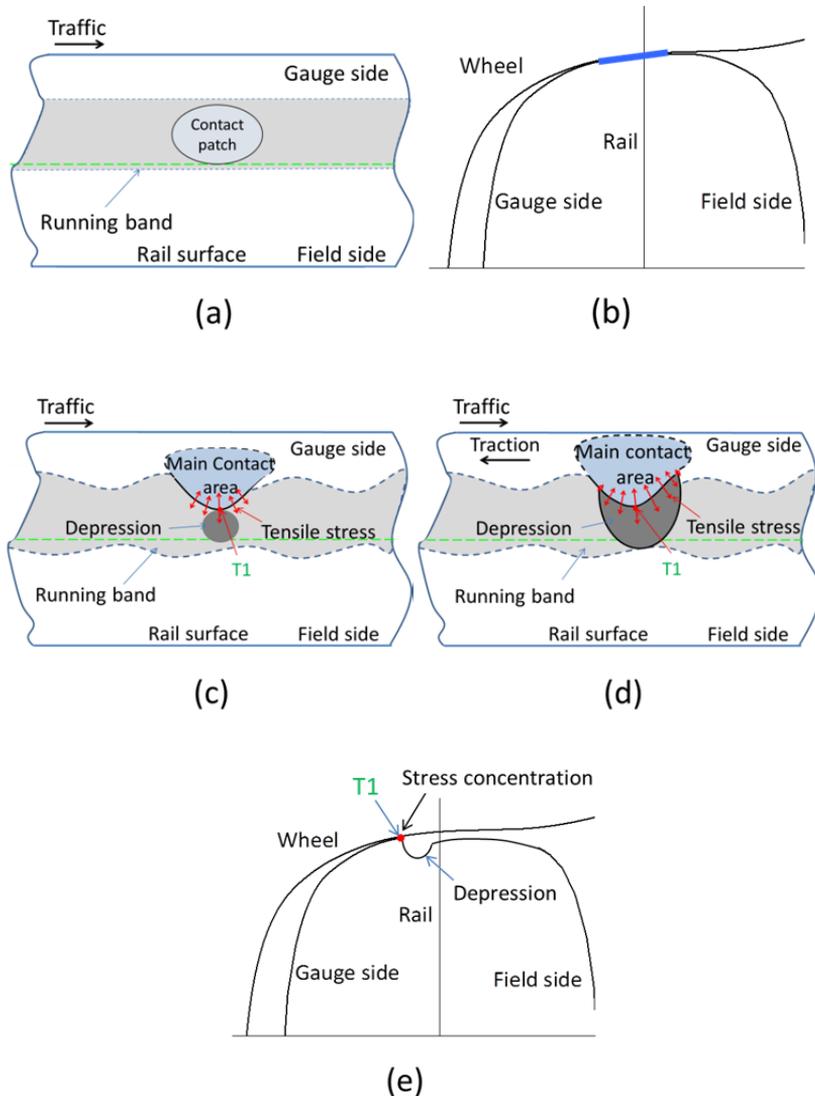


Fig. 4.12. Wheel-rail contact patch. The green line represents the rail centerline; the blue dashed lines mark the running band and corrugation; and the blue solid line indicates the main contact area: (a)-(b) When the rail is smooth; (c)-(e) when corrugation and a small black depression exist. The red arrows indicate tensile stress along the border of the main contact area; the length of the arrows is schematically proportional to the magnitude of the tensile stress. (c) corresponds to Fig. 4.11(b), and (d) corresponds to Fig. 4.11(c) and (d).

#### 4.4.4. Influence of rail inclination on cracking and tertiary cracks

Primary cracks at the squats were typically observed on the gauge-side edge of the black depression. However, cracks could be observed on the field side and at the middle ridge as

well as on both the gauge and field sides. These possibilities would happen because the crack initiation location is determined by the position of the main contact area, and the main contact area may vary. The rail inclination plays an important role in determining the position of the main contact area, and the potential positions are shown in Fig. 4.13.

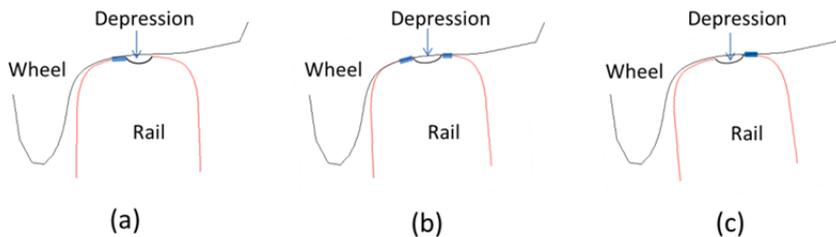


Fig. 4.13. Positions of the main contact areas (indicated by the blue line) that determine the locations of the primary cracks: (a) Rail inclination of  $1/40$ ; (b) a slight increase in the rail inclination; (c) a larger increase in the rail inclination.

4

When the rail inclination is at the nominal value of  $1/40$ , as shown in Fig. 4.13(a), the presence of the black depression shifts the main contact area to the gauge side. This shift is determined by the contact geometry, namely the profiles of the wheel and the rail, as illustrated in Fig. 4.12(b) and, especially, 4.12(e): the wheel-rail profiles are more conforming to each other on the gauge side than on the field side so that, when there is a depression, the main contact area is usually on the gauge side. As a result, the crack tends to initiate and propagate into a U shape on the gauge side edge of the black depression. This scenario in which the location of the primary crack is on the gauge side (schematically shown in Fig. 4.12) occurs at most of the monitored squats, such as those squats shown in Figs. 4.1, 4.3-4.10, 4.15 and 4.17, the only exception is the case shown in Fig. 4.14 corresponding to the situation in Fig. 4.13 (c).

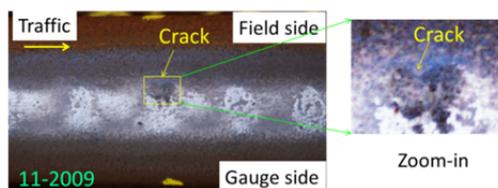


Fig. 4.14. A squat with a primary crack on the field side edge of the black depression.

Under operating conditions, the effective rail inclination can be different from the nominal state for various reasons, e.g., loose fastenings, which can influence the wheel-rail contact positions and the resulting cracking behavior. For instance, a large increase in the rail inclination causes the wheel and rail to come into the main contact on the field side, and the corresponding primary crack is formed on this side, as illustrated in Fig. 4.13(c). This scenario was occasionally observed in the continual field monitoring, and an example is shown in Fig. 4.14. This crack exhibited a U shape with the opening of the U facing toward the field side.

The case between the situations Fig. 4.13(a) and (c) is a slight increase in the rail inclination that slightly shifts the contact to the field side. This slight increase, which could render wheel-rail contact simultaneously at both the gauge side and the field side of the black depression, as illustrated in Fig. 4.13(b), and thus possible almost simultaneous initiation of the primary cracks on both the gauge and field sides, although such a situation has not yet been observed in real life.

For the situation in Fig. 4.13(a), the main contact area can sink significantly with the development of the primary crack and the growth of the depressed area in both the surface dimension and the depth. The growth of the depressed area increases the excited impact frictional rolling contact forces and contact stress on the main contact area, causing plastic deformation and wear. Therefore, the main contact area may progressively become lower than its less damaged surrounding area. As a result, the main contact shifts to the field side of the depression; the contact position becomes similar to that in Fig. 4.13(b) and further in Fig. 4.13(c). This outcome leads to the formation of a tertiary crack on the field side of the depression, as shown in Fig. 4.15. These tertiary cracks were observed at 24 squats at the monitored site at the time of the fifth and sixth observations. They exhibited a semi-circular crack along the edge of the elliptical edge of the black depressions with the opening of the semi-circular crack facing toward the rail centerline. These cracks could appear before or after the formation of the I-shaped crack described in Section 4.3.4.2, depending on the rail inclination: the increase in rail inclination could cause the wheel-rail contact at the field side sooner when the main contact area is sinking, which would promote the formation of the subsequent tertiary semi-circular crack on the field side of the depression sooner than the formation of the I-shaped crack.

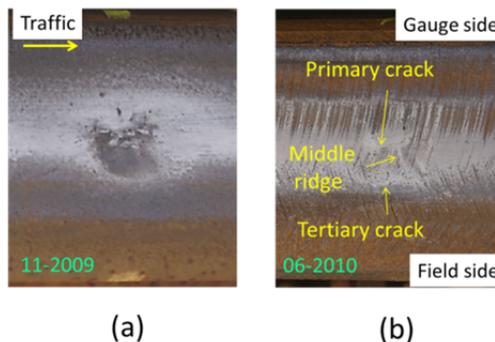


Fig. 4.15. A squat with two cracks, i.e., a primary crack and a tertiary crack: (a) Before rail grinding; (b) after rail grinding to a depth of 1.1 mm.

In summary, there were the following surface cracks observed: primary U-shaped cracks with the opening of the U opposite the black depression (they were usually observed on the gauge side); secondary I-shaped cracks in the middle ridge; and tertiary semi-circular cracks with the opening of the semi-circular cracks facing the black depression.

## 4.5. Discussion

### 4.5.1. Evolution of the squat geometry and comparison with other types of squats

In this investigation, the squat is induced by the rail surface irregularities of corrugation; in [3], observations of squats were made mainly based on indentation-induced squats. Works in [3, 4, 14] studied squat growth in the traffic direction and provided evidence that the squats also grew in the opposite-traffic direction. However, this aspect was not explored in detail. The current study proves that the geometry of squats initiated at corrugation grows significantly in the opposite-traffic direction, such as the squats shown in Figs. 4.4 and 4.5. The evolution process of squat geometry in the longitudinal direction is schematically illustrated in Fig. 4.16. In the beginning, a black depression is induced by impact force at corrugation. The black depression grows, and then, a primary U-shaped crack initiates on the edge of the depression. The crack continues to propagate with the U shape in both the traffic and opposite-traffic directions. The shape of the squat evolves following the primary crack in the two directions. When traction force is applied, the crack grows more rapidly in the traffic direction, and length  $L_2$  is greater than  $L_1$ . In the cases of braking, the evolution of the geometry in the traffic direction should have a lower rate, i.e.,  $L_2$  should be smaller than  $L_1$ . When the squat grows to a certain size, a peninsula forms. Roughly from this moment on, the development of the squats follows the growth process described in [3], and a middle ridge is produced due to the impact forces caused by the peninsula. As a result, an I-shaped crack is generated at this ridge. The squat then takes on the typical two-lung shape.

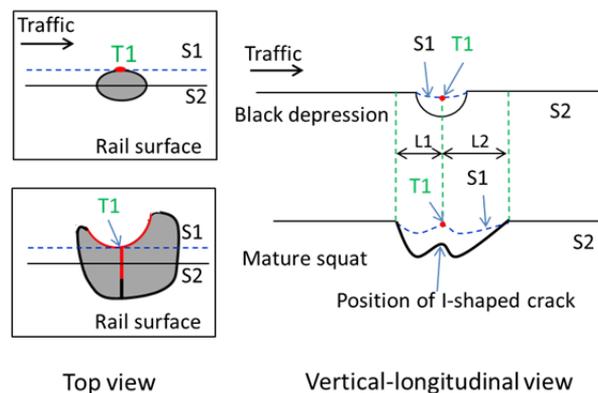


Fig. 4.16. Evolution of the squat geometry. S1 and S2 indicate the two vertical-longitudinal sections.

Some existing works, e.g., [10, 12], have studied crack propagation at squats. These studies concluded that a leading crack was generated first, and then propagated in the opposite-traffic direction to create a trailing crack. The leading crack and trailing crack form the U-shaped cracks at squats. The five-year continual field monitoring in the current study presented a different crack initiation and propagation process, where the cracks initiate at the bottom of a U-shaped crack and grow along the two arms of the U in both the traffic direction and opposite-traffic direction at the same time. The mechanism based on tensile stress

reasonably interprets the cracking phenomenon. It is not yet known from the present study if this process and mechanism of the crack initiation and initial growth are applicable to squats from other initiation sources.

Not only corrugation-induced squats but also those squats induced by other sources are developed from initial surface irregularities. These initial irregularities generally do not contain initial cracks. They involve the passive type of squats initiation sources, e.g., indentation and wheel burns, the active type, e.g., corrugation and welds, and the design type, e.g., insulation joints and crossing. The resulting cracks start from the primary virgin cracks. Existing cracks left over by insufficient grinding can develop into squat-like defects which are sometimes also called squats. These initiation sources are very different, but their resulting squats bear the same characteristic of two-lung shape for mature squats with U, V and Y-shaped macro cracks. The reason for all these disparate initiation sources to converge to the same appearance in the end is their common nature of mechanics. That is, they all excite dynamic contact forces at irregularities, resulting in the two-lung shape, the wavelength of which is determined by that of the dynamic force in relation to the local vehicle-track system. The study regarding to the relation between the wavelength of squat and that of dynamic force has been provided in [3, 4] with numerical analysis and field observations. The field observation in the present study again confirms this relation. The characteristic of the wavelength can be used as signature for dynamics-based early detection and track health condition monitoring [16, 30]. Timely maintenance based on early detection can not only guarantee operation safety, but also greatly reduce life cycle costs of railway infrastructure [31].

Despite the convergence to the common appearance of mature squats because of the same nature of mechanics, there could be differences in the initiation stages of the squats cracks from the different irregularities. The differences in the process and mechanisms of the initiation of primary cracks and early growth with those of corrugation-induced squats remain to be investigated. Such investigation will facilitate rail material design and selection in relation to loading condition and damage.

#### **4.5.2. Squat development after rail grinding**

During the ten continual observations, rail grinding to a depth of 1.1 mm was performed after the fifth observation. Some of the squat cracks had already propagated more than 1.1 mm into the rails; therefore, the rail grinding did not remove the cracks completely. The remaining cracks grew further and developed into a new squat, as shown in Fig. 4.17.

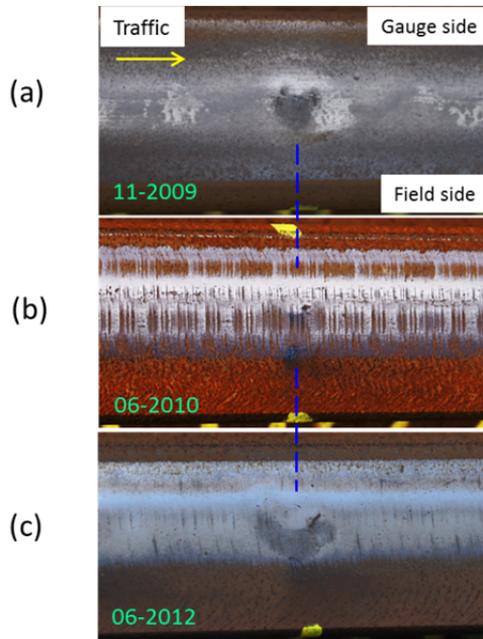


Fig. 4.17. Squat (the same as in Fig. 4.4) development after rail grinding: (a) Fifth observation before rail grinding; (b) sixth observation after grinding to a depth of 1.1 mm; (c) tenth observation after another grinding of approximately 0.3 mm between the seventh and eighth observations.

## 4.6. Conclusions

Five years of continual field monitoring have captured various stages of the life cycle of corrugation-induced squats, from very small black depressions without cracks to mature two-lung-like squats typically accompanied by Y-shaped cracks. Based on the observations and analyses of 30 squats, the following conclusions can be drawn.

- 1) Short pitch corrugation led to the formation of the squats. The cracks initiated in the surface or at least within the top layer of rail material not deeper than 1.1 mm.
- 2) A squat induced by short pitch corrugation started from a small elliptical black depression at the corrugation. The small black depression was formed by the local collapse of a network of microcracks resulting from large plastic deformation and work hardening due to impact forces.
- 3) When the black depression grew to approximately 8 mm in size with the maximum distance of the depression edge on the gauge side edge approximately 6 mm from the rail center, the large tensile stress initiated the first macrocrack at this point of maximum distance from the depression. These 8 mm correspond to the critical size for squat initiation proposed in [13], derived by comparing the maximal von Mises stress with tensile strength, the material being R260Mn. The crack initiation location of 6 mm from the rail center could depend on the contact geometry of wheels and rails, the local track conditions and the material property.

- 4) These macrocracks propagated along the border of the main contact area in both the traffic direction and the opposite-traffic direction, forming a U-shaped primary crack in/near the rail surface on the gauge side edge, with the U facing toward the gauge side. The crack initiation point is at the bottom of the U-shaped crack. The crack also propagates into the rail toward the field side at an angle of approximately  $20^\circ$ . The primary crack grew more rapidly in the traffic direction when traction force was applied.
- 5) After crack initiation, the squat geometry could not remain fully elliptical. Its gauge side edge became bordered by the U-shaped crack, while the part without cracking remained elliptical, causing the squats at this stage to take on a kidney-like shape. With further growth of the squats along the two arms of the U in the rail surface and the propagation deeper into the rail, the primary cracks gradually formed a peninsula-shaped part that protruded from its depressed surrounding materials. The peninsula-shaped part formed an obstacle to the rolling wheels, causing two impact forces, the first arose when wheels impacted the peninsula-shaped part and the second arose when the wheels fell from the peninsula-shaped part onto the leading edge of the squats. This outcome corresponds to the squat growth process described in [3, 4, 13, 14].
- 6) These impact forces, especially the first one, led to a ridge in the middle of the depressed black part of the squats, where a secondary I-shaped crack appeared subsequently. Together with the primary U-shaped crack, it constituted a Y-shaped crack. This Y-shaped cracks divided the black depression into two parts so that the typical two-lung shape of mature squats was formed.
- 7) A further tertiary semi-circular crack could occur along the elliptical edge of the depression on the opposite side of the primary U-shaped crack due to sinking of the main contact area.
- 8) The change in the rail inclination could shift the wheel-rail impact region so that occasionally the primary U-shaped crack was found on the field side, and the corresponding tertiary semi-circular crack could be found on the gauge side.
- 9) In general, the opening of the primary U-shaped crack, regardless of whether on the gauge or field side, is opposite to the black depression. The opening of the tertiary semi-circular cracks faces the black depression.
- 10) The findings in this chapter contribute to a new understanding of rail RCF. The squat initiation and cracking processes are elaborated and better revealed. The results could help rail engineers perform more effective inspections and plan maintenance more efficiently to reduce costs and improve operation safety. The pin-pointing of the crack initiation location and the identified tensile stress state for crack initiation and propagation should benefit the design and manufacturing of more squat-resistant materials.

## References

- [1] P. Clayton, M. Allery, Metallurgical aspects of surface damage problems in rails, *Canadian Metallurgical Quarterly*, 21 (1982) 31-46.
- [2] R. Nakamura, S. Owaku, N. Enomoto, The rail shelly crack in Japan, *Railway Technical Research Institute, Quarterly Reports*, 6 (1965).
- [3] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats—Correlation analysis and numerical modeling, *Wear*, 265 (2008) 1349-1355.
- [4] Z. Li, R. Dollevoet, M. Molodova, X. Zhao, Squat growth—Some observations and the validation of numerical predictions, *Wear*, 271 (2011) 148-157.
- [5] S. Pal, C. Valente, W. Daniel, M. Farjoo, Metallurgical and physical understanding of rail squat initiation and propagation, *Wear*, 284-285 (2012) 30-42.
- [6] S. Pal, W.J.T. Daniel, M. Farjoo, Early stages of rail squat formation and the role of a white etching layer, *International Journal of Fatigue*, 52 (2013) 144-156.
- [7] S.L. Grassie, D.I. Fletcher, E.A. Gallardo Hernandez, P. Summers, Studs: a squat-type defect in rails, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226 (2011) 243-256.
- [8] S. Simon, A. Saulot, C. Dayot, X. Quost, Y. Berthier, Tribological characterization of rail squat defects, *Wear*, 297 (2013) 926-942.
- [9] S. Bogdański, P. Lewicki, 3D model of liquid entrapment mechanism for rolling contact fatigue cracks in rails, *Wear*, 265 (2008) 1356-1362.
- [10] S. Bogdański, Quasi-static and dynamic liquid solid interaction in 3D squat-type cracks, *Wear*, 314 (2014) 20-27.
- [11] M. Farjoo, S. Pal, W. Daniel, P.A. Meehan, Stress intensity factors around a 3D squat form crack and prediction of crack growth direction considering water entrapment and elastic foundation, *Engineering Fracture Mechanics*, 94 (2012) 37-55.
- [12] M. Steenbergen, Squat formation and rolling contact fatigue in curved rail track, *Engineering Fracture Mechanics*, 143 (2015) 80-96.
- [13] Z. Li, X. Zhao, R. Dollevoet, An approach to determine a critical size for rolling contact fatigue initiating from rail surface defects, *International Journal of Rail Transportation*, 5 (2017) 16-37.
- [14] Z. Li, Squats on railway rails, *Wheel–Rail Interface Handbook*, Woodhead Publishing, 2009, pp. 409-436.
- [15] M. Oregui, S. Li, A. Núñez, Z. Li, R. Carroll, R. Dollevoet, Monitoring bolt tightness of rail joints using axle box acceleration measurements, *Structural Control and Health Monitoring*, 24 (2017).
- [16] Z. Wei, A. Núñez, Z. Li, R. Dollevoet, Evaluating degradation at railway crossings using axle box acceleration measurements, *Sensors*, 17 (2017) 2236.
- [17] S.L. Grassie, Squats and squat-type defects in rails: the understanding to date, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226 (2011) 235-242.

- [18] R. Andersson, P.T. Torstensson, E. Kabo, F. Larsson, A. Ekberg, Integrated analysis of dynamic vehicle-track interaction and plasticity induced damage in the presence of squat defects, *Wear*, 366–367 (2016) 139-145.
- [19] M. Hiensch, Improving Track-Friendliness of Rolling Stock, Proceedings of the International Heavy Haul Association, Australia,, (2015).
- [20] M. Naeimi, Z. Li, Z. Qian, Y. Zhou, J. Wu, R.H. Petrov, J. Sietsma, R. Dollevoet, Reconstruction of the rolling contact fatigue cracks in rails using X-ray computed tomography, *NDT & E International*, 92 (2017) 199-212.
- [21] J. Morrow, Cyclic plastic strain energy and fatigue of metals, Internal friction, damping, and cyclic plasticity, *ASTM International*, 1965.
- [22] W.R. Tyfour, J.H. Beynon, A. Kapoor, Deterioration of rolling contact fatigue life of pearlitic rail steel due to dry-wet rolling-sliding line contact, *Wear*, 197 (1996) 255-265.
- [23] M.E. Fine, Fatigue resistance of metals, *Metallurgical and Materials Transactions A*, 11 (1980) 365-379.
- [24] P. Forsyth, D. Ryder, Some results of the examination of aluminum alloy specimen fracture surfaces, *Metallurgia*, 63 (1961) 117.
- [25] K.L. Johnson, Contact mechanics, Cambridge university press, Cambridge, 1987.
- [26] F. Frank, B.R. Lawn, On the theory of Hertzian fracture, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, The Royal Society, 1967, pp. 291-306.
- [27] B. Alfredsson, J. Dahlberg, M. Olsson, The role of a single surface asperity in rolling contact fatigue, *Wear*, 264 (2008) 757-762.
- [28] G.M. Hamilton, L.E. Goodman, The Stress Field Created by a Circular Sliding Contact, *Journal of Applied Mechanics*, 33 (1966) 371-376.
- [29] B. Lawn, Partial cone crack formation in a brittle material loaded with a sliding spherical indenter, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, The Royal Society, 1967, pp. 307-316.
- [30] M. Molodova, Z. Li, A. Núñez, R. Dollevoet, Automatic Detection of Squats in Railway Infrastructure, *Ieee T Intell Transp*, 15 (2014) 1980-1990.
- [31] Z. Li, M. Molodova, X. Zhao, R. Dollevoet, Squat treatment by way of minimum action based on early detection to reduce life cycle costs, 2010 Joint Rail Conference, American Society of Mechanical Engineers, 2010, pp. 305-311.



# 5

## **PRE-CRACKING DEVELOPMENT OF WELD-INDUCED SQUATS DUE TO PLASTIC DEFORMATION: FIVE-YEAR FIELD MONITORING AND NUMERICAL ANALYSIS**

*Weld-induced squats are a major damage type in high-speed railways as well as in conventional railways. They incur high maintenance costs and endanger operational safety. This chapter first presents and analyzes five-year continual field monitoring observations and measurements of squats at rail welds. A hypothesis of the formation and development process of the squats is proposed, which includes three steps. Steps 1 and 2 are pre-cracking, and Step 3 is post-cracking. To verify the pre-cracking process, a 3D FE model is then built up to simulate the vehicle-track interaction with detailed consideration of the local wheel-rail frictional rolling contact. Not only dynamic contact forces but also plastic deformation and wear are calculated. Starting from a smooth rail surface with varying yield stress derived from field-measured hardness, the numerical analysis confirms the hypothesis that the varying hardness at heat-affected zones (HAZs) leads to initial V-shaped irregularities due to differential plastic deformation. Afterward, the surface irregularities excite the dynamic longitudinal contact force, which in turn produces a W-shaped surface pattern through further differential plastic deformation. The growth of the W-shaped pattern leads to the formation of squats. This chapter provides insight into the squat formation process at rail welds and suggests that welding quality control in terms of hardness variation in the HAZs could reduce or even avoid squats. Early detection of squats with dynamics-based methods is possible.*

## 5.1. Introduction

Welding has been widely employed to join rails in railway tracks. However, the welding process can introduce changes in the rail microstructures and its mechanical properties, as well as geometric irregularities. Particularly, during the welding process, heat-affected zones (HAZs) are produced in the vicinity of the weld metal (WM) [1]. The heat inevitably alters the microstructures in the rail steel, leading to phase transformation, mostly causing a lower hardness distribution within the HAZs [2]. As a result, a spatially inhomogeneous material properties distribution is introduced along the rail [3]. Moreover, geometric irregularities often excite high dynamic wheel-rail contact forces. These factors contribute to localized rolling contact fatigue (RCF) damage both at and under the rail surface. Several types of RCF damage have been identified at welds, including squats in the rail surface and the damage in the rail web.

Many efforts have been made to limit and reduce RCF at welds. Ilić et al. [4] proposed that heat treatment could improve the mechanical properties of welds. The influences of residual stress and porosity on crack initiation at rail web were studied by Skyttebol et al. [5] and Fry et al. [6]. Moreover, some studies, such as the investigation performed by Desimone and Beretta [7], focused on crack propagation in the rail web at welds. A further concern is the vehicle-track dynamic interaction at the railhead in the vicinity of welds because the dynamic forces resulting from the initial irregularities of welding would significantly influence RCF damage. Steenbergen and Esveld [3] studied the interaction at welds using numerical simulations with consideration of the measured geometry of welds. Wen et al. [8, 9] studied the vehicle-track interaction and the plastic deformation at both a straight track and a curved track. They found that the initial irregularities of welds had a significant influence on the dynamic contact forces. For instance, Gao et al. [10] analyzed measured rail surface geometry of welds in a Chinese high-speed railway and proposed a model to study the dynamic interaction at the welds. They found that the weld irregularities dominantly appeared in short-wavelength form, which have considerable impacts on wheel-rail dynamic responses. Recently, Sh. Sichani and Bezin [11] presented a model to estimate the evolution of geometry at welds.

Despite these works on rail welds, few studies are found on rail squats at welds. Squats are a main type of rail RCF defect [12, 13]. They are identified as local depressions in the rail surface with cracks [14]. Squats are mostly observed on tangent tracks and shallow curved tracks. A series of investigations by Li et al. [15-20] reported that squats initiate from various sources, such as welds, indentations, short pitch corrugation, insulation joints and crossings. As one of the main initiation sources, welds have been shown to contribute to the formation of a large population of squats [15]. For instance, it was found through field surveys in the Netherlands that 10 – 15% of squats were caused by welds [17]. They were observed at both thermite welds and flash butt welds, as shown in Fig. 5.1. At a thermite weld, duo squats are frequently observed, and at a flash butt weld, there is usually only a single squat. The accompanying cracks at squats can propagate in the rail and may ultimately lead to rail break. On a European railway network, approximately 15,000 welding repairs have to be performed every year because of squat damage [20]. Grinding is a technique widely used to remove squats, protecting the rail from further degradation. However, grinding is expensive and influences the availability of tracks. A good understanding of squat formation and development helps in developing and deploying efficient control measures against squats in the early stage, reducing maintenance costs and improving availability.

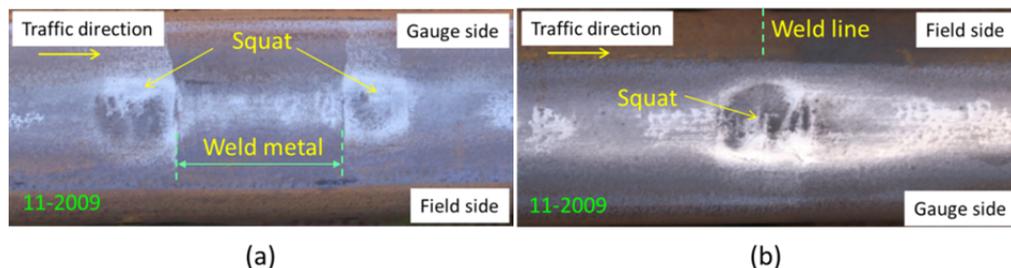


Fig. 5.1. Typical squats at welds: (a) duo squats at a thermite weld; (b) a single squat at a flash butt weld.

Through field observations and numerical analysis, Li et al. [15, 16] suggested that the initiation and development of squats are related to high-frequency vehicle-track dynamic interaction forces. The frequency is determined by certain eigen characteristics of the vehicle-track system and the length of rail surface irregularities. They explained how a small surface irregularity under dynamic forces develops into a typical squat with the characteristic of a two-lung-like shape. Li et al. [16] also proposed that squats at welds are caused by differential plastic deformation and wear and the resulting surface irregularity because of HAZs. However, the specific formation and development process observed from continual field monitoring has not been provided, and the causes taking into account differential plastic deformation and wear have not been studied either.

Field observation is an effective approach to study rail surface damage. Chapter 4 presented an extensive five-year continual field monitoring study on corrugation-induced squats. The development process of squats from early small surface depressions to the typical two-lung shape, as well as the growth process of the associated geometry and cracks, was presented, alongside a discussion of the causes of the impact force of wheels on rails that is responsible for the development of the geometry. These processes and causes can explain the formation of corrugation-induced squats. However, squats induced by welds have initiation sources other than corrugation, and their formation needs further investigation.

This chapter investigates the mechanism of squat initiation and growth at rail welds through both field observations and numerical pre-cracking simulations. To this end, continual field monitoring observations and measurements were first performed to identify the development features of the squats. Based on the observations and considering the effects of the dynamic wheel-rail forces on squat development presented in [15], a hypothesis for the development process of the squats is proposed. Afterward, a 3D FE dynamic vehicle-track interaction model with detailed consideration of wheel-rail contact is employed to verify the hypothesis by computing the dynamic contact forces, stresses, plastic deformations and wear. The FE model considers field-measured rail hardness and surface irregularities.

## 5.2. Field monitoring observations and measurements

To investigate squat initiation and development, a five-year field monitoring was performed in the period from 2007 to 2012 on a straight track located at Steenwijk in the Netherlands. One-directional traffic ran on the selected track. A detailed description of the monitoring is presented in Chapter 4. Briefly, ten rounds of observations were made at intervals of approximately six months during the monitoring. Squats caused by various sources, such as

corrugation, welds and wheel burns, were monitored. During each observation round, selected squats were photographed, and their cross-sectional profile and vertical-longitudinal profiles were measured using Miniprof and RailProf devices, respectively. RailProf measures at every 5 mm along the rail with a maximum error of  $\pm 0.03$  mm for profile deviation up to 0.5 mm and a maximum error of  $\pm 5\%$  of the measured value for profile deviation between 0.5 mm and 1.5 mm. In this chapter, squats caused by welds on the monitored track are the focus. Moreover, another five-round field monitoring was carried out near Assen, the Netherlands, from 05-2010 to 10-2012. Some data in this monitoring is also used in this chapter. Some other welds from a number of one-off field observations and measurements at Nijeveen and Zaandam in the Netherlands are further employed to illustrate the development of squats at welds. Both thermite welds and flash butt welds are discussed.

Below, it is shown in Section 5.2.1 that the weld-induced squats occurred at the HAZs. Section 5.2.2 shows that geometrical irregularities had already developed at the HAZs before the squats occurred. Section 5.2.3 shows that the squats and the irregularities have a strong correlation with the hardness variation in the HAZ. These observations form the basis for the hypothesis made in Section 5.3.

## 5.2.1. Squats with cracks at welds

### 5.2.1.1. Squats at thermite welds

Fig. 5.2 shows duo squats at a thermite weld in the first, fifth and sixth observations of the monitoring at Steenwijk. The rail was ground between the fifth and the sixth observations, as mentioned in Chapter 4. The photos are aligned with the two fusion lines (the red dashed lines) that border the WM and the parent rail material. The distance between the two lines outside the running band should be constant during the service life of the weld (approximately 59 mm in the current case) so that it can be used as a reference for aligning the photos.

In 11-2009, the squat on the left side already had the typical two-lung shape of squats with a slightly shining middle ridge, b1', between the two lungs, which were dark depressions. Because the depressions were lower than the normal rail surface and the ridge was higher than the depressions, the vertical-longitudinal profile across the squat would have a W-shaped pattern, as illustrated in Fig. 5.2(c). The vertical-longitudinal profile measured in 09-2007 (Fig. 5.2(b)) shows that a W-shaped pattern a1b1c1 already existed as early as at the first observation. A W-shaped pattern consists of two V-shaped dips, with one between the WM and the ridge (b1c1) and the other (a1b1) on the other (far) side of the ridge.

The defect on the right side did not yet bear the typical appearance of mature squats in 11-2009, and a middle ridge could be seen only by experienced eyes in the photo. The 06-2010 photo, however, reveals a primary crack typical of squats, as discussed in Chapter 4. The primary cracks of squats usually develop into a U-shape when squats are in their severe stage, e.g., see Li et al. [16] and Chapter 4, though in the 06-2010 photo the primary cracks of both defects did not yet have a full U-shape, but only an early and, thus, shallow and incomplete U-shape. The middle ridge of a squat usually starts from the bottom point of the U and extends to the opposite side of the squat, as observed in Chapter 4, and this is the case for both defects in Fig. 5.2. Thus, the right defect was also a squat.

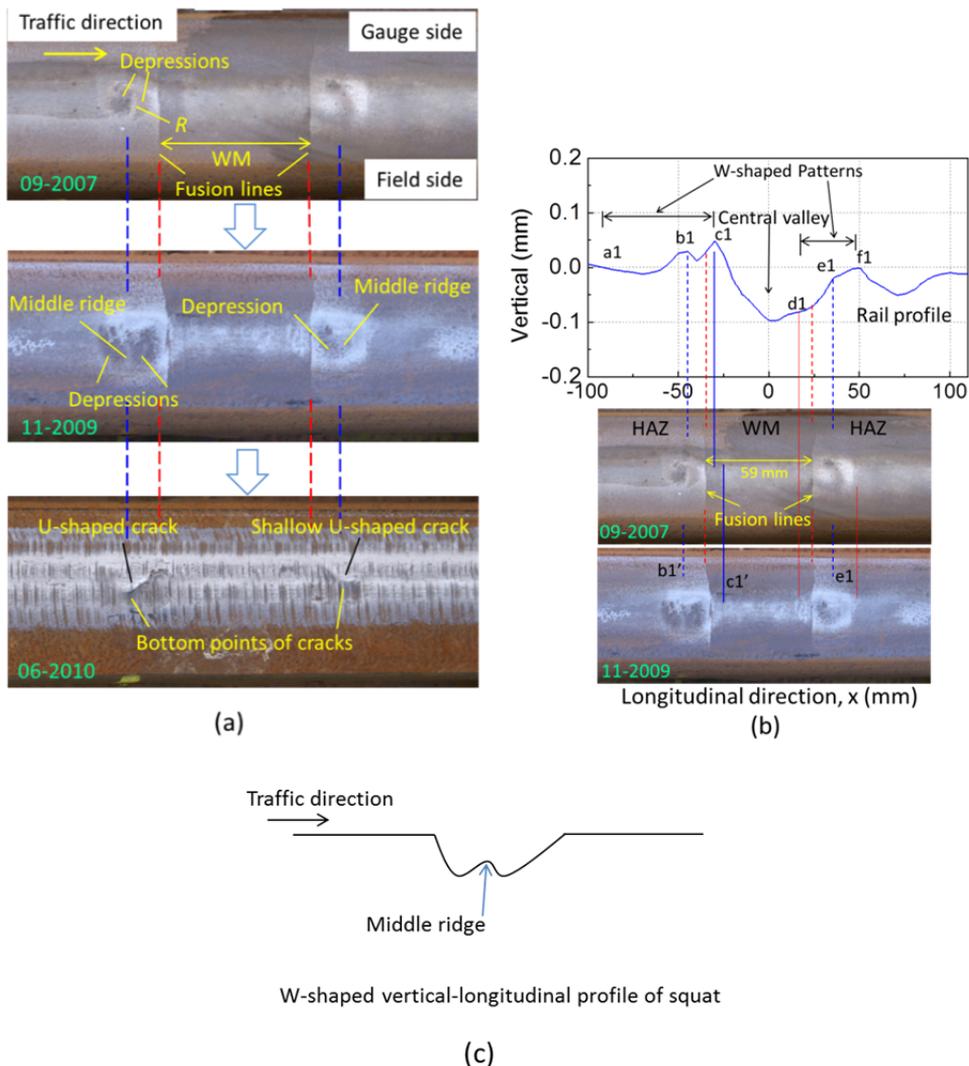


Fig. 5.2. Observations and measurements of a thermite weld with duo squats. Red dashed lines indicate fusion lines and blue dashed lines indicate the locations of middle ridges. (a) The squat in the first (09-2007), fifth (11-2009) and sixth (06-2010, after a rail grinding) observations; (b) Measured rail profile (the blue solid line) in the 09-2007 observation (Red solid lines align d1 and f1 with features in the photos to determine the dimension of the right squat); (c) Illustrative W-shaped vertical-longitudinal profile across a typical squat.

In the measured profile of Fig. 5.2(b), a clear W-shaped pattern was missing at the right squat. A peak, e1, however, existed at the location of the middle ridge. Looking at the photos, an indication of a dark depression can clearly be seen in the 11-2009 photo between d1 and e1. In view of this depression and the crack that had a size comparable to that of the left squat, it is reasonable to conclude that a V-shaped dip should exist between d1 and e1 in 09-2007.

Indeed, a shallow V-shape can be distinguished between d1 and e1 in the big central valley. A V-shaped dip should also be in formation between e1 and f1, as indicated by the less bright area between them in the 11-2009 photo. Thus, there was a W-shaped pattern at the right squat, too.

Further, a bright strip marked by *R* is visible in the left squat in the 09-2007 photo. It looks like a middle ridge but was not precisely at the location of the middle ridge b1' (see the 11-2009 photos). This bright strip would probably be caused by a varying hardness distribution, and it became erased or shifted to the location of the middle ridge during the growth of the squats, determined by the wavelength of the dynamic contact force, as will be discussed in Section 5.6.1.

At the time of 09-2007, cracks visible to the naked eye already existed in the surface of both squats. The cracks were in the V-dips between the ridges and fusion lines and had not developed into the V-dips on the far sides in 06-2010. It is therefore reasonable to conclude that these two V-dips on the far sides were not caused and influenced by the cracks that had developed by 09-2007. In Section 5.2.2, it will be shown that such V-shaped dips already exist before any visible surface cracks exist.

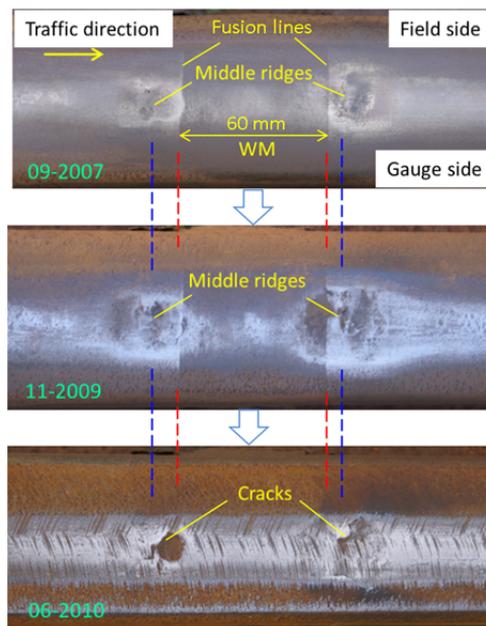


Fig. 5.3. Another example of thermite welds with duo squats in the first (09-2007), fifth (11-2009) and sixth (06-2010, after a rail grinding) observations, where the WM was approximately 60 mm in length. This is the same squat as that shown in Fig. 5.1(a).

Fig. 5.3 shows another instance of duo squats at a thermite weld from the same monitoring site at Steenwijk. The length of the WM is approximately 60 mm. Note, although this weld is located on the same track as the one in Fig. 5.2, the length of WM is different because of the inherent limitations of the welding process used and operator dependency [21]. The W-shaped

pattern was more evident than that in Fig. 5.2 at the time of the first observation. They continued to grow and became two typical squats by the time of the fifth observation. Cracks also grew with the development of the geometry.

It is not necessary to always have two squats simultaneously at a thermite weld. Single squats were also observed, such as the example in Fig. 5.4. At this thermite weld, the length of the WM looks to be much shorter than those in Fig. 5.2 and 5.3, so that only a single squat could occur. The length of WM and the characteristic length of the hardness variation at welds often determine whether one or two squats occur at a weld. This issue will be further discussed in Sections 5.2.3 and 5.6.



Fig. 5.4. One single squat at a thermite weld with shorter WM (from [15]).

Another explanation for having only one squat at a thermite weld is that the welding quality is good in the non-squatting HAZ of the weld, where there is no hardness variation or the magnitude of the variation is very small. The length of the WM in Fig. 5.5, for instance, was approximately 65 mm, which is not smaller than that in Figs. 5.2 and 5.3. However, only a single squat occurred, and there was no sign of squatting in the right HAZ. This observation thus inferred that the welding quality in the right HAZ was good, so that the resulting differential deformation and wear needed to fulfil the hypothesis in Section 5.3 was not large enough to induce a squat.

Indeed, many thermite welds do not cause squats. Thus, if the welding quality is well controlled, e.g., in terms of hardness variation, it is possible to greatly reduce squats at thermite welds.

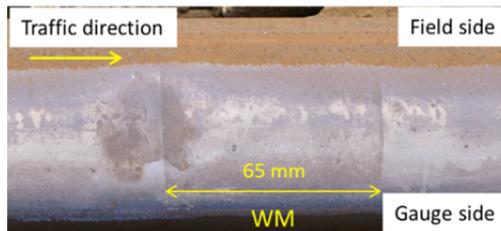


Fig. 5.5. Single squat at a thermite weld observed at Nijeveen in the Netherlands.

### 5.2.1.2. Squats at flash butt welds

Fig. 5.6(a) shows a flash butt weld with a single squat. In a flash butt weld, the HAZs are nominally symmetrical about the weld line (WL) because of the absence of WM [22]. The squat should have developed from a W-shaped hardness variation distribution, see Fig. 5.6(b) for a typical W-shaped hardness distribution of virgin flash butt welds, where each of the V-dips of the W corresponds to one of the HAZs. At the time of the first observation (09-2007), the W-shaped pattern of the squat was identifiable by the shining middle ridge that was along the weld line. The W-shaped pattern then grew to a two-lung-shaped squat with a much clearer middle ridge by the time of the fifth observation (11-2009).

A network of cracks can be seen after grinding in the 06-2010 photo, with a U-shaped crack (dashed line in the inset of Fig. 5.6(a)) on the gauge side, 2 transverse cracks (solid lines) along the ridge and a longitudinal crack (dotted line) on the field side. In this case, it is not clear whether it was the U-shaped crack or those along the ridge that were the primary cracks, as locations of both the U-shaped crack and of the ridge were shining in the 09-2007 photo, indicating large impacting stress, as discussed in Chapter 4. The two cracks along the middle ridge were relatively apart from each other, most probably due to the width  $d$  of the hardness plateau (ABC) in the middle of the weld (Fig. 5.6(b)). The width  $d$  is determined by the characteristic length of the hardness variation  $L_d$  that will be discussed in Section 5.2.3 with Fig. 5.9. Large sudden hardness change can be a source of cracking.

Rail surface evolves due to differential plastic deformation and wear; the corresponding locations of ABC might change and they became A'B'C' in the vertical-longitudinal rail profile in Fig. 5.6(c) measured at the weld in 09-2007. Its length  $d'$  is determined by the wavelength of the dynamic force, which is about 30 mm and is determined by the eigen characteristics of the local vehicle-track system, see [15, 23] and Section 5.3 of this chapter. The pattern of the measured profile followed the typical hardness distribution of Fig. 5.6(b), with a narrow central valley at B' corresponding to the hardness drop B at the WL in the middle of the hardness plateau ABC. As the distance of approximately 30 mm between A' and C' is much narrower than those of the thermite welds ( $> 45$  mm), there was only a single squat in this location.

So far in our observations, only single squats have been found at flash butt welds. This is unlike the cases for thermite welds where two squats were usually observed.

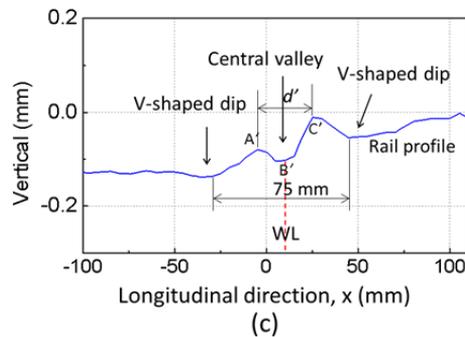
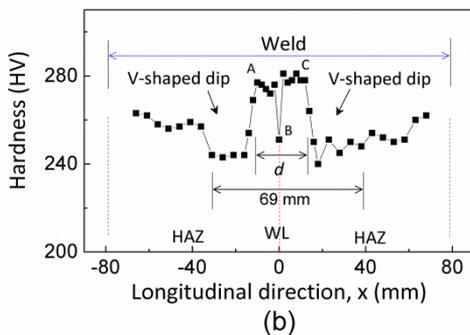
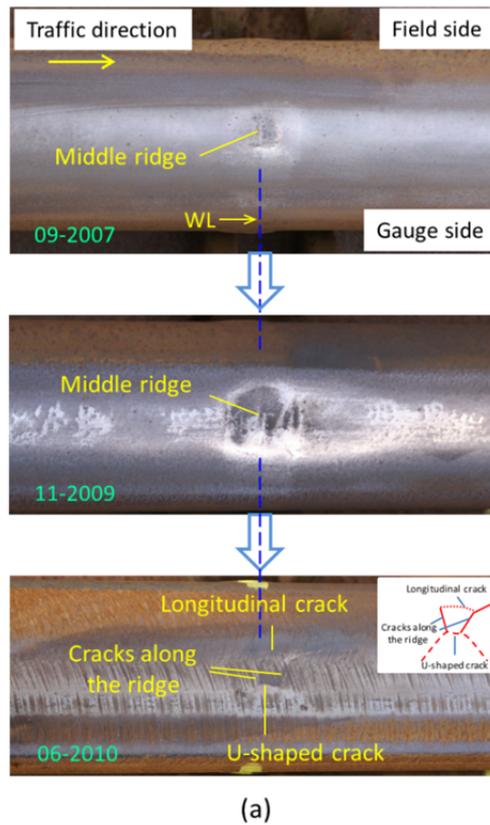


Fig. 5.6. (a) The growth process of a squat at a flash butt weld (with only one single squat). This is the same squat as that shown in Fig. 5.1(b). WL stands for Weld Line; (b) W-shaped hardness distribution measured at a virgin flash butt weld. The two V-shapes are on the 2 sides of the weld line; (c) Measured vertical-longitudinal rail profile in 09-2007. The length  $d'$  of  $A'B'C'$  corresponds to the width  $d$  in the hardness plateau  $ABC$  of (b), but they are not necessarily the same.

### 5.2.1.3. Summary

These field monitoring observations show that squats initiated at both thermite welds and flash butt welds. They should have developed from W-shaped pattern in the HAZs, and then grown into the typical two-lung shape with apparent cracks. Duo squats usually occurred at thermite welds, though cases of a single squat were also observed, probably depending on the length of the WM and the quality of the weld. Single squats were usually observed at flash butt welds, probably because the length between the two hardness dips of the two HAZs was too small to allow two squats to develop separately on the two sides of the WL.

### 5.2.2. Welds without cracking

In the aforementioned continual field monitoring, squats in the stages before crack initiation were not observed at welds. The profiles shown in Fig. 5.2(b) and Fig. 5.6(c) were measured when visible cracks were already present. To see the development of squats before crack initiation, an one-off field observation and measurement of several thermite welds were performed on the track at Zaandam in the Netherlands at the end of 2011. The surface irregularities of the examined welds are considered representative of pre-cracking squats, as they had in common the V-shaped dips in HAZs, which are presumably the initiation source of squats, as discussed in Section 5.2.1.

Fig. 5.7 shows two of the thermite welds and their measured profiles, respectively. The profiles are aligned with the welds. Two V-shaped dips (irregularities) in the HAZs and one central valley at the WM were observed. It is worth mentioning that the maximum depth of the irregularities might be larger than the measured depth because the profile was measured at the rail center, whereas the irregularities usually have an offset from the rail center because the running band is usually not precisely in the middle of rails. At the time of the observation, no cracks were found in the surface at the two welds, which indicates that surface irregularities at welds indeed exist before crack initiation.

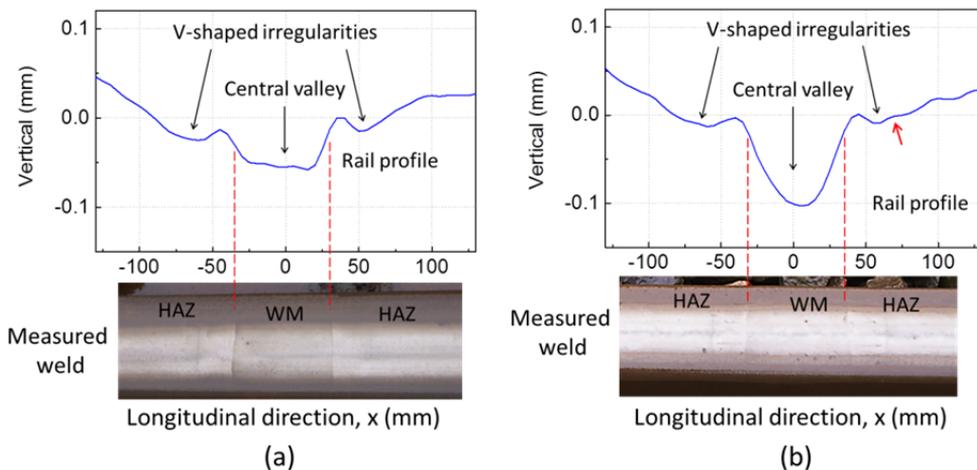


Fig. 5.7. Two thermite welds without cracks and their profiles. The length of the WM is approximately 64 mm in (a) and 67 mm in (b). If the small peak indicated by the red arrow is considered, the V-dip on the right side in (b) was evolving towards a W-shape.

Fig. 5.8 shows another thermite weld and its measured profile. No cracks were found in the rail surface. The profile was generally similar to those in Figs. 5.2 and 5.7, with a central valley and a V-shaped dip on each side. In Fig. 5.7(b) and 5.8, three of the V-shaped dips were developing towards W-shaped patterns.

Comparing the photos and measured profiles in Figs. 5.2, 5.7 and 5.8, it is observed that the severity of the surface irregularities is in the following decreasing order. (1) Fig. 5.2 is the most severe, with squats and cracks. Its left squat is more severe than the right, as judged from both the photos and the profile – the left squat already had a clear W-shaped profile and the W-shape of the right squat was not yet so clear. (2) Fig. 5.8 had clear and continuous waves on the rail surface, with indications in the profile that W-shaped patterns were in formation in both HAZs. (3) Fig. 5.7 did not have a clearly continuous wave pattern on the rail surfaces, with only one of the four HAZs showing an indication that W-shaped patterns were in formation. These consistent observations further support the inference that weld-induced squats initiate from V-shape without cracks, which further evolve into the W-shape and then the typical two-lung shape with cracks.

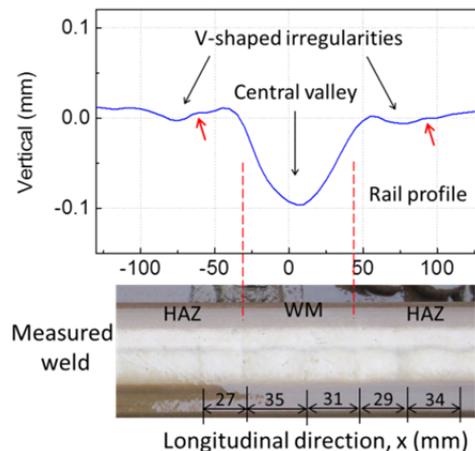


Fig. 5.8. Measured profile of a thermite weld without cracks: V-shaped irregularities were in both HAZs (the length of WM was approximately 73mm). If the 2 small peaks indicated by the 2 red arrows are considered, each of the V-dips was evolving towards a W-shape. Waves were visible on the rail surface with the wavelength indicated in the photo.

In the profile measurements of Figs. 5.2, 5.7 and 5.8, large central valleys existed at the WM. In practice, this is not necessarily always the case. Depending on the welding process and finish grinding, random surface geometry deviations are created. Bumps, instead of valleys, could occur at the WM, such as that shown in in Fig. 4 of [10]. In view of the fact that mature squats all have the characteristic two-lung shape, irrespective of such randomness, it is inferred that squats should have not been caused by the large central valleys or bumps at the WM and that surface deviations are not essential for the formation of squats. Instead, it should be the HAZ hardness distribution and the associated wavelength that are decisive. This issue will be examined with the help of numerical simulation in Sections 5.4, 5.5 and 5.6.

### 5.2.3. Hardness distribution at welds

Li et al. [15] proposed that the initiation of squats at welds is related to the hardness distribution. They also provided the general longitudinal distributions of the hardness in the HAZs for both thermite welds and flash butt welds, as schematically illustrated in Fig. 5.9(a). The hardness at welds varies because the welding material is different from the parent material and the welding process produces a lot of heat. The heat changes the hardness of the parent rail materials in the HAZs, mostly causing a lower hardness distribution within the HAZs. As a result, a spatially varying hardness distribution with two dips is introduced along the rail. This distribution may be characterized by the length  $L_d$  in Fig. 5.9(a), the distance between the two dips. This characteristic length is for thermite welds affected by the WM and HAZs, and it can be larger than 65 mm. This is because that the distance (length  $L$  in Fig. 5.9(a)) between the fusion line and the dip of the hardness is approximately 20 mm [24] on each side, and the WM can be longer than 25 mm, since the standard weld gap is 25 mm and wider gaps may occur [2]. The characteristic length of flash butt welds is  $L_d = 2L$  (Fig. 5.9(a)) and it is nominally 20 – 45 mm [22]. The measured hardness distributions for a virgin thermite weld

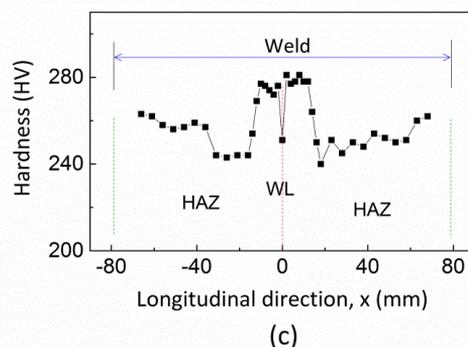
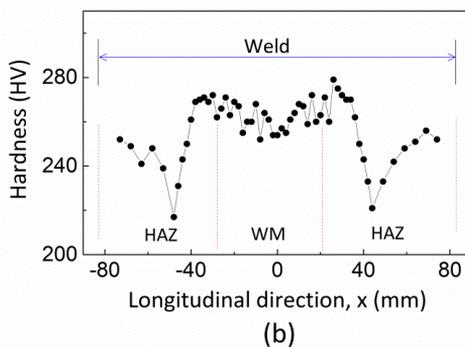
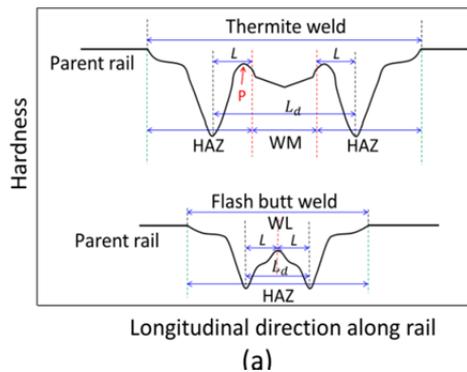


Fig. 5.9. Distributions of the hardness: (a) at a thermite weld and a flash butt weld, respectively (schematic diagram; the red dashed lines indicate the fusion lines); (b) at a virgin thermite weld (measured); (c) at a virgin flash butt weld (measured; this hardness distribution is the same as that in Fig. 5.6(b)).

and a virgin flash butt weld shown in Fig. 5.9(b) and Fig. 5.9(c) confirm these schematic distributions and nominal lengths of hardness variation.

The hardness dips in the HAZs correspond to the surface irregularities in the HAZs in Fig. 5.7 and Fig. 5.8 and at the squats in Fig. 5.2. This result shows that the formation of the initial surface irregularities and squats can be related to the hardness distribution. This issue will be analyzed in Sections 5.4 and 5.5.

Considering the discussions of sections 5.2.1, 5.2.2 and 5.2.3, the correlation between the hardness distribution and squat formation suggests that the frequent duo squats at thermite welds and single squats at flash butt welds should be mainly affected by the characteristic length  $L_d$  of the hardness variation, which is influenced by the lengths of the WM and HAZs. This issue will be further discussed in Section 5.6.

#### 5.2.4. Summary of the observations and measurements

Squats were observed at both thermite welds and flash butt welds. They initiated from V- and W-shaped surface irregularities, which were related to the varying hardness of HAZs. Other initial surface geometry deviations due to characteristics such as imperfect welding and finish grinding should not be essential for the formation of irregularities and, therefore, squats. Duo squats usually occurred at thermite welds due to a longer characteristic length of hardness variation, though single squats were also observed at times. Single squats were observed at flash butt welds due to the absence of WM and, thus, a shorter length of hardness variation. Thus, the number of squats (duo or single) caused by a weld is determined by the characteristic length of the hardness variation, which is related to the length of the WM and of the HAZs.

### 5.3. Hypothesis of squat formation at welds

Based on the field observations and measurements discussed above and considering the effect of the dynamic wheel-rail contact forces on squats development presented in [15], a formation process of squats at welds is hypothesized as shown in Fig. 5.10. It can be roughly divided into three steps according to the dominant driving mechanisms.

In the first step, surface irregularities initiate in the HAZs from the smooth rail surface through the accumulation of differential deformation and wear due to the varying material properties under cyclic loadings. The surface irregularities have a shape similar to that of the hardness distribution along the rail. Usually, a roughly V-shaped irregularity appears in each HAZ.

In the second step, a V-shaped irregularity excites a high dynamic contact force with a wavelength determined by the eigen characteristics of the local vehicle-track system and the length of the irregularity [15]. The dynamic force in turn produces further differential plastic deformation and wear so that the V-shaped irregularity develops into a W-shaped pattern, the wavelength of which is determined by that of the dynamic force [15]. The continuous development of the W-shaped patterns increase the dynamic force and the resulting stresses and strains at the same location, leading to a loop of positive feedback, with increasing W-shaped damage and progressive rail fatigue.

In the third step, crack initiation eventually occurs at the location where the values of the damage parameters are the highest, and then cracks propagate in the surface and into the rail.

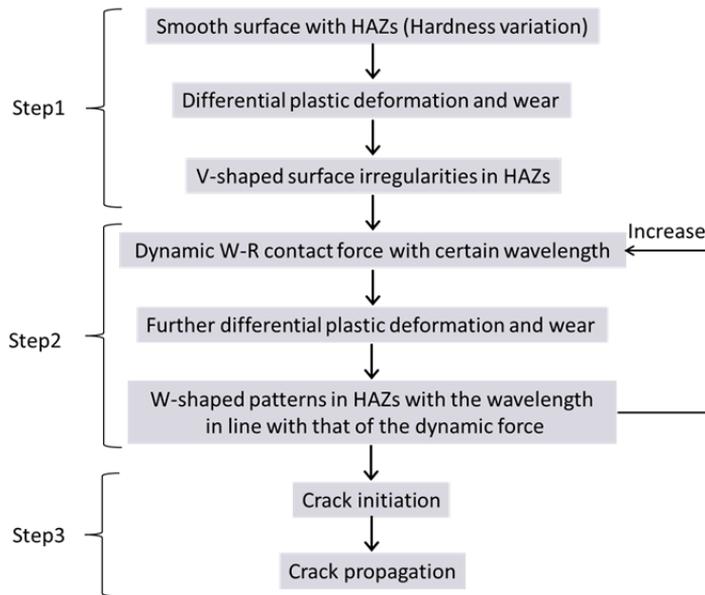


Fig. 5.10. Development process of squats at welds.

The distance between the two dips in the hardness distribution is 20 – 45 mm for flash butt welds, according to Section 5.2.3. This characteristic length is in close agreement with the typical wavelength of 20 – 40 mm of squats determined by the eigen characteristics of the local vehicle-track system [15]. The dynamic contact force excited by the differential plastic deformation and wear at one of the two dips can be in-phase with the differential plastic deformation and wear at the other dip. The two V-shapes at the two dips can combine with each other to form a W-shaped pattern that eventually becomes a single typical squat.

On the other hand, the characteristic length of the hardness variation at thermite welds is often greater than 65 mm. This length is much greater than the 20 – 40 mm wavelength of squats. As a result, the two V-shapes at the two dips cannot combine with each other to form a W-shaped defect. Instead, each V-shaped dip will develop into a W-shaped pattern, so that two W-shaped patterns are generated.

Thus, according to the hypothesis, the number of squats caused by a weld will be determined by the characteristic length of the spatially varying hardness distribution. If the characteristic length is much larger than the wavelength of typical squats, two squats may occur at a weld. This is frequently the case with thermite welds. Otherwise, only one squat occurs, for instance, at flash butt welds.

In this chapter, the numerical simulations in Sections 5.4 and 5.5 will verify this hypothesis of squat development before crack initiation, i.e., Step 1 and Step 2. The crack initiation and propagation will be studied in another investigation.

## 5.4. Numerical model

In this section, a 3D FE dynamic vehicle-track interaction model is applied to verify the hypothesis of the development of squats at welds due to differential plastic deformation and wear before crack initiation. The thermite weld is the focus.

### 5.4.1. FE model

A 3D FE dynamic vehicle-track interaction model is created using the ANSYS/LS-DYNA software to calculate the contact forces, the stresses and the deformation at a rail weld. A schematic view of the model is shown in Fig. 5.11(a). In this model, half of a typical wheelset and a straight track are considered. The primary vehicle suspension is modeled as a group of springs and dampers, and the vertical load from the carriage is imposed through a mass block. Between the rail and the sleepers, groups of springs and dampers are inserted to represent rail pads. Below the sleepers, the ballast is modeled as a group of springs and dampers. The stiffness and damping coefficients of these springs and dampers are taken from [25]. The static wheel load is 116.48 kN based on [25] including the 544 kg weight of the wheel. The standard wheel profile S1002 and a worn rail profile measured near the squat of Fig. 5.2 are employed. The nominal rail inclination of 1/40 is applied. The dynamic response arising from the vehicle-track interaction is considered. The accuracy of the approach is validated by Molodova et al. [26] when studying the detection of squats.

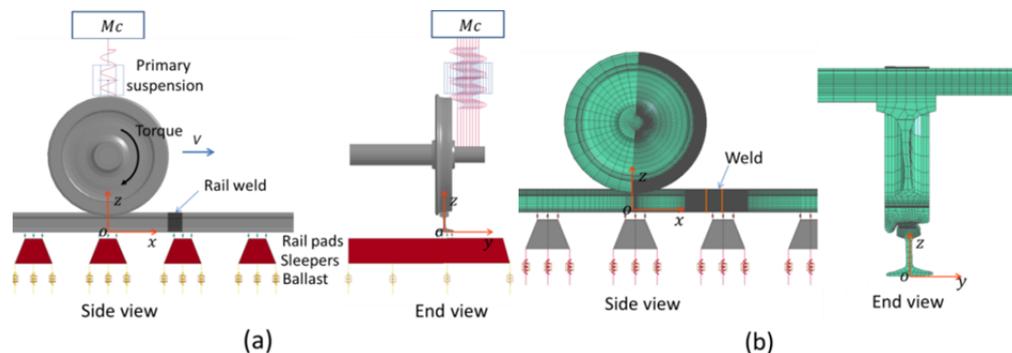


Fig. 5.11. FE vehicle-track interaction model: (a) FE model of wheel-rail rolling contact at a rail weld; (b) mesh of the model.

Contact between the wheel and the rail is determined by a surface-to-surface contact-searching scheme [27]. It is based on a master-slave algorithm. The kinematic contact condition is enforced by a penalty method [28]. The Coulomb friction law is employed. The accuracy of this calculation method for contact problems has been validated by Zhao et al. [29, 30] and in Chapter 3.

The 3D 8-node solid brick element is applied to discretize the wheel, the rail and the sleepers. To improve the accuracy of the stress calculation while maintaining a reasonable computational time, a non-uniform mesh is employed to both the rail and the wheel, see Fig. 5.11(b). The mesh fineness at the weld and at the part of the wheel that is in contact with the weld is approximately 1 mm per element. The Lagrangian dynamic non-linear analysis is

conducted using an explicit integration approach, as it is intrinsically suitable for simulating dynamic rolling contact processes [29].

A torque is applied to the axle of the wheel to produce traction to drive the wheel rolling. The traveling speed of the wheel is 140 km/h. The frictional coefficient  $f$  is 0.6, and the traction coefficient is 0.35 [19]. In this study, one wheel passage is simulated, and the resulting stresses and deformation in the rail at the weld are calculated and analyzed.

#### 5.4.2. Material properties at the weld

In this chapter, a thermite weld with the measured hardness distribution shown in Fig. 5.9(b) is analyzed. The hardness variation is distributed over a length of approximately 160 mm, with the distance between the two hardness dips of the HAZs being 92 mm. It thus should be sufficiently long to produce two squats at the two HAZs, according to the preceding discussion. The hardness distribution is measured to determine the yield stress by the formula below [31],

$$\sigma_y = \frac{H_V}{3} (0.1)^n \quad (5.1)$$

where  $H_V$  is the Vickers hardness,  $\sigma_y$  is the yield stress, and  $n$  is the strain hardening coefficient. The value of  $n$  for Grade 900A rail steel is 0.225, which is determined by the initial yield stress of 546 MPa and the corresponding hardness of 275 [32, 33]. The hardness is converted to the yield stress by equation (5.1). The converted yield stress is linearly interpolated at approximately 0.5 mm, then smoothed using the Adjacent-Averaging method with a window of 10 points. The resulting distribution of the yield stress is shown in Fig. 5.12(a) and serves as an input to the FE model (Fig. 5.12(b)). The middle of the distribution is at the rail position of 580 mm. This yield stress distribution is applied throughout the whole rail height. The hardness of the surface layers of both the parent rail and the wheel is set to  $280 H_V$  based on field measurements, and therefore the corresponding yield stress is 556 MPa. The initial yield stress of 546 MPa is applied to the subsurface of the wheel and rail.

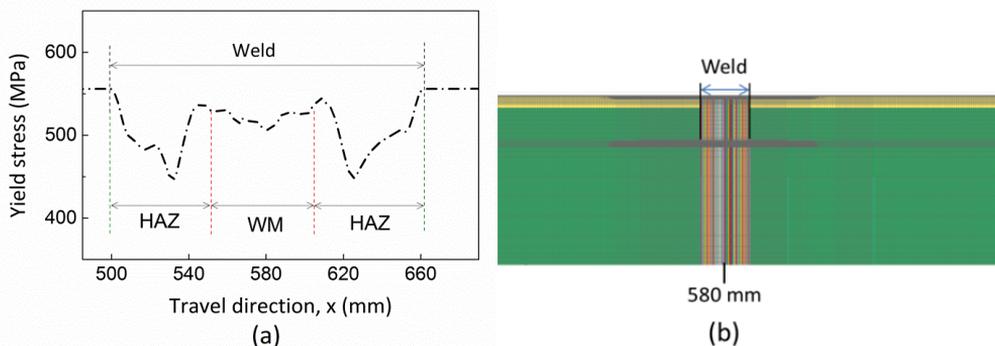


Fig. 5.12. Distribution of yield stress: (a) at the thermite weld derived from Fig. 5.9(b) and equation (1); (b) its implementation in the FE model. (The middle point of the varying yield stress is at the rail location of 580 mm.)

The non-linear isotropic/kinematic hardening model originally proposed by Lemaitre and Chaboche [34] is applied to both the wheel and the rail. In this model, five parameters are defined to characterize the material behavior: the Young's modulus  $E$ , the yield stress  $\sigma_y$ , the isotropic plastic hardening modulus  $H$ , the kinematic hardening modulus  $C$  and the kinematic hardening parameter  $\gamma$ . The yield stress is determined as shown above, and the other parameters are obtained according to [35, 36]. They are given in Table 1 and are applied to both the wheel and the rail for Steps 1 and 2, as will be discussed in Section 5.4.4.

Table 1. Material properties of the wheel and the rail.

| Components                 | Elastic modulus (GPa) | Poisson's ratio | Yield stress (MPa)            | $H$ (GPa) | $C$ | $\gamma$ |
|----------------------------|-----------------------|-----------------|-------------------------------|-----------|-----|----------|
| Surface of wheel & rail    | 210                   | 0.3             | 556                           | 41        | 58  | 10       |
| Subsurface of wheel & rail | 210                   | 0.3             | 546                           | 41        | 58  | 10       |
| Weld                       | 210                   | 0.3             | Varied according to Fig. 5.12 | 41        | 58  | 10       |

### 5.4.3. Surface deformation and wear

Plastic deformation and wear are the two direct causes of rail surface deformation [37], and thus the causes of the evolution of rail surface irregularities. The surface nodal permanent displacement is the plastic deformation. It is computed directly by the FE simulation of a wheel passage, and the wear is calculated based on the surface shear stress and micro-slip using the Archard wear model [38, 39] as follows.

The surface shear stress  $\tau_i$  and micro-slip  $v_i$  at each time step  $i$  are obtained from the FE simulation. According to the Archard wear model, a material removal of depth  $d$  due to wear is calculated for each node in the running band by:

$$d = \frac{K}{H_v f} \int_0^{t_0} \tau(t) ds = \frac{K}{H_v f} \int_0^{t_0} \tau(t) v(t) dt = \frac{K}{H_v f} \sum_0^M \tau_i v_i \Delta t \quad (5.2)$$

where  $t_0$  is the duration that a node passes a contact patch.  $\tau(t)$  is the shear stress at a node on the rail surface at time  $t$  and is updated after each time step in the solution procedure.  $ds$  is the increment of the sliding distance of the rail node relative to the wheel contact surface, and is calculated by multiplying the micro-slip  $v(t)$  by the time increment  $dt$ .  $\Delta t$  is the discretized form of  $dt$ .  $M$  is the total number of time steps.  $H_v$  is the material hardness defined in Section 5.4.2. The wear coefficient  $K$  is a constant, the value of which is taken as  $3.56e-4$  according to [40].

### 5.4.4. Simulation steps

It takes thousands of wheel passages to form a squat through cyclic plastic deformation and wear. Simulating all these wheel passages is too expensive in terms of computational resources. Therefore, two different rail surface geometries are simulated to verify the first two steps during the development process of squats hypothesized in Section 5.3.

In the first step, the rail surface is assumed to be smooth without any geometric irregularities but with measured spatially varying yield stresses at the weld, as described in

Section 5.4.2. The rolling of a wheel over a rail is simulated with the model and loading described above. It is expected that geometric irregularities will arise, due to the HAZs and the loading, on the rail surface after the wheel passage across the weld. These simulated surface irregularities are compared with field observations and measurements.

In the second step, the surface irregularities derived from the simulation in Step 1 are applied to the rail surface of the FE model to examine the effects of the dynamic force resulted from the irregularities and the trend of the V-shaped dips developing into W-shaped surface patterns at the HAZs. The varying yield stress distribution is assumed to be the same as that of Step 1. In this step, the fastening system is modeled by two groups of springs and dampers to include fastening degradation under wheel-rail interaction [41]. In Step 1, the fastening system is modeled by twelve groups of springs and dampers, ensuring more uniform contact between the rail bottom and the sleeper.

## 5.5. Simulation of rail profile evolution considering deformation and wear

### 5.5.1. Step1: Formation of V-dips from smooth rail with varying yield stress

#### 5.5.1.1. Calculated deformation and wear

In this section, the contact stress, plastic deformation and wear simulated at the thermite weld are presented for the smooth rail profile.

Fig. 5.13 shows the von Mises stress distribution when the wheel-rail contact patch starts to enter the left HAZ. Large stress exceeding the yield stress occurs at the rail surface. This finding suggests that plastic deformation primarily occurs at the rail surface. Therefore, the plastic deformation at the surface should be closely examined.

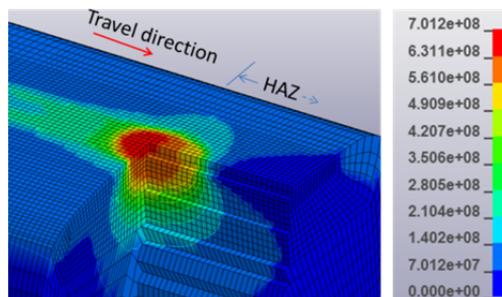


Fig. 5.13. Von Mises stress distribution when the contact patch is entering the left HAZ.

The calculated vertical plastic deformation  $dz$  and wear exhibit spatially (3D) differential distributions at the weld. Fig. 5.14 shows the 2D deformation and wear along the center of the running band. It is considerably greater in the HAZs than in the parent rail and the WM. The greatest value is approximately 50% larger than that in the parent rail surface. Plastic deformation also occurs in the WM. The differential deformation generally follows the variation of the hardness with the two V-shaped dips in the HAZs.

The varying yield stress also gives rise to differential wear distribution, as shown in Fig. 5.14. Its pattern, however, does not follow the variation of the hardness. For instance the

greatest wear is in the middle part of the left HAZ rather than at the location of the hardness dip. Therefore, there is a phase shift between the wear and the plastic deformation. A possible reason for this shift is that the micro-slip and surface shear stress that determine the wear are disproportionately affected by the varying hardness distribution, unlike the case for plastic deformation. To fully understand the root cause, further detailed study of the contact behaviors is needed.

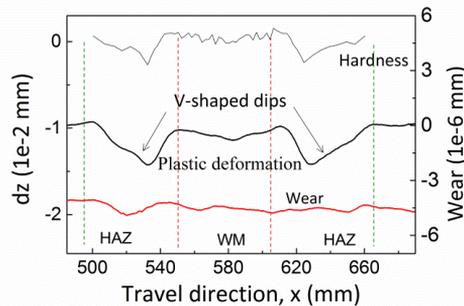


Fig. 5.14. Surface vertical plastic deformation  $dz$  and wear (average smoothing within a radius of 5.5 mm) of Step 1 along the center of the running band at the thermite weld (negative ordinates mean that the directions of deformation and wear are downwards relative to the initial (smooth) rail surface).

### 5.5.1.2. Deriving V-dip irregularities from deformation

Rail surface irregularities can be formed by superposition of the accumulated differential plastic deformation and wear. Here, they can be derived from the simulation results of plastic deformation and wear in Fig. 5.14. In the results, the magnitude of plastic deformation is much larger than that of the wear, and the wear does not follow the hardness distribution. This finding indicates that in the first contact cycles, plastic deformation would dominate the formation of the surface irregularities. With an increase in wheel passages, the plastic deformation per passage will decrease and shakedown will set in due to work hardening. This can be further investigated by simulation of multiple cycles of wheel passages to compare the accumulated plastic deformation and wear. In the present study, it is assumed that the generated irregularities follow the pattern of the plastic deformation. This assumption will be verified in Step 2 by checking whether the typical W-shaped pattern of squats could be formed based on this assumption. Thus there are two V-shaped irregularities due to the plastic deformation in the two HAZs. Additionally, the differential plastic deformation produces a slight central valley in the WM, corresponding to the hardness distribution in this area.

### 5.5.1.3. Comparing the simulation with field measurement

The derived geometric irregularities of the two V-shapes and central valley are in agreement with measured surface irregularities such shown in Fig. 5.7 and Fig. 5.8, especially that of Fig. 5.7(b), with which a comparison is made in Fig. 5.15. The occurrence of the two V-shaped irregularities due to the varying material properties in the HAZs and their correspondence to the variation of the hardness distribution confirm Step 1 of the hypothesis of Section 5.3.

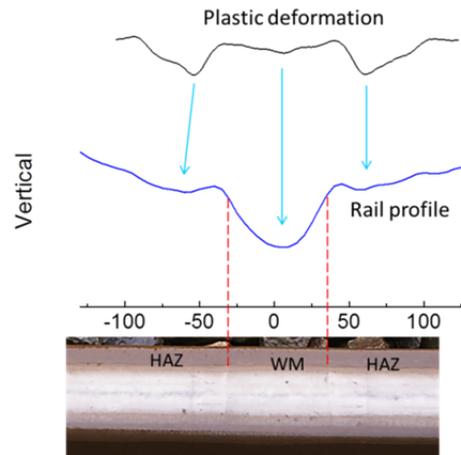


Fig. 5.15. Comparison of the calculated surface irregularities due to plastic deformation (upper curve) in Step 1 with the surface irregularities measured at the thermite weld of Fig. 5.7(b) (lower curve).

## 5.5.2. Step 2: Formation of W-shapes from V-dips with varying hardness

### 5.5.2.1. Generating rail surface with V-dips

At the beginning of Step 2, the surface irregularities due to plastic deformation derived in Step 1 are applied to the rail surface. The derivation is done by setting the largest depth of the left V-dip to 0.05 mm, which is close to the measurement in Fig. 5.2(b). This means that the deformation is scaled up by a factor of 10.4 to account for multiple wheel passages. The rest of the profile is set proportionally by the same scaling factor. The resulting rail surface is shown in Fig. 5.16 in 2D and 3D.

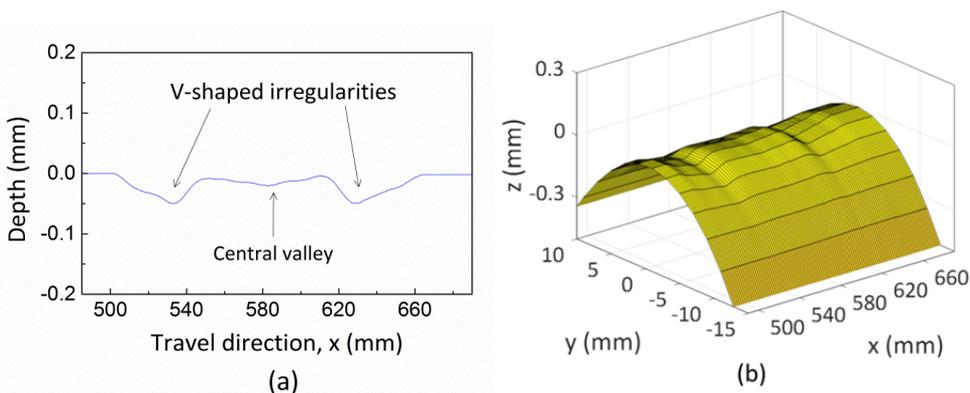


Fig. 5.16. Rail surface at the beginning of Step 2, obtained from the plastic deformation calculated in Step 1: longitudinal rail profile with two V-shaped irregularities at HAZs and one slight central valley at the WM. (a) Vertical-longitudinal profile and (b) the rail surface in 3D.

### 5.5.2.2. Formation of W-shaped patterns and comparison with field measurement

In Fig. 5.17(a) the plastic deformation (black line) and wear (orange line) calculated from the rail surface of Fig. 5.16 and average-smoothed within a radius of 5.5 mm are shown together with the start rail profile (dashed blue line) of Step 2, i.e., the profile of Fig. 5.16(a), along the center of the running band. Again the deformation and wear are differential, and the deformation per cycle is much larger than the wear. By multiplying the deformation by a factor of 10 to account for the effect of multiple wheel passages and adding it to the initial profile (i.e., Fig. 5.16(a)), an updated rail profile is obtained in Fig. 5.17(b). The following observations can be made from this profile:

- 1) The two V-shaped dips of Fig. 5.16(a) develop into two W-shaped patterns (Fig. 5.17(b)), with the shortest wavelength being 24 mm between  $b_2 - c_2$  and the longest being 38 mm ( $a_2 - b_2$ ). The central valley becomes deeper.
- 2) The wavelength of the W-shapes follows that of the longitudinal contact force, see Fig. 5.17(c), whereas the fluctuation of the vertical force follows the profile that excites it, and the wavelength of the vertical and longitudinal contact forces are different, as indicated by the vertical dashed lines in Fig. 5.17(c). These are in agreement with the findings of [23], where it was shown that the longitudinal and vertical forces of wheel-rail contact are not necessarily of the same wavelength. The vertical contact force has a strong dependence on the excitation, i.e., rail surface irregularities – rail corrugation, and the longitudinal contact force dominates the initiation of the corrugation. In [15, 16] and Chapter 4, it was pointed out that squats are related to short pitch corrugation.
- 3) Correspondence can be found between the simulated W-shaped patterns of Fig. 5.17(b), indicated by the peaks  $a_2 - f_2$ , and the measured W-shaped patterns with peaks  $a_1 - f_1$  of Fig. 5.17(d), though the wavelengths differ, with the wavelengths in the simulation being more uniform (the shortest is 24 mm and longest 38 mm). Among the measured wavelengths, the shortest is 15 mm between  $e_1 - f_1$ , and the longest is 47 mm between  $a_1 - b_1$ . This difference in wavelength will be discussed in Section 5.6.1. Note that the WM does not belong to the W-shapes.

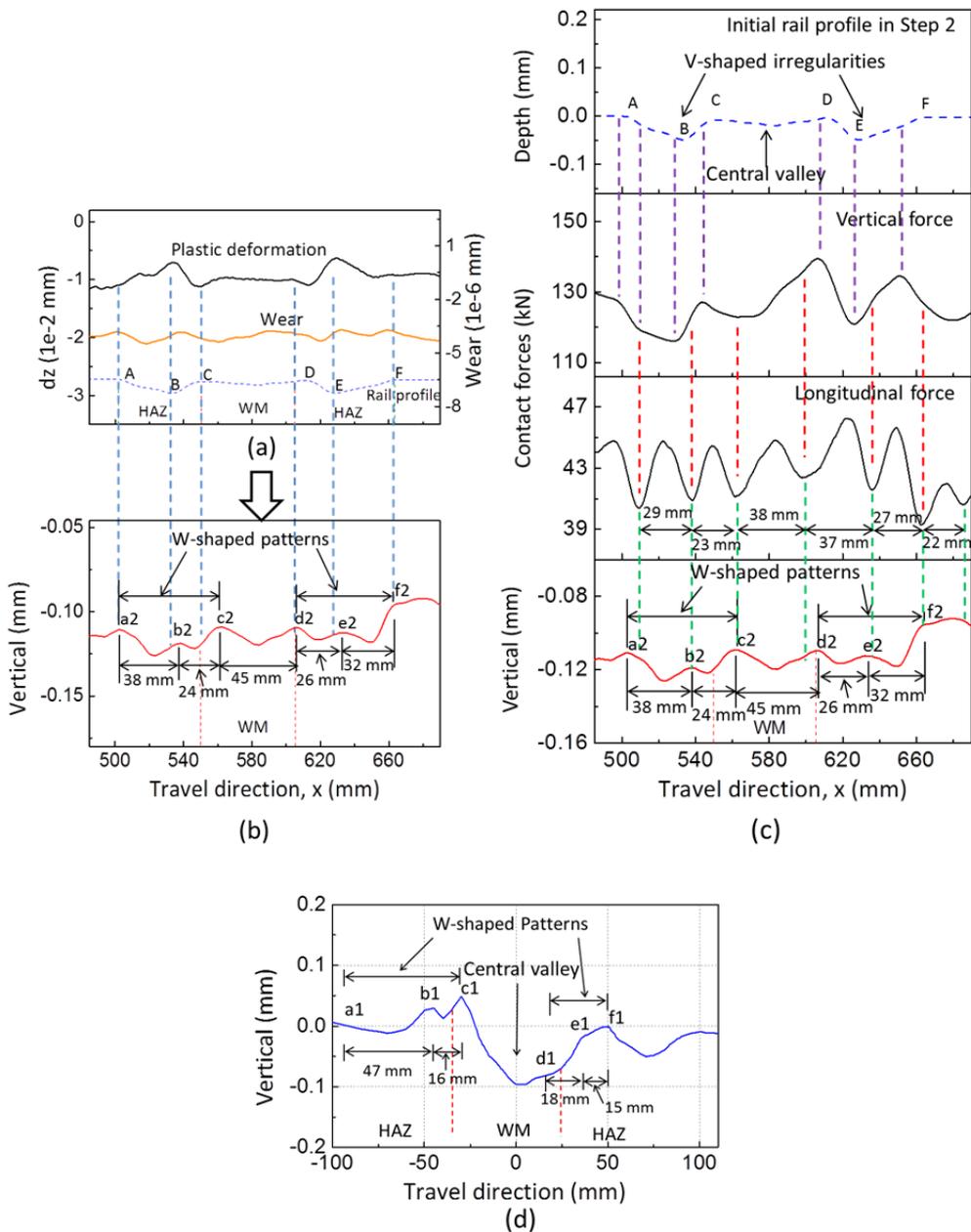


Fig. 5.17. Development of the profile along the center of the running band in Step 2 and comparison with a field measurement: (a) calculated plastic deformation and wear vs. rail profile at the beginning of Step 2; (b) the resulted profile (6 points average smooth) by adding 10 times of the plastic deformation to the initial profile of Step 2; (c) vertical and longitudinal contact forces vs. the excitation (initial rail profile in Step 2) and the resulted profile (the

lowest profile is the same as that in Fig. 5.17(b)); (d) measured profile at a weld (the same as Fig. 5.2(b)).

## 5.6. Discussions

### 5.6.1. W-shape is determined by the wavelength of the local dynamic system

The difference in the wavelengths between  $a1 - f1$  and  $a2 - f2$  can be explained by considering the irregularities  $a1 - f1$  as being in a stage different from that of  $a2 - f2$  in the development.

The irregularities started as V-shaped dips between ABC and DEF arising from a smooth rail surface, following the hardness variation in the HAZs (Fig. 5.17(a)). The V-shaped dips excite a dynamic longitudinal contact force of a certain wavelength determined by the local wheel-track system [23], causing differential deformation of the same wavelength. This deformation adds to the irregularities, causing them to evolve from V-dips to W-shapes, with new peaks arising in the initially large V-dips, as discussed for Figs. 5.2, 5.7 and 5.8 in Section 5.2.2, highlighted with the red arrows in Fig. 5.7(b) and in Fig. 5.8. In Fig. 5.17 the new peaks are  $b2$  and  $e2$ . The new peaks divide an initially large V-dip into two smaller V's to form a W-shape.

The new W-shapes are in phase with the differential plastic deformation in Fig. 5.17(a), indicating that the resulting geometric irregularities follow the differential plastic deformation. This fact also confirms the assumption in Step 1, wherein V-shaped irregularities follow the pattern of plastic deformation, as discussed in Section 5.5.1.2.

The differential deformation due to the longitudinal force is not necessarily in phase with the deformation caused by the hardness variation. Thus, the new peaks do not necessarily divide an initially large V into two equal small V's. This makes the wavelength of the initial W-shape non-uniform, as seen in Fig. 5.17(b), where it varies between 24 - 38 mm, and in Fig. 5.17(d), where it varied between 15 - 47 mm. This non-uniform wavelength could be considered the result of the variation in the uniform wavelength of the longitudinal contact force disturbed by the V-dips of the hardness variation. For instance, the wavelength of the longitudinal force, frequently embodied in the wavelength of the short pitch corrugation, is usually around 30 mm in the Dutch railway network [23]. This wavelength is disturbed when there are rail surface irregularities, as shown in Fig. 5.17. Another example of disturbed wavelength is in Fig. 5.8, where it is between 27 - 35 mm.

Once a W-shape exists, it also excite the local wheel-track dynamic system, causing deformation and wear of the same wavelength and phase as those of the W-shape. This process forms a loop of positive feedback that leads to dominance of the W-shape over the V-shape. The uniform wavelength of the longitudinal contact force and the dominance of the W-shape would then gradually change the initially non-uniform wavelength to be more or less uniform.

This change of wavelength could be seen in Fig. 5.2: a shining R-strip corresponding to the position  $b1$  existed in the 09-2007 photo. It could have occurred due to a hardness jump next to the fusion line, as indicated by the  $P$  arrow in Fig. 5.9, that caused differential deformation. The shining brightness of the R-strip and the depressions on the two sides of it indicate impact-like wheel-rail contact [15]. This dynamic interaction obviously widened the

depression between b1c1 in both directions into the wider b1'c1' in 11-2009. Thus, the wavelength of b1c1 was extended. Gradually the width of the two dark depressions, i.e., the wavelength of the two V's in a W-shape, of a squat becomes approximately equal, as seen in the mature squats of Fig. 5.1 – 5.6.

In the development process of the two initial V-dips in the two HAZs of a weld, they have a similar process but possibly different growth rates and different morphologies. In the beginning, the two V-dips are roughly symmetrical about the WM. Then they simultaneously develop into W-shapes with two V-dips. The initial W-shapes are non-uniform because of the disturbed wavelength of dynamic forces by the initial V-dips. Therefore, the morphology of the initial W-shapes could be different. One may have a larger dimension at the front V-dip, while another one may have a larger dimension at the rear V-dip. Moreover, the two W-shapes could grow at different rates because of variations in the magnitudes of dynamic longitudinal contact forces. These forces depend on the size of the initial V-dips and local vehicle-track system. As a result, one W-shape could be larger than the other in dimension, e.g., the left squat in Fig. 5.2 looks larger than the right one. Eventually, however, all V-dips tend to become similar over time, and the two squats should become similar in both dimension and morphology.

### 5.6.2. The presence of single or duo squats is determined by length of welds

A flash butt weld usually has a shorter characteristic length of hardness variation. As a result, the two dips of the resulting surface irregularities are much closer than are those at thermite welds. This result can be confirmed by the differential plastic deformation in Fig. 5.18 calculated from the hardness distribution of Fig. 5.9(c). The distance between the two dips is approximately 40 mm, which is within the wavelength of 20 – 40 mm of typical squats. Thus the two dips combine to form one single W-shaped pattern and, eventually, a single squat.

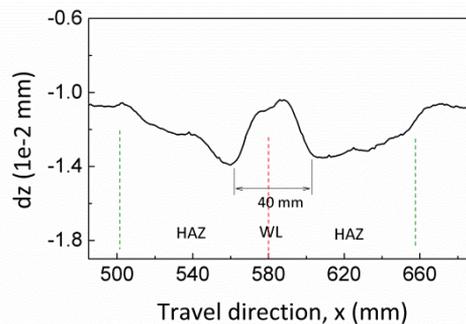


Fig. 5.18. Plastic deformation (average smoothing within a radius of 5.5 mm) along the center of the running band for a flash butt weld with the hardness distribution of Fig. 5.9(c).

Similarly, if the characteristic length of the hardness variation in a thermite weld is short, a single squat is likely to occur, such as the squat shown in Fig. 5.4.

### 5.6.3. Occurrence or not of squats is determined by both hardness and dynamic contact forces

Occurrence or not of squats at a weld is determined by both welding quality and the eigen characteristics of the local vehicle-track system. If a weld is in good condition where there is no hardness variation or the magnitude of hardness variation is very small in the HAZs, then V-shaped irregularities and subsequent squats are unlikely to take place. This could be the case in the right HAZ of the weld in Fig. 5.5, where no squat was observed. Even if V-shaped dips are formed due to hardness variation, they can develop into squats only if there exist dynamic forces with a certain wavelength determined by the local wheel-track system. In practice, the quality of welding varies from one weld to another due to inherent limitations of the welding processes and their dependency on human operators [21]. Moreover, the characteristics of the local vehicle-track system are determined by the quality of both tracks and vehicles. Therefore, only the welds that fulfill both conditions at the same time can develop into squats. As both conditions are stochastic and independent from each other, their combination limits the percentage of welds that can develop into squats. However, even a small percentage of welds potentially develop into squats, they significantly affect the safety of railway systems.

### 5.6.4. Summary

Field observations and measurements, as well as numerical simulations, show that weld-induced squats started from V-shaped dip plastic deformations that followed the hardness distribution at the HAZs. The deformations caused a dynamic longitudinal contact force and the resulting differential plastic deformation. The latter added onto the V-shaped dips to cause the initiation and growth of W-shaped irregularities that eventually developed into a typical two-lung-like squat. Each of the lungs was a depression, corresponding to one of the two V's of the W-shape.

Since the hardness-variation induced rail surface irregularities and the resulting dynamic contact forces are the precursor to the squats, it is possible to use a dynamics-based method to detect the squats at an early stage, or even before any fatigue cracks occur. The characteristic wavelength can be used as a signature for the detection and monitoring [26, 42].

## 5.7. Conclusions

Five-year continual field monitoring observations and measurements have been performed to investigate the initiation and development of squats at rail welds. Based on these, a three-step hypothesis of the development process of squats at welds is proposed. Steps 1 and 2 are the pre-cracking process, and Step 3 is the post-cracking process. The first two steps are the focus of this chapter. Numerical simulations using a 3D FE dynamic vehicle-track interaction model have been performed by considering the varying yield stress based on measured hardness and surface irregularities. The numerical results are in agreement with the field observations, verifying the first two steps of the hypothesis. The conclusions are as follows.

- 1) Squats at welds developed from surface irregularities without cracks.
- 2) In Step 1, initial surface V-shaped irregularities are predominately caused by differential plastic deformation that follows the varying hardness distribution in the

HAZs. Imperfect finish grinding of welding is not necessary the cause of squats, although it could produce geometric deviation, promoting the development of squats.

- 3) In Step 2, the V-shaped irregularities causes a dynamic longitudinal contact force and the resulting differential plastic deformation. The latter adds onto the V-shaped irregularities to cause the initiation and growth of W-shaped irregularities that eventually develop into typical two-lung-like squats. Each of the lungs is a depression, corresponding to one of the two V's of the W-shape.
- 4) In this process, the role of the varying hardness distribution is to cause the initial V-shaped rail surface irregularities. The irregularities subsequently cause dynamic longitudinal contact force which determines the W-shape and the eventual typical two-lung shape of mature squats. Here, the dominant characteristic is the wavelength of the longitudinal contact force, which is determined by the local wheel-track system. From the numerical simulation and field observation, this characteristic wavelength of the contact force of the squats appears to be the same as that of the short pitch corrugation.
- 5) The development of the W-shaped irregularities continuously increases the dynamic contact force at the same location, resulting in a positive-feedback growth loop of the irregularities and the dynamic force, leading to rail fatigue and cracks.
- 6) The numerical simulations showed that the irregularity magnitude caused by plastic deformation in a wheel passage is much larger than that caused by wear. This finding needs further investigation to verify. Besides, hardening will occur under repeated wheel-rail rolling contact. As a consequence, hardness and yield stress will evolve over time. The effects of this evolution on the development of surface irregularities is another topic for further research. Further research should also investigate whether white etching layers occur on rail surface around welds and how they influence the hardness variation since they are sometimes observed on rail surfaces, e.g., [43]. Moreover, the material of WM is considered the same type as the parent rail steel. Further work is needed to investigate the effects of more material parameters, such as elastic modulus, plastic hardening modulus and  $n$  in equation (5.1).
- 7) The number of squats induced by a weld is determined by the characteristic length of the hardness variation or the quality of welding. Duo squats are frequently observed at thermite welds because the characteristic length is usually much larger than the typical wavelength of squats, i.e., 20 – 40 mm. A single squat usually occurs at flash butt welds because the characteristic length is in close agreement with that of typical squats. Single squats occurs at some thermite welds due to short characteristic length or good quality of welding, indicating that by improving welding quality, weld-induced squats can be reduced or even avoided.
- 8) Since the hardness variation-induced rail surface irregularities and the resulting dynamic contact force are the precursors to the squats, it is possible to use a dynamics-based method to detect squats at an early stage or even before any fatigue

cracks occur. The characteristic wavelength can be used as a signature for detection and monitoring.

## References

- [1] J. Myers, G. Geiger, D. Poirier, Structure and Properties of Thermite Welds in Rails, *WELDING J.*, 61 (1982) 258-268.
- [2] P. Micenko, H. Li, Double dip hardness profiles in rail weld heat-affected zone: literature and research review report, Brisbane, Australia, (2013).
- [3] M.J.M.M. Steenbergen, C. Esveld, Relation between the geometry of rail welds and the dynamic wheel - rail response: Numerical simulations for measured welds, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 220 (2006) 409-423.
- [4] N. Ilić, M.T. Jovanović, M. Todorović, M. Trtanj, P. Šaponjić, Microstructural and mechanical characterization of postweld heat-treated thermite weld in rails, *Materials characterization*, 43 (1999) 243-250.
- [5] A. Skyttebol, B.L. Josefson, J.W. Ringsberg, Fatigue crack growth in a welded rail under the influence of residual stresses, *Engineering Fracture Mechanics*, 72 (2005) 271-285.
- [6] G.T. Fry, F.V. Lawrence, A.R. Robinson, A model for fatigue defect nucleation in thermite rail welds, *Fatigue & Fracture of Engineering Materials & Structures*, 19 (1996) 655-668.
- [7] H. Desimone, S. Beretta, Mechanisms of mixed mode fatigue crack propagation at rail butt-welds, *International Journal of Fatigue*, 28 (2006) 635-642.
- [8] Z. Wen, G. Xiao, X. Xiao, X. Jin, M. Zhu, Dynamic vehicle-track interaction and plastic deformation of rail at rail welds, *Engineering Failure Analysis*, 16 (2009) 1221-1237.
- [9] W. Li, G. Xiao, Z. Wen, X. Xiao, X. Jin, Plastic deformation of curved rail at rail weld caused by train-track dynamic interaction, *Wear*, 271 (2011) 311-318.
- [10] J. Gao, W. Zhai, Y. Guo, Wheel-rail dynamic interaction due to rail weld irregularity in high-speed railways, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 232 (2018) 249-261.
- [11] M. Sh. Sichani, Y. Bezin, Differential wear modelling – Effect of weld-induced material inhomogeneity on rail surface quality, *Wear*, 406-407 (2018) 43-52.
- [12] A. Al-Juboori, D. Wexler, H. Li, H. Zhu, C. Lu, A. McCusker, J. McLeod, S. Pannil, Z. Wang, Squat formation and the occurrence of two distinct classes of white etching layer on the surface of rail steel, *International Journal of Fatigue*, 104 (2017) 52-60.
- [13] D. Cannon, K.O. EDEL, S. Grassie, K. Sawley, Rail defects: an overview, *Fatigue & Fracture of Engineering Materials & Structures*, 26 (2003) 865-886.
- [14] P. Clayton, M. Allery, Metallurgical aspects of surface damage problems in rails, *Canadian Metallurgical Quarterly*, 21 (1982) 31-46.
- [15] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats—Correlation analysis and numerical modeling, *Wear*, 265 (2008) 1349-1355.
- [16] Z. Li, R. Dollevoet, M. Molodova, X. Zhao, Squat growth—Some observations and the validation of numerical predictions, *Wear*, 271 (2011) 148-157.
- [17] Z. Li, Squats on railway rails, *Wheel-Rail Interface Handbook*, Woodhead Publishing, 2009, pp. 409-436.

- [18] Z. Wei, A. Núñez, Z. Li, R. Dollevoet, Evaluating degradation at railway crossings using axle box acceleration measurements, *Sensors*, 17 (2017) 2236.
- [19] Z. Li, X. Zhao, R. Dollevoet, An approach to determine a critical size for rolling contact fatigue initiating from rail surface defects, *International Journal of Rail Transportation*, 5 (2017) 16-37.
- [20] Z. Li, X. Zhao, R. Dollevoet, M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, *Vehicle System Dynamics*, 46 (2008) 237-246.
- [21] P. Mutton, E. Alvarez, Failure modes in aluminothermic rail welds under high axle load conditions, *Engineering Failure Analysis*, 11 (2004) 151-166.
- [22] EN 14587-2. Railway Applications—Track-Flash Butt Welding of Rails—Part 2: New R220, R260, R260Mn and R350HT Grade Rails by Mobile Welding Machines at Sites Other Than a Fixed Plant, European Committee for Standardization, Brussels, 2009.
- [23] S. Li, Z. Li, A. Nunez Vicencio, R. Dollevoet, New insights into the short pitch corrugation development enigma based on 3D-FE dynamic vehicle-track coupled modelling in frictional rolling contact, *Applied Sciences (Switzerland)*, 7 (2017).
- [24] K. Saita, K. Karimine, M. Ueda, K. Iwano, T.Y.K. Hiroguchi, Trends in rail welding technologies and our future approach, *Nippon steel & sumitomo metal technical report*, (2013) 84.
- [25] X. Zhao, Z. Li, J. Liu, Wheel-rail impact and the dynamic forces at discrete supports of rails in the presence of singular rail surface defects, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226 (2011) 124-139.
- [26] M. Molodova, Z. Li, A. Núñez, R. Dollevoet, Validation of a finite element model for axle box acceleration at squats in the high frequency range, *Computers & Structures*, 141 (2014) 84-93.
- [27] D.J. Benson, J.O. Hallquist, A single surface contact algorithm for the post-buckling analysis of shell structures, *Computer Methods in Applied Mechanics and Engineering*, 78 (1990) 141-163.
- [28] J. Hallquist, G. Goudreau, D. Benson, Sliding interfaces with contact-impact in large-scale Lagrangian computations, *Computer Methods in Applied Mechanics and Engineering*, 51 (1985) 107-137.
- [29] X. Zhao, Z. Li, The solution of frictional wheel-rail rolling contact with a 3D transient finite element model: Validation and error analysis, *Wear*, 271 (2011) 444-452.
- [30] X. Zhao, Z. Li, A three-dimensional finite element solution of frictional wheel-rail rolling contact in elasto-plasticity, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 229 (2015) 86-100.
- [31] J. Cahoon, W. Broughton, A. Kutzak, The determination of yield strength from hardness measurements, *Metallurgical and Materials Transactions B*, 2 (1971) 1979-1983.
- [32] C. Tapp, *Mechanische Eigenschaften von Schienenstählen nach zyklischer Verformung*, VDI-Verlag, 2005.
- [33] C. Esveld, *Modern Railway Track*, MRT-Productions, The Netherlands, 2001.

- [34] J. Lemaitre, J.-L. Chaboche, *Mechanics of solid materials*, Cambridge university press, 1994.
- [35] E. Kabo, A. Ekberg, P.T. Torstensson, T. Vernersson, Rolling contact fatigue prediction for rails and comparisons with test rig results, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 224 (2010) 303-317.
- [36] G. Johansson, J. Ahlström, M. Ekh, Parameter identification and modeling of large ratcheting strains in carbon steel, *Computers & Structures*, 84 (2006) 1002-1011.
- [37] U. Olofsson, R. Nilsson, Surface cracks and wear of rail: A full-scale test on a commuter train track, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 216 (2002) 249-264.
- [38] J. Archard, Contact and rubbing of flat surfaces, *Journal of applied physics*, 24 (1953) 981-988.
- [39] J. Perez, J. Sinclair, J. Tunna, A review of wheel wear and rolling contact fatigue, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 221 (2007) 271-289.
- [40] U. Olofsson, T. Telliskivi, Wear, plastic deformation and friction of two rail steels—a full-scale test and a laboratory study, *Wear*, 254 (2003) 80-93.
- [41] X. Zhao, Z.L. Li, R. Dollevoet, Influence of the Fastening Modeling on the Vehicle-Track Interaction at Singular Rail Surface Defects, *Journal of Computational and Nonlinear Dynamics*, 9 (2014).
- [42] M. Molodova, Z.L. Li, A. Nunez, R. Dollevoet, Automatic Detection of Squats in Railway Infrastructure, *Ieee T Intell Transp*, 15 (2014) 1980-1990.
- [43] A. Kumar, G. Agarwal, R. Petrov, S. Goto, J. Sietsma, M. Herbig, Microstructural evolution of white and brown etching layers in pearlitic rail steels, *Acta Materialia*, 171 (2019) 48-64.

# 6

## **CONCLUSIONS, APPLICATIONS AND RECOMMENDATIONS**

## 6.1. Conclusions

Chapters 2-5 study the performance of an FE explicit method for rolling contact, and analyze the formation and development process of squats through both field monitoring and numerical analysis. These works address the research questions in Chapter 1.

*Q1: What is the applicability of half-space-based methods for contact problems with respect to the significant dimensions?*

Chapter 2 provides the answer to question Q1. In this chapter, the FE method was used to assess the applicability of half-space-based methods to static normal contact problems. Different combinations of significant dimensions of the contact bodies in terms of height, width, and length were studied. The scenarios considered effectively covered the entire range of interest from the nearly smallest possible significant dimension (1.1x the characteristic size of the contact patch) to infinite dimension. Various contact patch eccentricities were considered. Both cylindrical and non-cylindrical contact geometries were examined. The following main conclusion was drawn:

- Provided the significant dimensions of the contact bodies are 1.1x larger than the characteristic size of the contact patch, half-space-based methods can deliver reasonably accurate solutions with an error within 9% for most non-conforming elastic contact problems.

*Q2: How well does the explicit FE method work in frictional rolling contact with large spin? What are the advantages of the explicit FE method in rolling contact with spin?*

Chapter 3 addresses question Q2. The 3D explicit FE method based on a Lagrangian formulation was applied to spin-rolling contact. An explicit time integration scheme and a penalty contact algorithm are employed in this method. The accuracy of this method was demonstrated through a detailed analysis and a comparison with the solution of the Hertz theory and the Kalker's variational theory. Moreover, the effect of plastic deformation was investigated by comparing the elasto-plastic solution with the elastic solution. The conclusions are as follows:

- This explicit approach is able to address systems with an arbitrary geometry and continuum dynamics, in which large geometric nonlinearity and non-linear material properties, such as plasticity, can be considered.
- When large geometric spin is taken into account, a quasi-static solution can be obtained through dynamic relaxation in the explicit FE modeling.
- The FE solutions are well consistent with the Hertz solution and the Kalker's solution.
- Some fluctuations in stress and micro-slip are observed in the FE solutions. These fluctuations occur because the moving wheel excites physical vibrations in the system and the system is insufficiently relaxed to a steady state. Such high-frequency vibrations can be damped out to a certain degree by plastic deformation.
- In frictional rolling contact problems with large geometric spin, the effect of the tangential contact stress on the normal solution is negligible in the elastic case. However, this effect may need to be considered in the elasto-plastic case.
- Spin could change the distributions of stick and slip. It can also affect the distribution of surface tangential tractions.

- Plasticity has a strong effect on the size and shape of the contact patch, the areas of stick and slip, and the magnitudes of the pressure, shear stress and von Mises stress. Plasticity can lead to the shakedown of stress after repeated wheel passages. It can also cause contact geometry to become more conforming. An analysis of rolling contact with plasticity can provide more realistic insights.
- When spin is considered, the stress history of the critical point is different from that obtained in the case of the Hertz contact in two dimensions and that in frictional contact without spin.

*Q3: What are the fatigue crack initiation location and cracking behaviors at rail squats? How does the geometry of squats evolve? What are the causes of the crack initiation and propagation, as well as the evolution of geometry?*

Chapter 4 aims to address question Q3. In this chapter, a five-year continual field monitoring was analyzed for squats induced by corrugations. This monitoring has captured various stages of the life cycle of squats, from small black depressions without cracks to typical mature two-lung-like squats typically accompanied by Y-shaped cracks. A squat development process was proposed based on the observations and analysis of 30 squats, as well as the mechanisms of the crack initiation, crack propagation and geometry evolution. The conclusions are as follows:

- Short pitch corrugation can lead to the formation of the squats. The cracks initiate in the surface or at least within the top layer of rail material not deeper than 1.1 mm.
- A squat induced by short pitch corrugation starts from a small elliptical black depression at the corrugation. The small black depression is formed by the local collapse of a network of microcracks resulting from plastic deformation and work hardening due to impact forces.
- When the black depression grows to approximately 8 mm in size with the maximum distance of approximately 6 mm between the depression edge on the gauge side and the rail center, the large tensile stress initiates the first macrocrack at this point of maximum distance from the depression. These 8 mm corresponds to the critical size for squat initiation proposed in [1], derived by comparing the maximal von Mises stress with tensile strength, the material being R260Mn. The crack initiation location of 6 mm from the rail center could depend on the contact geometry of wheels and rails, the local track conditions and the material property.
- These macrocracks propagate along the border of the main contact area in both the traffic direction and the opposite-traffic direction, forming a U-shaped primary crack in/near the rail surface on the gauge side edge, with the U facing toward the gauge side. The crack initiation point is at the bottom of the U-shaped crack. The crack also propagates into the rail toward the field side at an angle of approximately 20°. The primary crack grows more rapidly in the traffic direction when traction force is applied.
- After crack initiation, the squat geometry could not remain fully elliptical. Its gauge side edge became bordered by the U-shaped crack, while the part without cracking remains elliptical, causing the squats at this stage to take on a kidney-like shape. With further growth of the squats along the two arms of the U in the rail surface and the propagation deeper into the rail, the primary cracks gradually form a peninsula-shaped part that protruded from its depressed surrounding materials. The peninsula-shaped part forms an obstacle to the rolling wheels, causing two impact forces, the first arises when wheels

impact the peninsula-shaped part and the second arises when the wheels fall from the peninsula-shaped part onto the leading edge of the squats. This outcome corresponds to the squat growth process described in [1-4].

- These impact forces, especially the first one, leads to a ridge in the middle of the depressed black part of the squats, where a secondary I-shaped crack appeared subsequently. Together with the primary U-shaped crack, it constitutes a Y-shaped crack. This Y-shaped crack divides the black depression into two parts so that the typical two-lung shape of mature squats is formed.
- A further tertiary semi-circular crack could occur along the elliptical edge of the depression on the opposite side of the primary U-shaped crack due to sinking of the main contact area.
- The change in the rail inclination could shift the wheel-rail impact region so that the primary U-shaped cracks were occasionally found on the field side, and the corresponding tertiary semi-circular crack could be found on the gauge side.
- In general, the opening of the primary U-shaped crack, regardless of whether on the gauge or field side, is opposite to the black depression. The opening of the tertiary semi-circular cracks faces the black depression.

*Q4: What cause the rail surface irregularities that develop into squats? How do the rail surface irregularities develop before crack initiation? How do the mechanisms produce squats in term of not only dynamic forces but also differential plastic deformation and wear?*

Chapter 5 aims to address question Q4. The five-year continual field monitoring observations and measurements were analyzed to investigate the initiation and development of squats at rail welds. Based on these, a three-step hypothesis of the development process of squats at welds is proposed. Steps 1 and 2 are the pre-cracking process, and Step 3 is the post-cracking process. The first two steps are the focus of the present work. Numerical simulations using a 3D FE dynamic vehicle-track interaction model have been performed by considering the varying yield stress based on measured hardness and surface irregularities. The numerical results are in agreement with the field observations, verifying the first two steps of the hypothesis. The conclusions are as follows:

- Squats at welds developed from surface irregularities without cracks.
- In Step 1, initial surface V-dip irregularities are predominately caused by differential plastic deformation that follows the varying hardness distribution in the heat affected zones. Imperfect finish grinding of welding is not necessary the cause of squats, although it could produce geometric deviation, promoting the development of squats.
- In Step 2, the V-dip deformation causes a dynamic longitudinal contact force and the resulting differential plastic deformation. The latter adds onto the V-dips to cause the initiation and growth of W-shaped irregularities, that eventually develop into typical two-lung-like squats. Each of the lungs is a depression, corresponding to one of the two V's of the W-shape.
- In this process, the role of the varying hardness distribution is to cause the initial V-shaped dip rail surface irregularities. The irregularities subsequently cause a dynamic longitudinal contact force which determines the W-shape and the eventual typical two-lung shape of mature squats. Here the dominant characteristic is the wavelength of the longitudinal contact force, which is determined by the local wheel-track system. This

characteristic wavelength of the contact force of the squats appears from the numerical simulation and field observation to be the same as that of the short pitch corrugation.

- The development of the W-shaped irregularities continuously increases the dynamic contact force at the same location, resulting in a positive feedback growth loop of the irregularities and the dynamic force, leading to rail fatigue and cracks.
- The numerical simulations showed that the irregularity magnitude caused by plastic deformation in a wheel passage is much larger than that caused by wear. This needs further investigation to verify.
- The number of squats induced by a weld is determined by the characteristic length of the hardness variation or the quality of welding. Duo squats are frequently observed at thermite welds because the characteristic length is usually much larger than the typical wavelength of squats, i.e., 20 - 40 mm. A single squat usually occurs at flash butt welds because the characteristic length is in close agreement with that of typical squats. A single squat occurs at some thermite welds due to short characteristic length or good quality of welding, indicating that by improving welding quality, weld-induced squats can be reduced or even avoided.
- Since the hardness-variation induced rail surface irregularities and the resulting dynamic contact force are the precursors to the squats, it is possible to use a dynamics-based method to detect the squats at an early stage or even before any fatigue cracks occur. The characteristic wavelength can be used as a signature for detection and monitoring.

## 6.2. Applications of research findings

The research findings have potential applications in the calculation of rolling contact problem, rail design, maintenance and railway safety.

### 6.2.1. Applications from the study on applicability of half-spaced based methods

It was found that half-space-based methods are sufficiently accurate if the 1.1x condition is satisfied. The conclusion leads to greater confidence in studies of contact problems in engineering science where half-space-based methods are applied. In engineering practice, this study can be a guide to assist engineers in choosing appropriate methods when solving contact problems. Should the 1.1x condition be satisfied and an error tolerance of 9% acceptable, the half-space-based methods are preferred, as they are relatively easier to apply and more computationally efficient. Otherwise, the FE method should be considered [5]. Moreover, this research gives the criterion in term of the dimensions when half-space methods are used to verify other methods.

### 6.2.2. Applications of the explicit FE method for rolling contact

Although the explicit FE method for rolling contact was not new, their capability and accuracy had not been completely justified previously. The findings in this thesis proved the explicit FE to be a quite effective solution method that treat complicated transient wheel-rail rolling contact with various parameters. It can be readily used to solve various dynamic wheel-rail contact problems in, e.g., dynamic vehicle-track interaction, noise, wear and RCF.

### 6.2.3. Practical applications of the research findings on squat formation

In this thesis, the corrugation-induced squat development process including geometry evolution, crack initiation and propagation processes has been better revealed and elaborated. The findings could help rail engineers perform more effective inspections and plan maintenance more efficiently to reduce costs and improve railway safety. The pin-pointing of the crack initiation location and the identified tensile stress state for crack initiation and propagation should benefit the design and manufacturing of more squat-resistant materials. For instance, it is possible to increase the fatigue lifespan of rails by improving their tensile strength. Furthermore, the fatigue mechanism of crack initiation and propagation could be applied to explain other types of RCF of rails.

According to the findings on the initiation mechanism of surface irregularities at welds, it is possible to avoid or delay the formation of squats at welds by controlling the finish hardness distribution during welding. Welds with uniform hardness are less likely to produce squats.

Moreover, the vehicle-track dynamic interaction at both corrugation and welds has provided the physical basis required for accurate estimation of welds quality via on-board measurements. The characteristic of the wavelength can be used as a signature for dynamics-based early detection and track health condition monitoring of squats.

## 6.3. Recommendations on future research

### 6.3.1. Extension of the explicit FE method for rolling contact

The explicit FE method has been widely applied to rolling contact problems. It has been verified in non-conforming rolling contact with and without large geometric spin. Conforming rolling contact can occur at the contact between the wheel tread and the rail gauge corner [6]. This should be further studied with the explicit FE method.

### 6.3.2. Study of rail squats

Considerable development process and mechanism of rail squats have been provided in the present thesis. To fully understand rail squats, a number of aspects are suggested for further study.

#### 1) Contact mechanism of squat formation

In the current research, based on a five-year continual field monitoring, the formation process of corrugation-induced squats was one of the focuses. The contact mechanism in terms of tensile stress and the dynamic force has been proposed. However, the stress analysis through numerical modeling of dynamic vehicle-track interaction was not provided. To better understand the proposed formation mechanism, this analysis is suggested to conduct in further study.

#### 2) Crack initiation at weld-induced squats

A three-step hypothesis of the development process of weld-induced squats was proposed. The first two steps for pre-cracking process were focused on in this thesis by field monitoring and numerical simulation. However, the third step for cracking process was not included. Further work should study the crack initiation.

#### 3) Crack propagation

In the present research, crack propagation was studied in a qualitative manner based on field observation. To better plan rail maintenance, it is suggested to predict the remaining life of rail by calculating the crack growth path and rate through dynamic fracture mechanics in further study. Influencing parameters of crack growth should be considered.

#### 4) Parametric study of squat formation

A parametric study should be carried out to clarify various parameters that influence the formation of squats. The parameters that could increase dynamic force and promote the formation of surface irregularities and crack initiation should be considered, such as profiles of wheel and rail, the material parameters of rail and welds, fastening and rail pads. Additionally, frictional heating may generate interface thermal stress, which can have a significant effect on the mechanical properties of rail and the propensity for fatigue crack initiation [7]. It is also suggested to take into account this factor in the prediction of crack initiation and propagation.

Moreover, regarding the development of weld-induced squats, hardening will occur under repeated wheel-rail rolling contact. As a consequence, hardness and yield stress will evolve over time. The effects of this evolution on the development of surface irregularities is a topic that needs further research. Further research should also investigate whether white etching layers occur on rail surface around welds and how they influence the hardness variation since they are sometimes observed on rail surfaces, e.g., [8].

#### 5) RCF crack initiation model

Classic RCF models were only applicable to some ideal contact conditions. When rail surface irregularities exist, their development continuously changes the contact geometry and therefore the contact stresses. In this case, those classic RCF models have a lot of difficulties to estimate crack initiation. Therefore, a more accurate RCF model should be developed to take into account the variation of contact geometry.

### 6.3.3. Future study on RCF damage

It is suggested that the following aspects should be studied not only on squats but also on other types of RCF damage.

- A better way to understand RCF damage is to reproduce them by taking into account cyclic permanent deformation using experimental modeling devices or the FE dynamic vehicle-track interaction model. In either way, the work hardening, pre-existing residual stress and geometry evolution due to plastic deformation and wear can be included automatically. Existing test rigs can be applied to implement the experimental modeling. Although the FE model currently still have a lot of difficulties in cyclic modeling nowadays due to the high computation cost, it should be feasible in the near future with the development of computer performance.
- To extend rail lifespan and reduce maintenance cost, another interesting research topic could be optimization of the microstructure of rail material to improve the rail fatigue resistance.

## References

- [1] Z. Li, X. Zhao, R. Dollevoet, An approach to determine a critical size for rolling contact fatigue initiating from rail surface defects, *International Journal of Rail Transportation*, 5 (2017) 16-37.
- [2] Z. Li, R. Dollevoet, M. Molodova, X. Zhao, Squat growth—Some observations and the validation of numerical predictions, *Wear*, 271 (2011) 148-157.
- [3] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats—Correlation analysis and numerical modeling, *Wear*, 265 (2008) 1349-1355.
- [4] Z. Li, Squats on railway rails, *Wheel–Rail Interface Handbook*, Woodhead Publishing, 2009, pp. 409-436.
- [5] Z. Wei, C. Shen, Z. Li, R. Dollevoet, Wheel–Rail Impact at Crossings: Relating Dynamic Frictional Contact to Degradation, *Journal of Computational and Nonlinear Dynamics*, 12 (2017) 041016-041011.
- [6] Z. Li, Wheel-rail rolling contact and its application to wear simulation (Ph.D. thesis), Delft University of Technology, Delft, The Netherlands, 2002.
- [7] I.M. Widiyarta, F.J. Franklin, A. Kapoor, Modelling thermal effects in ratcheting-led wear and rolling contact fatigue, *Wear*, 265 (2008) 1325-1331.
- [8] A. Kumar, G. Agarwal, R. Petrov, S. Goto, J. Sietsma, M. Herbig, Microstructural evolution of white and brown etching layers in pearlitic rail steels, *Acta Materialia*, 171 (2019) 48-64.

# ACKNOWLEDGEMENT

After a long journey, this book is about to be printed. I would like to take this great opportunity to thank all the people who helped me at every point during this journey to the degree.

Firstly, I would like to express my deepest gratitude to my supervisor and promotor, prof. Zili Li for his endless guidance and continuous support in all the time of research and writing of this thesis at TU Delft, as well as for his patience and motivation. Prof. Li always has penetrating insights into railway related issues. Particularly, his valuable insights into squats made me always on the right track of this journey. Moreover, his careful reviews and critical thinking would greatly not only improve the quality of my research but also influence me in my future career. I will be always grateful.

Besides, I would like to thank my another promotor prof. Rolf Dollevoet; he became my promotor since the second year of my PhD; I appreciate his support and encouragement. I would like to also express my appreciation to Prof. Andre Molenaar for his guidance as promotor during the first two years of my PhD.

My PhD research was sponsored by TU Delft. The support from TU Delft is greatly acknowledged.

Further, I would like to present my special thanks to the people who helped me quite a lot during my PhD journey. I would like to start with Dr. Xin Zhao. Before I came to Delft, he spent time to visit a room and paid the deposit for me, so that I had a place to stay as soon as I arrived in Delft. Also, in the first two years of my PhD, he helped me a lot with numerical modelling. Thank you for the valuable help in Delft. Also, I would like to thank Dr. Zhiwei Qian, who joined the railway group in half way of my PhD. He shared a lot of knowledge and gave me a lot of constructive suggestions in my writing. Another important person who I would like to thank is Dr. Alfredo Nunez Vicencio. He always can quickly give critical feedback on everything. I very much appreciate your help and support in the last stages of this journey.

I would like to thank all the supporting staffs in our group. Field monitoring is one of the most important parts of my research. Jan Moraal made the monitoring happen. Thank you for the arrangement of the monitoring and for giving me the ride to the sites each time. I also would like to thank Jacqueline Barnhoorn for arranging a lot of administrative issues.

I wish to thank other colleagues and friends in Railway Group and Road Group. Many thanks to Shaoguang Li, Maider Oregui, Nico Burgelman, Maria Molodova, Chang Wan, Haoyu Wang, Lizuo Xin, Chen Shen, Siamand Rahimi, Zhen Yang, Siamak Hajizadeh, Anthonie Boogaard, Yuewei Ma, Xiangming Liu, Zilong Wei, Dr. Ling Chang, Ali Jamshidi,

Pan Zhang, Hongrui Wang and Li Wang, Omid Hajizad and Meysam Naeimi. We spent a lot of unforgettable moments together. We shared offices, we played badminton and table tennis and we worked on Moerdijk bridge at nights. I enjoyed every moment we had together. Thanks should also go to new PhD students: Chunyan He, Bin Zhu, Yang Jin, Yunlong Guo and Siwarak Unsiwilai. You started during my last year, but you organized a lot of fun activities. I also express my thanks to the all visiting PhDs: Lei Xu, Wenqiang Liu, Song Liu and Bin Zhu. Although you only stayed for a short period of time, you leave a lot of good memories behind. I further express my appreciation to some colleagues in Road engineering. They are Dongya Ren, Yuan Zhang, Gang Liu and Mohamad Mohajeri. I enjoyed your companion. I am also grateful to all the visiting scholars in Railway Group: Dr. Yu Zhou, Dr. Pingrui Zhao, Prof. Qian Xiao, Dr. Xianman Chen, Prof. Rong Chen, Dr. Hongqin Lian, Prof. Xiaoping Wu, Dr. Xiubo Liu and Dr. Chunyu Han, they brought a lot of knowledge of China railway. I further give my thanks to my office mates Jeroen Wegdam and Daan Hoekstra for editing my Dutch summary of this thesis.

Moreover, I would like to express an extra thanks to Zhen Yang, Lin Xiao, Xiangming Liu and Zhiwei Qian for organizing a lot of events, which made this journey more enjoyable. Also, thank Xiangming for giving me a daily ride to TU in one winter.

I also give my appreciation to other friends I met in Delft. They are Liyuan Fan, Liyuan Shan, Xiaoman Xu, Hua Dong and Dr. Yuguang Yang. Thank all of them for giving me kind help.

Finally, I must express my very profound gratitude to my parents and my sisters for supporting me spiritually throughout writing this thesis and my life in general. My sincere gratitude also goes to my girlfriend Zhao Lian for her patience and understanding when I got stressed and became depressed.

## **ABOUT THE AUTHOR**

Xiangyun Deng was born in Xuanwei, Yunnan, China in 1984. He received his MSc degree in Engineering Mechanics from the Department of Modern Mechanics at University of Science and Technology of China in 2010. His Master research topic was on the mechanical behaviour of brittle material under impact loading. He started to work as a PhD in the section of railway engineering at Delft University of Technology at the end of 2010. His PhD research focused on the numerical analysis of wheel-rail contact mechanics and formation mechanism of rail squats.

# LIST OF PUBLICATIONS

## Journal papers:

1. X. Deng, Z. Li, Z. Qian, W. Zhai, Q. Xiao, R. Dollevoet. Pre-cracking development of weld-induced squats due to plastic deformation: five-year field monitoring and numerical analysis. Accepted for publication by International Journal of Fatigue (2019).
2. X. Deng, Z. Qian, Z. Li, R. Dollevoet. Investigation of the formation of corrugation-induced rail squats based on extensive field monitoring. International Journal of Fatigue 2018;112:94-105.
3. X. Deng, Z. Qian, Z. Li, R. Dollevoet. Applicability of half-space-based methods to non-conforming elastic normal contact problems. International Journal of Mechanical Sciences 2017;126:229-34.
4. X. Deng, Z. Qian, R. Dollevoet. Lagrangian Explicit Finite Element Modeling for Spin-Rolling Contact. Journal of Tribology, 2015. 137(4): pp. 041401.
5. Z. Yang, X. Deng, Z. Li, Numerical modeling of dynamic frictional rolling contact with an explicit finite element method, Tribology International, 129 (2019) 214-231.
6. X. Chen, X. Deng, L. Xu, A Three-Dimensional Dynamic Model for Railway Vehicle–Track Interactions, Journal of Computational and Nonlinear Dynamics, 13 (2018) 071006.

## Conference papers:

1. X. Deng, M. Naeimi, Z. Li, Z. Qian, R. Dollevoet. Residual fatigue life evaluation of rail at squats seeds using 3D explicit finite element analysis. 0 Breugel, Kvan Koenders, EAB(Eds.) Proceedings of the 1st international conference on ageing of materials and structures, 2014. AMS'14 (pp. 1-10) Delft: TU Delft.
2. Z. Qian, X. Deng, Z. Li, R. Dollevoet. Influences of characteristic size of contact bodies on rolling contact. Lakusic, S (Eds.) Proceedings of the 3rd international conference on road and rail infrastructure, 2014. (pp. 1-10) Zagreb: University of Zagreb.
3. X. Deng, Z. Li, R. Dollevoet. Analysis of the impact-contact states of stresses and strains at crack initiation of squats. Proceedings of the 3rd international conference on computational contact mechanics, 2013. Lecce, Italy.
4. X. Deng, Z. Li, R. Dollevoet. Finite element simulation of rolling contact with spin in elasticity and elasto-plasticity. Zhang, Weihua, Jin, Xuesong (Eds.) Proceedings of the 9th international conference on contact mechanics and wear of wheel/rail systems, 2012. (pp. 113-119) s.l.: State Key Laboratory of Traction Poser, Southwest Jiaotong University.