Micro-CT Scanning in the Investigation of Squat Defects in Rail Steel

Shaun Earl

Submitted for the degree of Doctor of Engineering in the Department of Materials Science and Engineering at the University of Sheffield, UK.

January 2021

Summary

This thesis contains the results from an investigation into Squats, a discrete rail steel defect. Micro-computed tomography (micro-CT) scanning was used to scan the entire structure of four out of five defects removed from railway track. The fifth was not scanned as it shelled in track. The difficulty turning these raw X-ray images into a segmented 3D model were overcome by developing a new technique. Isolating separate regions of the scans created areas where voxel value variation was at a minimum (i.e. the histogram became one narrow peak rather than multiple broad peaks). This allowed the automatic crack segregation module to work fairly well, and then enhanced using the region growing modules within the software. These scan results were then verified to be accurate using metallographic sample preparation, optical and electron microscopy.

Micro CT scanning was performed on a custom 450KeV scanner, allowing the capture of the first entire Squat crack network morphology. Full defect imaging allowed the different defects to be compared to each other, highlighting differences and similarities. The defects came from metro, mixed and high-speed railways and one was found within an aluminothermic weld. The verification process and investigations of the scan volumes yielded further information about the defect's origins. These origins were used to determine that, for the five defects investigated, there were four different causes. The two that shared a cause were from the same track section. Based on the causes, the defects were identified as a Stud, two Grinding Induced Squats (GIS), a Squat (caused by the legacy issue of MnS inclusions) and possibly a Squat or Stud in a slightly contaminated weld. None of the defects were considered to be a classic Squat, which is caused by Rolling Contact Fatigue (RCF), because there were other factors in their initiation.

One of the defects contained a transverse defect, which is a crack that grows down through the railhead and can break the rail. This transverse defect was ~9mm deep into the rail when it was removed, meaning it would not have returned to the surface to shell. A high-resolution volume of the transverse defect region was created and its origin gives an important insight into the potential causes of rail-breaking defects. The origin of this transverse defect was a cluster of debris-filled voids that had formed due to the corrosion and cyclic loading (fretting) of a crack branch. These voids aligned with a deep grinding mark on the surface of the rail, which acted as a stress raiser. Because corrosion is a factor in this transverse defect case, the age of a rail

and its environment are factors for defect development as well as traffic volume, given the correlation of corrosion with time.

Results of this work highlight both the importance of a good surface finish and the diversity of causes found within the term "Squat". Thus the identification of the Stud variant may be the beginning of a more comprehensive group of Squat type defects being established. This refining of the category could lead to fruitful big data analyses of the Squat type defect occurrences.

The CT volumes of the defects created in this work can easily be stored for comparison in future investigations. The virtual nature of the volumes allows the sharing of defect information more readily than physical and sectioned defects, which deteriorate with time and require physical storage and transport.

Acknowledgements

Finally, my heart goes out to my family, who kept me strong on my turbulent and winding path. Their unwavering support of my studies over the years, on top of this project, means the world to me and is never overlooked.

In science, we stand on the shoulders of giants. Therefore, we must support each other to hold our balance, as well as drive each other forwards. The people mentioned here were instrumental in that act. As I move onwards, they will be dearly missed, or cherished in future ventures together.

Shaun Earl

"Born and bred in Bolton", forged in Sheffield.

Contents

Acronyms

CoF   Coefficient of Friction

CT   Computed Tomography

GCC   Gauge Corner Cracking

GIS   Grinding Induced Squat

IF   Infinite Focus

NDT   Non-Destructive Testing

PAG   Prior Austenite Grain

RCF   Rolling Contact Fatigue

WEL   White Etching Layer

WRI   Wheel-Rail Interface

Chapter 1: Introduction

Railways are a well-established part of the UK infrastructure. From their invention to today, the way that trains get from A to B has not fundamentally changed, but the details certainly have, such as train design, power sources (steam through to electric), materials used and the speeds at which trains travel. Passengers being afraid of travelling on a train that could reach the scary speed of 30mph seems unfathomable now. Especially as we enter a faster paced world, with the introduction of high-speed trains on dedicated high speed lines, the motorways of the rail network. These higher speeds, coupled with much more traffic and longer lasting track components has brought with it a new issue, the Squat.

Squats are classically a rolling contact fatigue (RCF) related defect in the running surface of a rail. A pair of cracks that grow beneath the surface, allowing the surface to deform and depress, resulting in two black lobes or spots to appear on the rail surface. These black spots are the typical signature that there is a Squat in the rail (Figure 1). The localised deformation also leads to a widening of the running band, which is the shiny part of the railhead. The initiation of the defect is usually on or very close to the surface so another typical feature, a V-shaped crack, also develops as the crack plane grows.
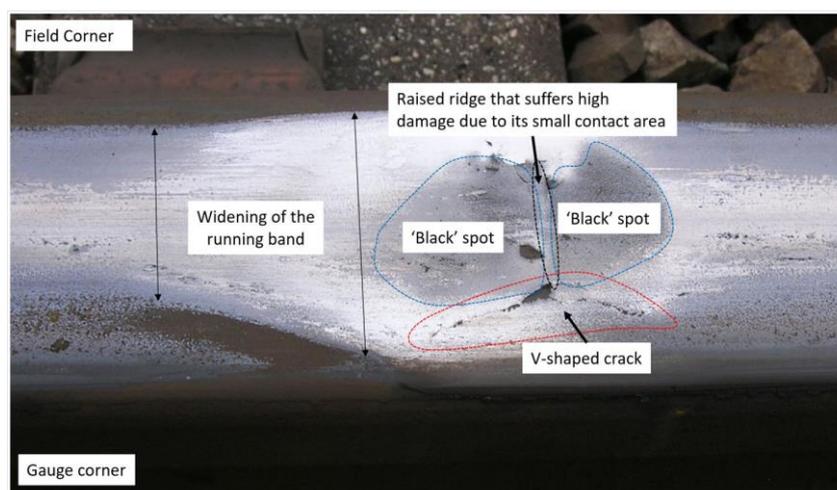
Figure 1: A typical Squat showing the associated annotated features. Original image courtesy of British Steel

When the defect reaches a critical size, the subsurface cracks propagate to the surface at the edges of the running band resulting in more surface cracks. The deformed surface surrounded by cracks will eventually shell leaving the rail severely damaged. The defect, both in its developing and shelled form, cause vibrations/strong-dynamic forces, that cause more damage to the track infrastructure and the trains passing over them. The worst-case scenario is that the subsurface cracks grow down into the rail rather than surface and shell, resulting in a transverse break of the rail, almost certainly resulting in a passing train being derailed.

An example of the catastrophic effects of a broken rail are exemplified in the case of the Hatfield derailment outside London in 2000. This rail break was due to gauge corner cracking (GCC), another form of RCF defect, which caused multiple rail breaks when the cracks grew into the rail, such that the rail could not support the heavy load of a train. This derailment led to fatalities and severe injuries to those on board the train. This type of risk to infrastructure and public safety, not to mention the cost of managing defects, make the elimination of RCF defects a high priority. Now GCC is better understood and more effectively managed, another problem has now become prominent. Squats are currently the main cause of early rail replacements for Network Rail [1]. Squats pose a conflict with the solution used for GCC. GCC was partially remedied using harder rails, but there is some opinion that Squats are more common on harder rails. Grinding trains were also used to combat GCC through artificial wear and this is currently the main technique used against Squats, though some Squat 'hot-spots' respawn Squats soon after their removal.

Determining the cause of Squats and the reason for some of them growing transversely down into the rail is crucial to managing and eliminating them as a problem and risk. This need has led to a strong research focus on Squats over the past few years. The current work continues the effort by looking at defects from different types of railway lines: light/metro, mixed and high-speed traffic.

Knowing what lies beneath a rail surface is a challenge, with the railhead being made from steel. This problem has been tackled using Non-Destructive Testing (NDT) methods, such as ultrasonic and eddy current tests, but these cannot profile the crack. Using X-rays is also an option if a rail has been removed from track. This can be expanded to X-ray micro-computed tomography (μ-CT) scanning, creating 3D data from thousands of x-rays at multiple angles.

The comparison of different Squat samples should give clues as to how they initiate, show how the different traffic types affect crack development and provide insight as to why some defects develop into transverse defects whilst others do not. In conjunction with X-ray (μ-CT) scanning as a non-destructive volumetric inspection tool [2], this work utilises traditional sample preparation techniques to validate the output of the μ-CT and further investigate areas of interest.

This chapter will cover the research projects aims, novelty and focus before giving an overview of Squats and Squat Type unidentified Defects (Studs), then concluding with the reasoning behind the construction of the thesis chapters. This thesis is not separate pieces of work per chapter, but a continuous story covering different aspects due to the forensic nature of this work.

1.1 Project Aim

The aim of this project was to evaluate the effectiveness of μ-CT scanning as a NDT method, for determining the size and morphology of entire squats type defects. This leads to a more informed method of sectioning a rail defect for investigation.

The main challenge for this task is the x-ray penetration of radio dense steel. The effective and verified use of μ-CT scanning will provide 3D models to assist in the understanding of Squat defects and their features.

1.2 Research plan

This research became defined by an early question: "Where in the Squat defect do you cut to maximise the information you can get from a defect?" This question arose because micrographs of Squat defects often showed a similar crack structure, but it was not always clear where in the rail the micrographs had been taken from. Most were probably taken from the centre of the defect.

This question lead to using μ-CT scanning to scan three Squat defects to map the crack network, ultimately resulting in 3D models being created from the 2D X-ray images. The ability to create full 3D models of actual rail defects will provide:

1. Comparisons for simulated models

2. A guide for investigating the scanned defect further

3. A way of comparing the severity of the crack network of different defects

4. A clear identification of any morphological features of interest.

All of this is possible before any destructive sample preparation that severely limit the options for future investigation.

μ-CT scanning has been used previously to successfully investigate the form of RCF crack networks by Jessop [3] and Squats by Naeimi et al. [4]. Naeimi et al. also give good justifications for why CT scanning is a powerful NDT tool that maximises the information that can be retrieved from a sample. The current work builds on this by expanding the scan to the entire defect, which required a high powered 450kV μ-CT scanner

Chapter 3 covers the methods and techniques used before Chapter 4 provides a comparison of the CT models created, along with verification that what the scans showed was true. Verification is needed because artefacts can appear in CT scans due to beam effects. These artefacts are not real and so a metallographic investigation was undertaken to determine what was real and what was artifact. Chapter 5 covers the metallurgical work that verified the work in Chapter 4, work conducted to determine if thermal damage was present and micrographs relating to defect initiations.

One of the three scanned defects showed very extensive damage considering its total traffic loading (MGT) and a transverse crack that had grown to a significant depth. The reasons for why this defect had developed so rapidly became the next research focus, covered in Chapter 6. I.e. Chapter 6 is an in depth look at one of the aspects noticed in the core work in Chapter 4. This Chapter also links surface features to subsurface features to provide clues for track inspectors as to what may be below the surface.

In summary the research questions for this work are:

1. Can µ-CT scanning accurately detect crack networks on a large scale within a rail?
2. Are all of the features captured during scanning real, or do artefacts make the results deceptive?
3. Are there differences between the crack morphologies of defects?
4. Is there a correlation between the microstructure and the defect?
5. What influence does the surface quality have on the defect growth behaviour?

Each of these points is addressed by a Chapter, from 4 to 6.

1.3 Novelty of project

This project is the first µ-CT scan of an entire rail defect, penetrating up to 11cm of steel and still achieving visible image of cracks. This full rail scanning allowed the first extraction of an entire crack network from a defective rail, shown as a CT volume/model. The CT model required the development of a technique for crack extraction within a 'noisy' sample such as steel, as the automatic function in the processing software, VGStudiosMAX, did not work. A new technique is developed through this project, the segmentation of isolated regions of low range voxels, to compensate for the noise created from scanning steel, covered in Chapter 3.

These volumes provided evidence of double initiations that were close together. Combined with the finding of small Squat like structures inside larger squats, as well as in areas without a large defect. Regarding early stage initiation, Y-shaped cracks were identified as an early surface sign of the development of Squat type defects.

Transverse cracks are traced back to their origin, then linked to grinding marks. This confirms the presence of Grinding Induced Squats (GIS), as detailed in the literature, through an NDT approach rather than the track observations and micrographs. Along with the grinding marks that obscured the surface of the rail, pitting covered the surface of two of the defects, which seems to have influenced the crack growth behaviour. This factor needs further investigation with more examples of defects with pitted surfaces to be scanned and checked for verification.

1.4 References

[1]   Network Rail, "Rail Head Squats," 2015. [Online]. Available: https://cdn.networkrail.co.uk/wp-content/uploads/2017/03/Challenge-Statement-Track-Plain-line-rail-head-squats.pdf. [Accessed: 13-Jun-2018].

[2]   E. N. Landis and D. T. Keane, "X-ray microtomography," *Mater. Charact.*, vol. 61, no. 12, pp. 1305–1316, 2010.

[3]   C. Jessop, J. Ahlström, L. Hammar, S. Fæster, and H. K. Danielsen, "3D characterization of rolling contact fatigue crack networks," *Wear*, vol. 366–367, pp. 392–400, 2016.

[4]   M. Naeimi *et al.*, "Reconstruction of the rolling contact fatigue cracks in rails using X-ray computed tomography," *NDT E Int.*, vol. 92, no. April, pp. 199–212, 2017.

[5]   J. J. Kalker, *Three-dimensional elastic bodies in rolling contact*. Kluwer Academic Publishers, 1990.

[6]   B. L. Singh and C. B. E. Sr, "Rolling Contact Fatigue in Rails – an Overview," 2008. [Online]. Available: http://www.iricen.gov.in/iricen/ipwe_seminar/2011/paper_2_lalloosingh.pdf. [Accessed: 17-Jul-2020].

[7]   W. Daniel, S. Pal, and M. . Farjoo, "Rail squats: progress in understanding the Australian experience," *J. Rail Rapid Transit*, vol. 227, no. 5, pp. 481–492, 2013.

[8]   M. Steenbergen, "Rolling contact fatigue: Spalling versus transverse fracture of rails," *Wear*, vol. 380–381, pp. 96–105, 2017.

Chapter 2 : Literature Review

2.1 Introduction

Classical Squats are rolling contact fatigue (RCF) related defects found on the running surface of railway tracks. Two dark spots or lobes on the running surface, sometimes referred to as "lung shaped", identify them. These dark spots are due to corrosion building up in the depressed regions of the rail. These depressions are due to cracks growing subsurface. They were originally noticed in Japan in the 1950s before being documented in the UK in the 1970s, followed soon after by other countries.

The cause of Squats is not clear, though their initiation has been associated with dents, welds, wheel burns and corrugation [1]. They are currently the leading cause of early rail replacement in the UK, costing ~£4m p.a. in repairs on the UK network [1]. They also pose a safety risk as some Squats can develop into a transverse defect, meaning the crack starts to grow downwards into the rail, which can break the rail and potentially derail a train. This high replacement cost and risk to public safety make Squats a high priority for railway research.

Their link to rail irregularities as mentioned above, along with their occurrence on straight and gentle curves means an understanding what happens in the contact patch between the wheel and the rail is crucial to determining why cracks occur and develop the way they do. It is also necessary to understand what other defects occur on the running surface, as Squats can be affected by other defects, especially corrugation. These factors, alongside the way that tracks are built and maintained, make up the literature review to establish the core knowledge used for this work.

The next few sections will discuss each of the factors in relative isolation before the meaning of all of the factors combined is summarised at the end. The literature covered is not an exhaustive coverage of the literature available on Squats, it is just what is needed to understand this work. A major omission from the review is the details of the computer modelling (e.g. finite element) of Squat development. In addition, some of the mechanics involved in railway operations is not covered in depth, such as train dynamics.

2.2 Contact Mechanics and the Wheel-Rail Interface (WRI)

Contact mechanics is a broad area covering many applications and uses Hertzian contact theory, named after Heinrich Hertz. In 1882, Hertz at the age of 24 published a classical paper on contact mechanics entitled "*On the contact of elastics solids*" [2]. In order to use Hertzian elastic contact theory, certain assumptions have to be met: the radii of the two contacting bodies are significantly larger than the contact area between the two bodies; each body is an elastic half space with a small elliptical or circular contact area; the surfaces are frictionless and the two bodies behave perfectly elastically. Hertzian elastic contact theory is used extensively, an provides an accurate view, provided contact is indeed elastic. Where contact is plastic and where friction values increase, then additional analysis to Hertzian contact is required. In reality, there is often varying friction between the wheel and the rail, plastic yielding often occurs when the wheel passes over the rail and the contact area often contains third bodies such as lubricants and contaminants.

Due to these limitations in the theory of Hertzian contact due to the assumptions mentioned above, in addition to the wheel and rail not being infinitely long cylinders, non-Hertzian contact can be more suitable. Non-Hertzian contact is used in a method called CONTACT [5]. This leads to the understanding that the distribution within a contact is not perfectly even the stress state changes toward the edge of the contact. Because of this variation in distribution, Finite Element Analysis can be useful for modelling contacts, especially if there are deviations in the smoothness of the contact (surface defects) that Hertzian ellipses would not account for.

Understanding the wheel-rail interface (WRI) is critical, as the contact between the rails and the train's conical wheels determines everything, including acceleration, braking and motion of the train around a curve. Figure 2 shows a wheel contacting a rail and the slight variation in the wheel depth (blue arrows) from left to right, making the wheel slightly conical rather than cylindrical. Note how the wheel makes contact with the worn rail at two points rather than across the whole profile. This is covered more in section 2.6.

The conical shape of the wheels allow the train to curve by creating a differential in the radius of the wheels as the train moves away from its centre position on the rail. This differential, in other words the outer wheel circumference being larger, means that the outer wheel covers more distance than the inner wheel leading to curving. Considering the importance of the

contact, it is necessary to understand the conditions and parameters that affect the contact: the tribology.
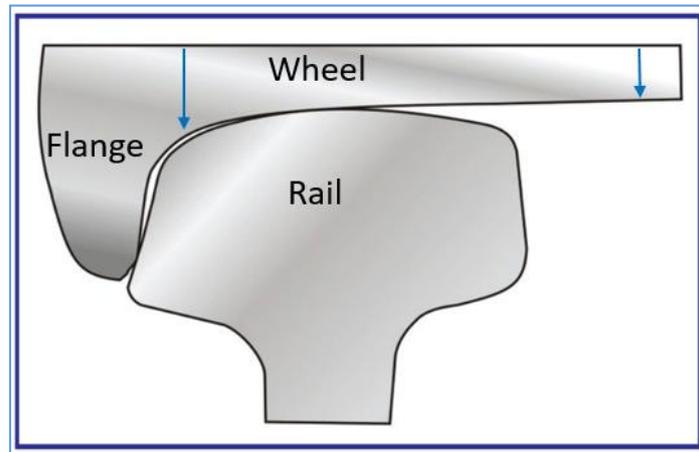
Figure 2: An example of a wheel and a rail in contact. Original image taken from [3].

2.2.1 Tribology

Tribology concerns itself with 3 main topics; friction, lubrication and wear. Friction and lubrication are directly linked as lubrication affects friction. These topics are covered in this section, whereas wear is covered in section 2.5.2.

For this work, the tribology of the wheel contacting the rail is the focus, but tribology covers any area where machines with moving parts are used. Critical failures can occur due to excessive wear if tribology is not understood and in the case of applications such as flight or rail travel, this can be catastrophic.

2.2.2 Line and Elliptical Contacts

When two solid bodies are brought into contact with each other, they can form conforming or non-conforming contacts. Conforming contacts fit each other well, whereas non-conforming contacts make point and/or line contacts, which is the case when a train wheel contacts with a rail. Two parallel cylinders in rolling contact with each other produce a line contact.

Some small scale testing equipment that simulate the WRI such as twin disc machines [4] use two discs with the same orientation, producing a line contact (Figure 3). The WRI between a

train wheel and a railway track have perpendicular curvatures to each other and so make an approximate elliptical contact (Figure 4). In reality, the contact patch is not perfectly elliptical due to deviated conformity between the wheel and rail. The contact patch can be measured directly using methods such as pressure sensitive films and ultrasonic measurements [5], or be modelled using CONTACT [6], FASTSIM [7] and finite element analysis. Each model provides varying levels of calculation speed and processing complexity and the results vary as Figure 5 shows.
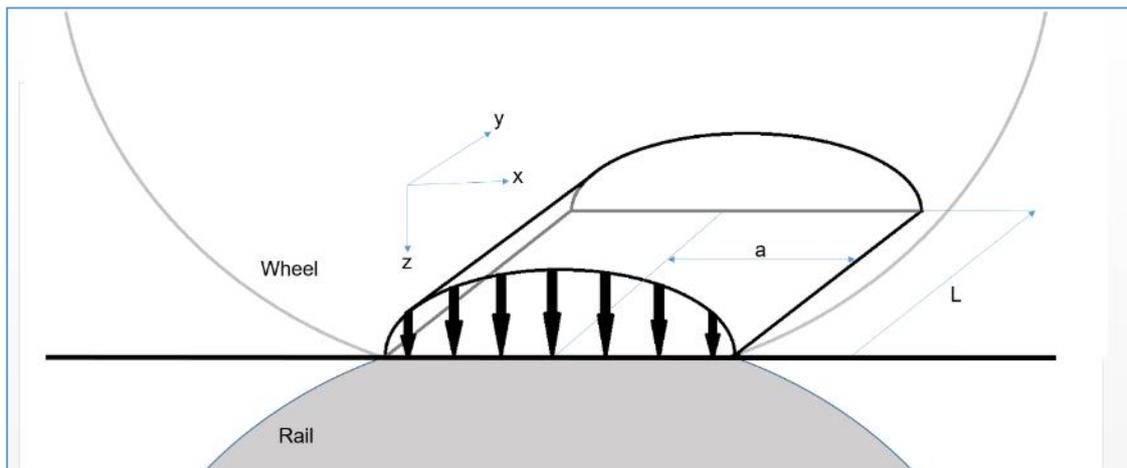
Figure 3: Contact patch and pressure distribution for a wheel/ rail line contact.
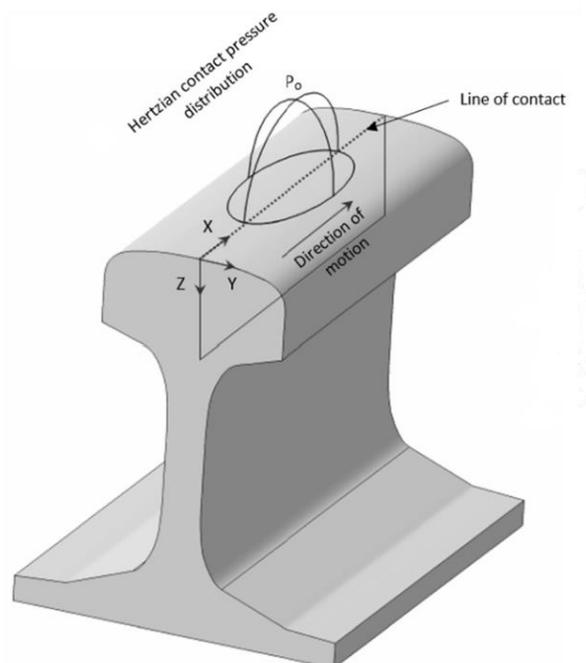
Figure 4: Elliptical contact patch and pressure distribution [8] .
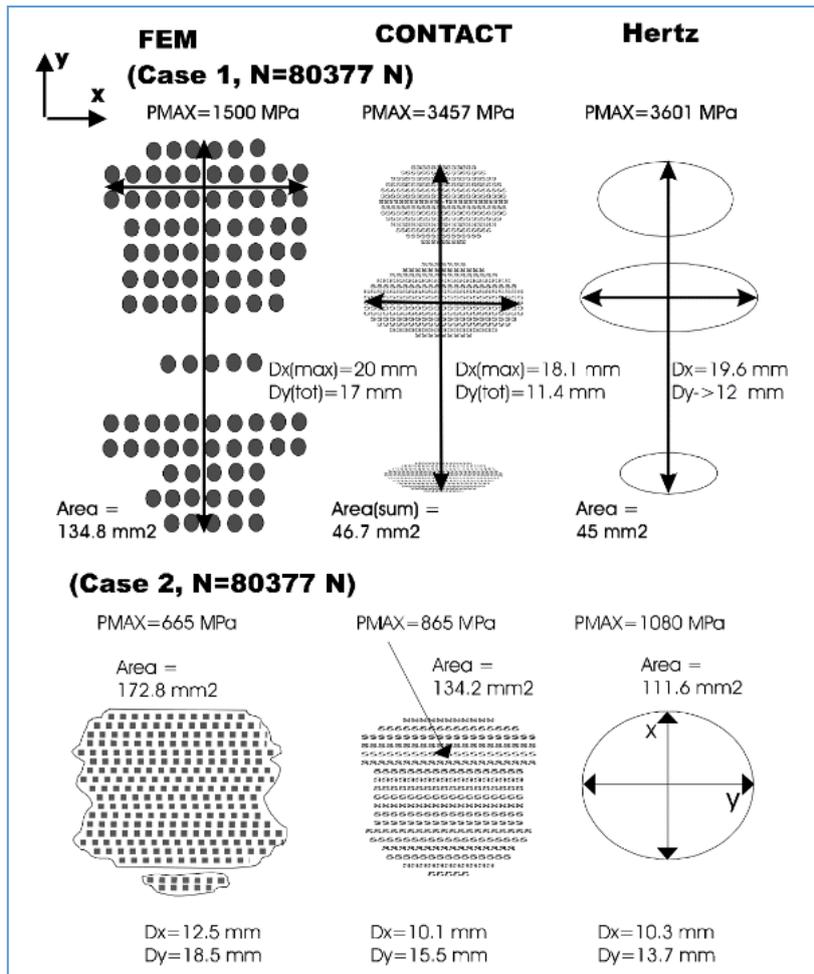
Figure 5: Comparison of contact patches and pressures for FEM, CONTACT and Hertz modelling [9].

Contact pressure affects the size of the contact patch, the loads on the contact, the material properties of the two bodies, any third body between the contacting bodies and the roughness of the surfaces. Third bodies such as lubricants and debris (oxides, sand) can change how the pressure is distributed. Moreover, the maximum pressure varies with depth into the material (Figure 6). Roughness is important because surfaces are never truly smooth despite appearances on the macroscopic scale. They are rough and so only the asperities of the surfaces actually make contact and large loads on a few asperities causes them to yield and deform. This means that the whole of the apparent contact area is not supporting the applied load, being distributed at asperities with many point and line contacts throughout the contact patch. Roughness also plays a role in the intensity of the peak stresses experienced by the rail as a high roughness can lead to peak pressures eight times higher than smooth contact conditions [10] (Figure 7). These peak pressures often occur subsurface in rails, dependant on the surface

friction. High subsurface peak pressures due to high surface roughness is a crucial consideration in grinding, a maintenance technique discussed in section 2.8. A smooth surface finish is desirable to reduce these peak pressures.
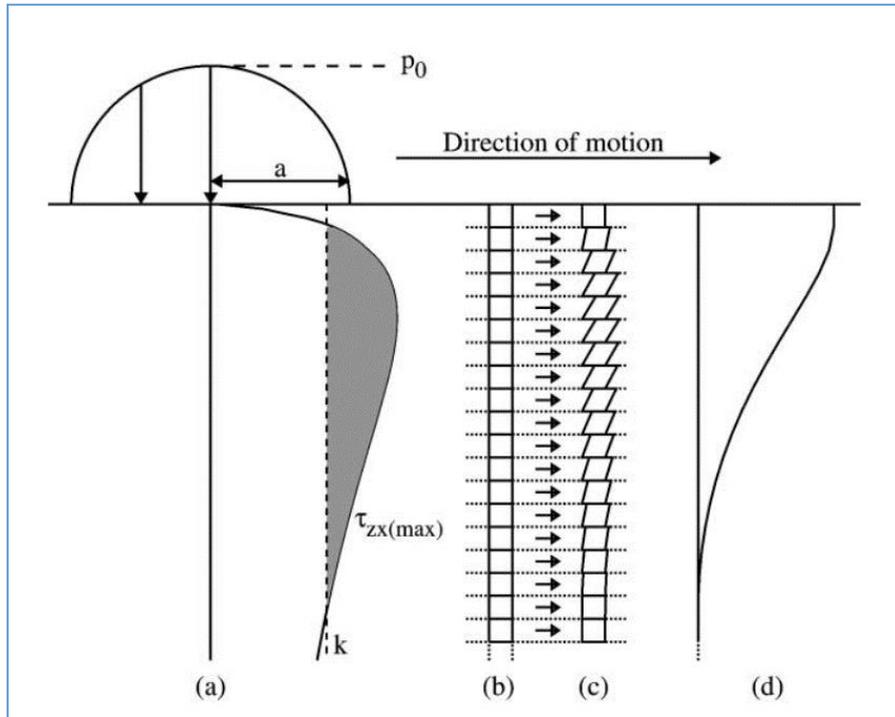
Figure 6: Variation of subsurface stress due to a contact. a) Variation with depth of maximum orthogonal shear stress (τ_{zx}). b) Unrestrained material elements. c) Plastic shear strain increment per cycle. d) Material displacement after 1 cycle [11].
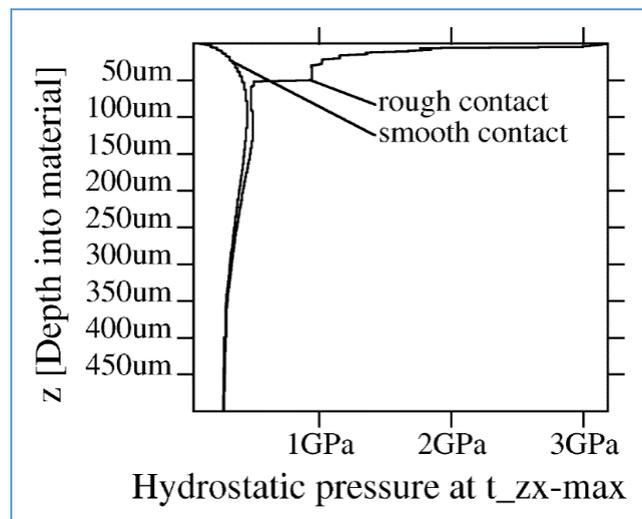
Figure 7: Comparison of stresses for rough and smooth contacts [10].

2.2.3   Friction

The friction that occurs between two bodies, such as a wheel and a rail, is known as wheel-rail adhesion [12]. This adhesion allows both traction and braking of the train. Friction is measured as coefficient of friction (CoF). CoF is a value that is typically between 0 and 1, although exceptional combinations of materials and contact conditions can produce higher numbers.

The Laws of Friction, ascribed firstly to Leonardo da Vinci c.1500 and then Amontons c.1700:

i.   The friction force F is independent of the apparent area of contact.
ii.   The friction force is proportional to the normal load between the bodies,
      i.e. $F \alpha W$, i.e. $F = \mu W$.

      μ is the friction coefficient, W is the load.

      Coulomb (1785) added a third law:
iii. The friction coefficient is independent of the sliding speed.

It is necessary to distinguish between the force needed to initiate the lateral motion (static friction) and the one needed to maintain it. The friction coefficient during sliding is called the kinetic friction coefficient. It is generally lower than static coefficient of friction [1].

The study of friction is complex and the origins of friction are still not fully understood. Its origins lie in adhesion and deformation, but the full description of friction is beyond the scope of this thesis.

Friction is important in the wheel rail interface because it needs to be high enough to allow acceleration and braking, but low enough that the train runs economically once it is up to speed. This desired economy is why such as small contact patch is needed. Ratchetting (covered in 2.2) is dependent on friction. The shakedown limit decreases when the CoF exceeds 0.3 and when the CoF equals 0.6, the strength of the material is halved compared to CoF < 0.3 [14].

Figure 8 portrays stress distributions, due to the wheel-rail contact, for both compressive load and compressive load with traction included. Introducing the concept of traction into Hertzian theory causes the shear stresses below the surface to distort. Note how the stress contour under

traction creates an inverted V-shape at the leading edge of the wheel. This will be important when looking at the subsurface shape of Squat defects.
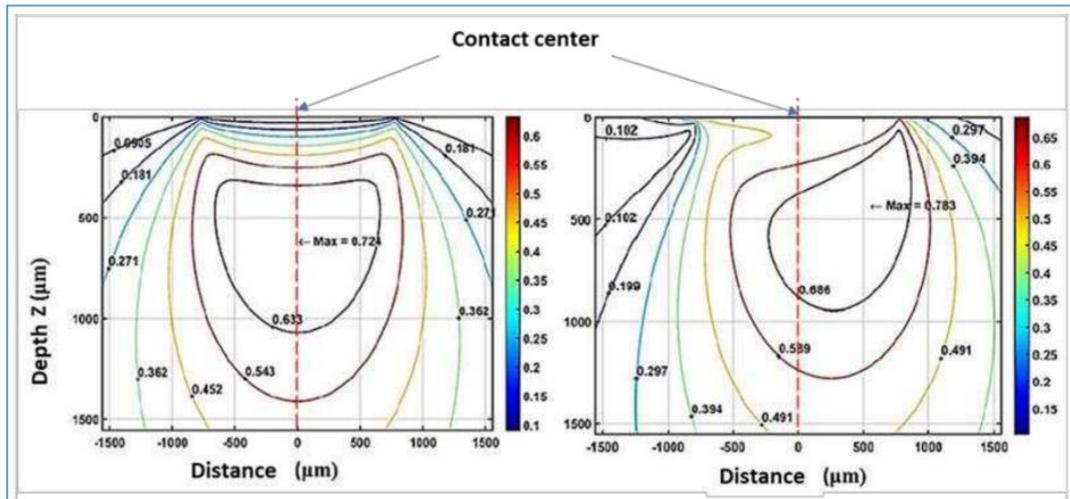
Figure 8: Stress distributions of maximum shear stress (in GPa) for compressive static loading (left) and compressive tractive loading under 1.8GPa of compressive loading and traction of 0.2% where applicable [13].

2.2.4   Shakedown and Ratchetting

Rails are exposed to high cyclic loads that exceed the yield point of a new rail. The rail goes through elastic deformation and into plastic deformation. When the rail is unloaded, the plastically stressed material will return to a low stress state along a path further up the strain axis of the stress strain curve than the original loading curve. This material has now been work hardened and the elastic shakedown limit has now increased. If any of the following loading cycles exceed the new elastic limit and enter the plastic deformation regime, more work hardening can occur and the resulting in the limit being raised further. The other possibility is that the material will fail and cracks will initiate. The plastic shakedown limit increases as well as the elastic shakedown limit.
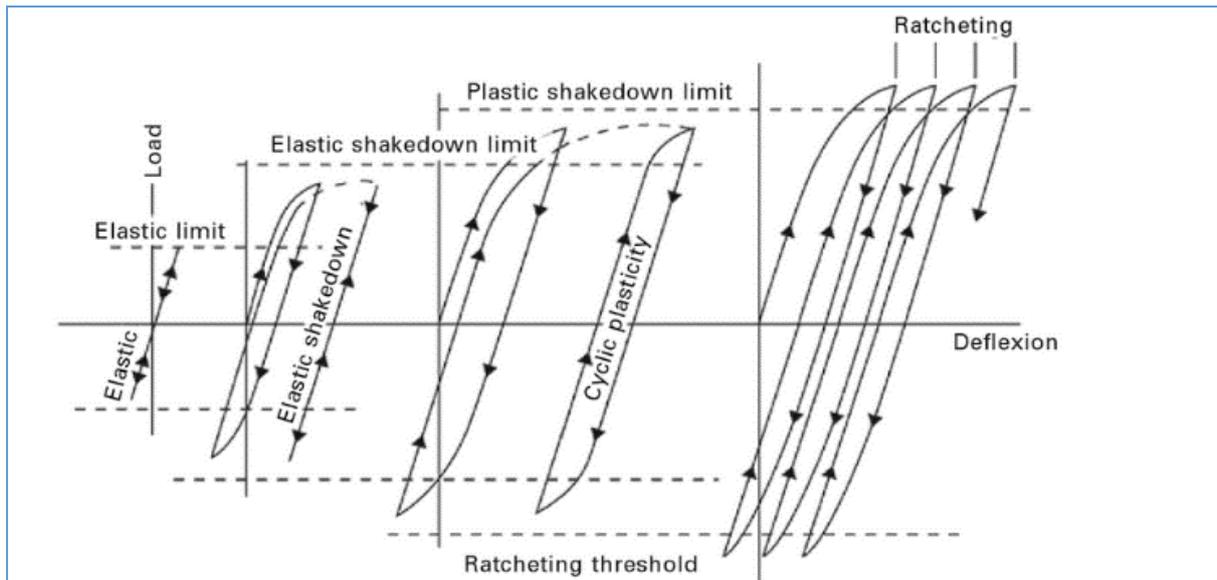
Figure 9: Material response to cyclic loading. Loading above the elastic shakedown limit causes plastic deformation each cycle [15].

Under Hertzian contact conditions the peak shear stress is below the surface. This requires that contact is elastic and that the friction between rail and wheel is low. However, when the friction rises to 0.3 or above, the peak stress moves to the surface. Shakedown is most severe where the peak stresses from the wheel contact occurs, and is exacerbated by traction forces. This shakedown of the subsurface material can lead to subsurface material flow that when combined with a metallurgical imperfection, can lead to subsurface crack initiation [16]. As seen in Figure 9, exceeding the plastic shakedown limit causes ratchetting. Ratchetting is an incremental collapse of the material at the surface of the body in contact. The material exhausts its ductility and deforms, folding over into a weakened microstructure that is prone to crack initiation and material loss. The deformed material seen in the surface region of Figure 10 is due to ratchetting.
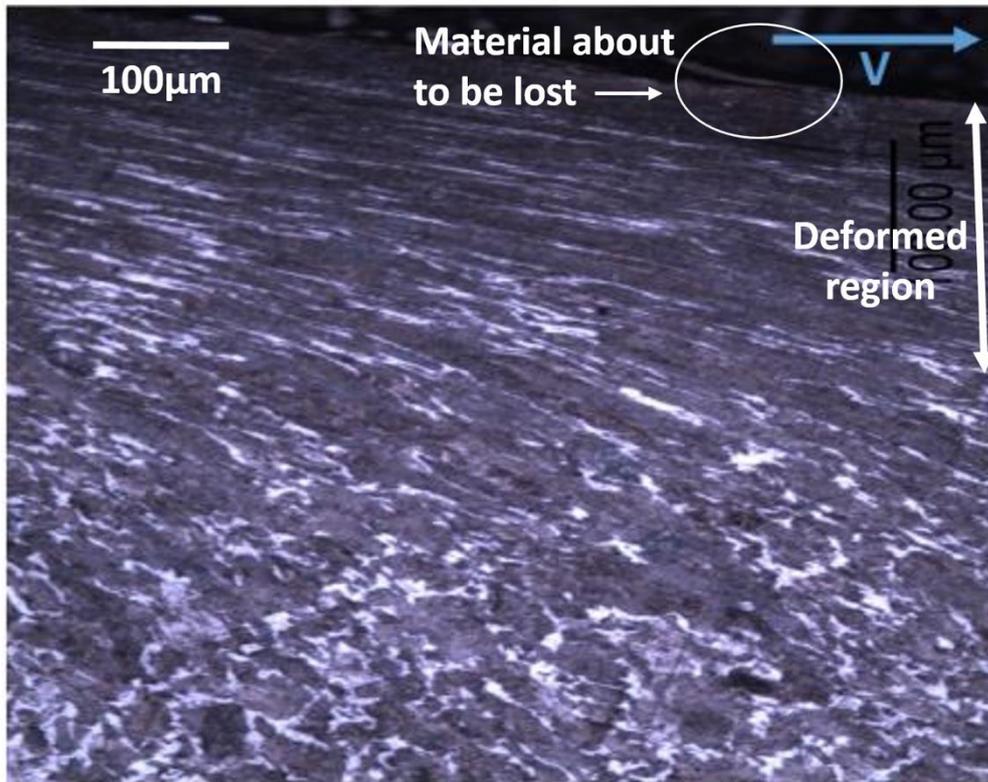
Figure 10: Ratchetting of rail steel in a disc sample. Material loss can be seen on the surface in the top of the image.

2.2.5 Slip and creep

Slip can occur between the wheel and rail, either partially or fully, dependant on CoF and the slip. Slip is presented as a percentage, which is the difference between the rolling distance of the wheel compared to the distance covered on the rail. Slip as low as 1% is enough to cause some damage to the rail [17] but slip is expected to reach as high as 20-30% during acceleration in order to form some of the surface damage observed on rails [18]. Slip is common during acceleration and braking because for a train to accelerate it needs the wheels to be moving slightly faster than the rail and in order to brake the wheel needs to move slower than the rail.

Research by Carter reported that the forces acting between the train wheel and the rail are proportional to the creep [19], which is another term for slip. Creepage occurs because as a wheel tries to navigate a curve, it is not in line with the rail and so it forms and angle between the wheelset and the rail (α in Figure 11).

Using Figure 11, the wheel is rotating with a circumferential velocity (V_C) and the linear velocity of the train along the rail would be the translational velocity (V_T). The final velocity to be considered would be the lateral velocity ($\dot{y}$) as centrifugal force tries to pull the train sideways because the track is turning the train even though the train wants to travel straight. The purpose of this explanation is just to show that the wheelset does not simply follow the track around, they sit on the rail at an angle (Figure 12), so there are multiple velocities at work that lead to creepage in the contact patch as the train curves. Therefore, creepage is the continuous relative movement of the wheel in relation to the rail.
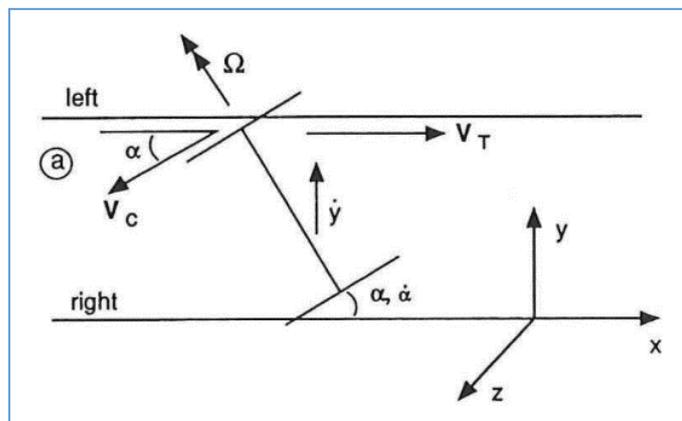
Figure 11: A train wheelset on a track of a left and right rail. Ω is the angular velocity of the wheelset [20].
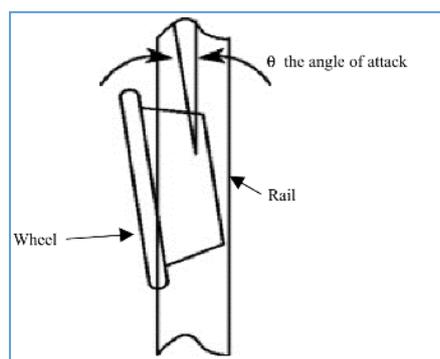
Figure 12: A plan view of a train wheel curving around a bend in a track. Θ=α in Figure 11 [21].

There is an effect called slip and stick, where stick means full traction, or no slip. Slip and stick goes further into explaining creep and slip. Simply expressed, the contact is not in full adhesion

with the rail at all times, as this is not possible due to the friction that this would require under such high forces. Therefore, the wheel can be partially sticking to the rail and partially slipping at the same time within the same contact area (Figure 13).
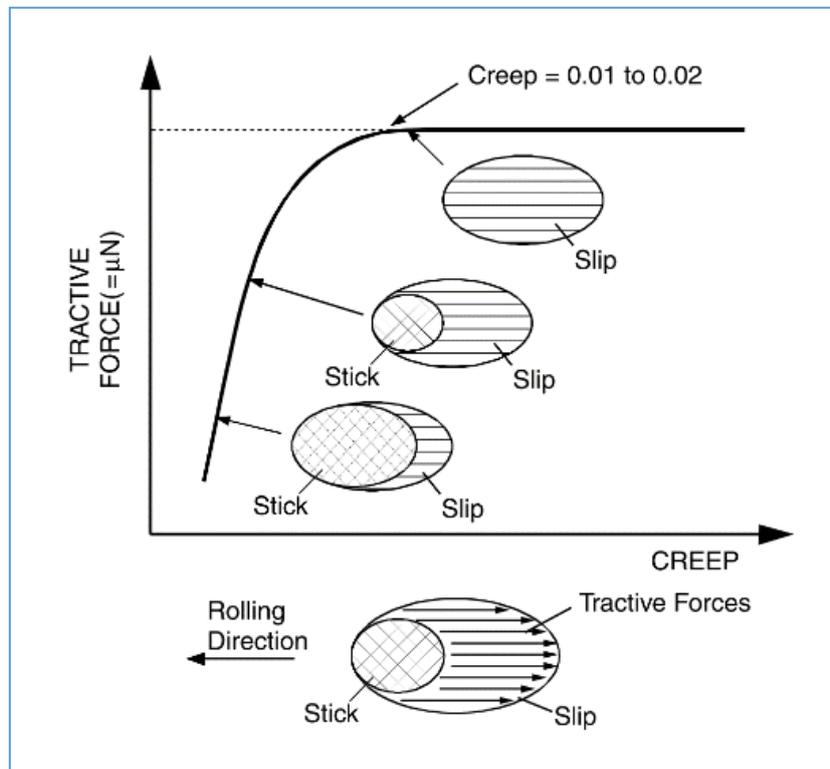
Figure 13: The relationship between tractive forces and creep in a wheel rail contact. The ellipses represent the regions of stick and slip in a contact patch [22].

Figure 13 shows that as tractive force increases the slip region grows: meaning that the wheel is more likely to slip during high traction events such as acceleration and braking. So slip and creepage (and so as a direct result, friction) must be understood as part of understanding what causes material overloading in a railhead. Overloading a material beyond its elastic limit leads to plastic deformation, meaning that the damage is permanent and cannot recover elastically.

2.2.6 Friction management

Third body layers (something between the initial two bodies in contact such as a wheel and a rail) can change the CoF, which can be used to create more desirable friction conditions: some third bodies such as sand will increase friction whereas a grease would reduce the friction. Increasing the friction improves a train's ability to brake, accelerate and the climb inclines.

Reducing friction reduces damage in the WRI, especially when forces are higher such as in curves. Some third bodies are deliberately applied such as products applied to the track. Others are not deliberate, but still have an effect, such as when steel reacts with oxygen to make oxide layers, and when leaves are crushed under the large forces in the WRI to create a very low friction coating, chemically similar to Teflon, or simply water that acts as a lubricant.

Friction management involves either increasing or decreasing the friction from its normal state. Friction enhancers increase friction and lubricants decrease friction. The effects of these modifiers are seen after a wheel enters full slip [22]. Figure 14 shows curves for positive friction and negative friction. Positive friction relates to products that enhance friction and negative friction relates to products that reduce friction (lubricants). Lubricants include commercially used grease used in trackside and train mounted applicators, water and biological contaminants such as leaves. Figure 15 demonstrates the variation in rail conditions due to third bodies that create a large range of friction conditions. In this work lubricant will refer to deliberate lubrication using products through applicators or be specified as water lubrication.
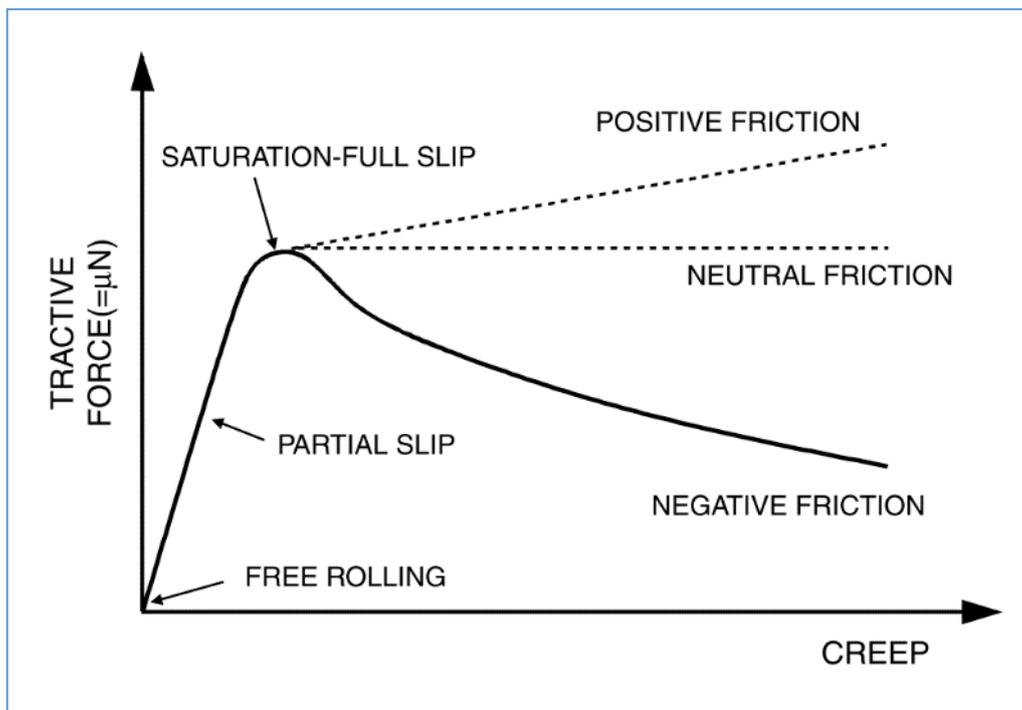
Figure 14: Traction-creep curve showing how friction modifiers play a role after full slip occurs [22].

There are also products known as friction modifiers that reduce friction to a preferred CoF value of 0.3-0.4, rather than a dry rail. This does not make them a lubricant though, as lubricants aim to reduce the friction to a minimum. Further developments have led to top of rail lubricants that provide a non or slow drying third body in the WRI to allow further control of the friction. Unlike friction modifiers that operate in their dry state, top of rail are either an oil, grease or water-oil hybrid that remain liquid [23].

Lubricants, top of rail lubricants and friction modifiers are deployed in order to reduce the damage to the rail that is caused by too much friction, i.e. wear and Rolling Contact Fatigue (RCF) [24]. This is the same principle as protecting steel roller bearings from abrasion [25] except in the WRI, the lubricant dries up, washes away and becomes contaminated with debris and other liquids. The correct amount of lubricant needs to be used to protect the rail long enough that the train navigates the whole curve. However, if too much is used then the lubricant could detrimentally affect traction. Therefore, it needs to be present when needed, but then absent for speed changes, making optimum selection and application of friction management products a challenge on a complex track layout.
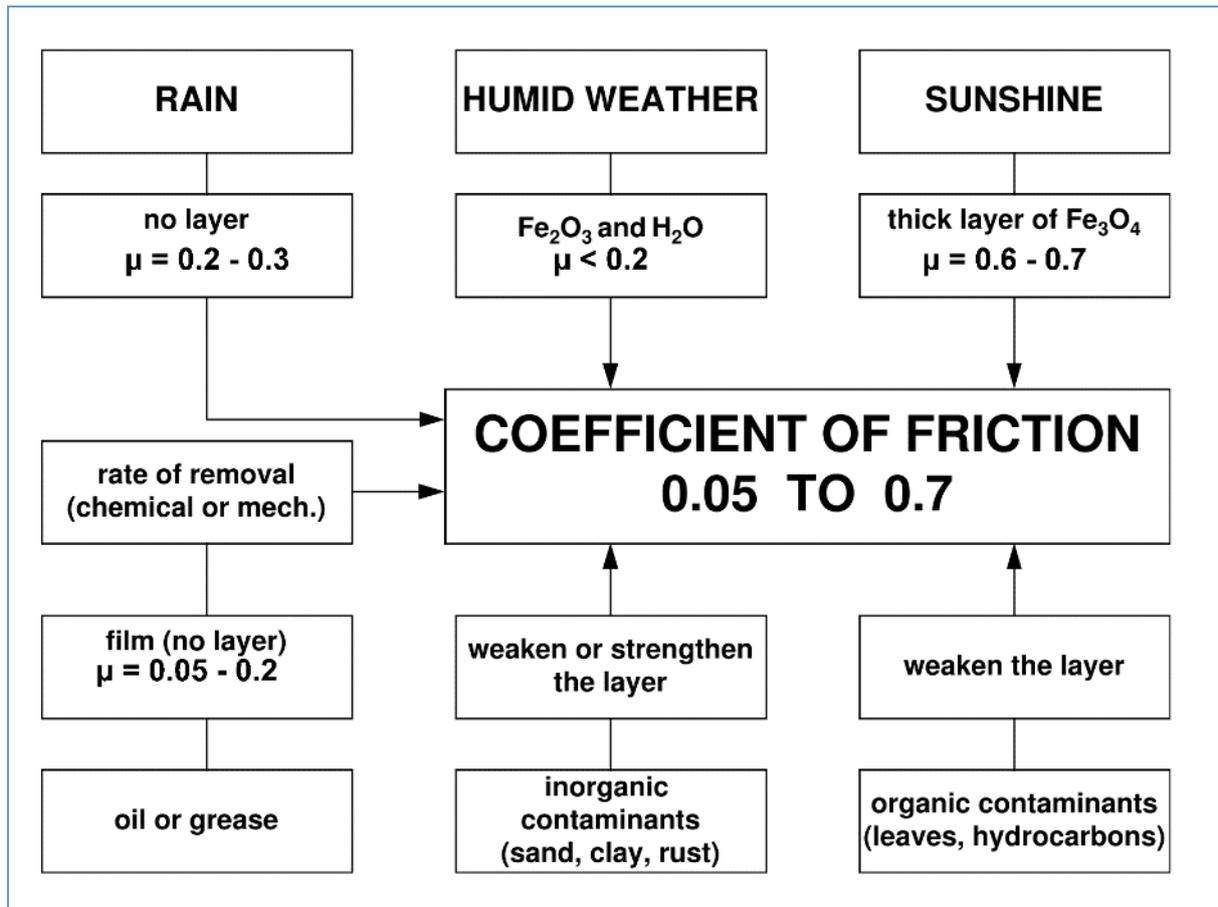
Figure 15: Factors that can cause the coefficient of friction to vary from 0.05 to 0.7 [16].

2.2.7 Lubricators

Lubrication can be added to the track automatically in two ways. Either by the train onto its wheels using a lubricator stick that presses against the wheel tread, transferring the lubricant. Alternatively, using trackside lubricators that dispense lubricant onto the railhead for the wheels to pick up as they pass over. Track side lubricators can be positioned before curves when the train needs the protection from wear and RCF the most but train mounted lubricators can cover the entire circumference of the wheel.

RCF and wear are undesirable forms of damage to the running surface of the rail and the wheel. RCF and wear are discussed more in section 2.6. Good friction management protects against this damage. However, studies have shown that intermittent lubrication can lead to an increase in RCF through the initiation of cracks during dry conditions followed by the lubrication of crack walls during wet conditions [26], [27]. There is also the concept of fluid becoming

trapped and then pressurised leading to crack growth via mode 1 rather than mode 2 [28]. The presence of surface defects also affects the detrimental effects that lubricants can cause [29]. On top of all of that, the presence of lubricant on one rail, but not the other can also lead to thermal damage to the rail [30] due to wheel spin.

The issue of liquid entrapment applies to friction enhancers too. The application of friction enhancers that remain liquid in the contact zone result in the exacerbation of any existing RCF damage. Therefore, the use of self-drying water based products (traction gels) that only leave the solid particles behind, are better where RCF already exists [31].

2.3 Track construction

Track construction has changed since Victorian times but the core details remain the same. There are many considerations to track construction, even when limited to the track without any of the supporting systems such as signals. To remain concise, most details are not going to be discussed here, only the details that have been identified as pertinent to understanding Squat type defects. This limits the content to track support, rail connection methods and the difference between high and low rail.

The rails are clipped to sleepers with pads between them. These pads help to reduce vibration (noise), distribute the stresses and reduce wear on the track system. The sleepers can be concrete, wooden, steel, twinbloc or slabtrack, which is continuous concrete rather than discrete sleepers. With the exception of slabtrack, sleepers are laid in ballast which is small rocks of a size that provides support and drainage. As the ballast breaks down under the forces of the passing trains, it needs the smaller pieces cleaning out and the rest resettling to maintain its effectiveness. A bad defect can cause the ballast to break down and this must be rectified to reinstate the desired track support. Under the ballast there is often another layer of subballast, then perhaps a sand layer which all sits on the formation which is either the natural ground or a constructed formation such as a cutting.

Figure 16 illustrates the construction layout of ballasted track. The construction allows the spreading of a very high load from the small contact area of the wheel on the rail, through slightly larger areas on the rail foot and sleepers into the ballast that spreads the load widely to inhibit high loads causing damage to the substructures, which could lead to long-term changes

in the track support. The ballast impedes movement of the sleepers, provides protection against frost, retards vegetation growth and allows maintenance of the track [22]. Techniques such as tamping and stone blowing can be used to maintain the ballast, but the ballast does degrade over time [32].
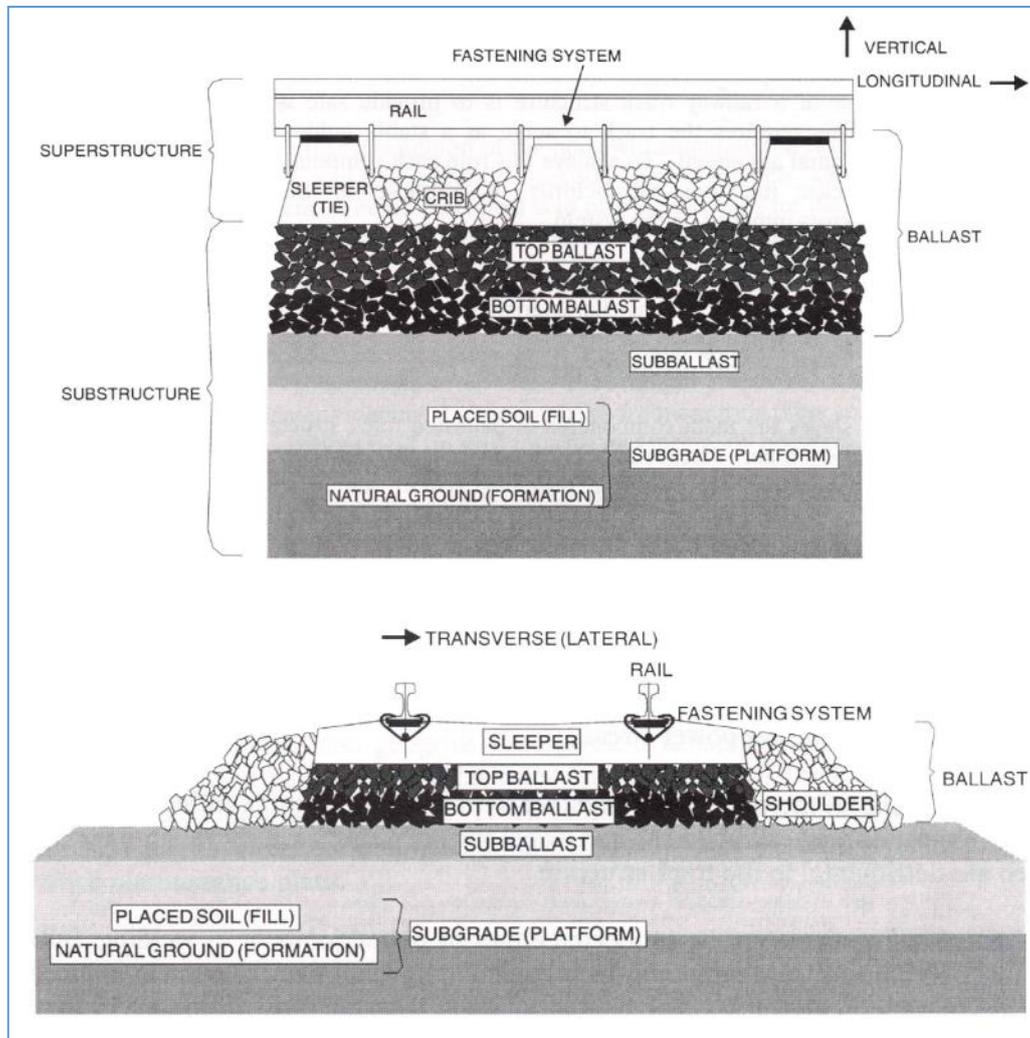
Figure 16: Typical construction of ballasted track [33].

2.3.1 Welds and fishplates

Welds and fishplates connect two rails together. The name "fishplate" has a naval historical origin from fish shaped plates that were used to connect two sections of a mast together. These are effective and useful as they can be unbolted to release a rail without the need for cutting. Figure 17 shows how fishplates connect rails and also shows an issue with this method. The issue is that because there needs to be an expansion gap left for temperature changes, the load

on the rail becomes very high as the wheel passes from one rail to another, causing material failure and plastic flow on the butt (end) of the rail. Another issue is that the bolt holes that in the fishplates and rail are exposed to high shear stress that leads to cracks forming at the edge of the bolthole and propagating outwards (Figure 18). This effect can be somewhat counteracted using a tapered drill-bit for drilling the holes, known as Cold Bolt Expansion. The drill-bit radius gets larger as it passes through the rail, introducing compressive stresses into the holes edge, thereby counteracting the tensile stresses experienced by the hole.

(1) Change in the running band at rail end

(2) Change of shape of rail end due to the discontinuity and loads applied
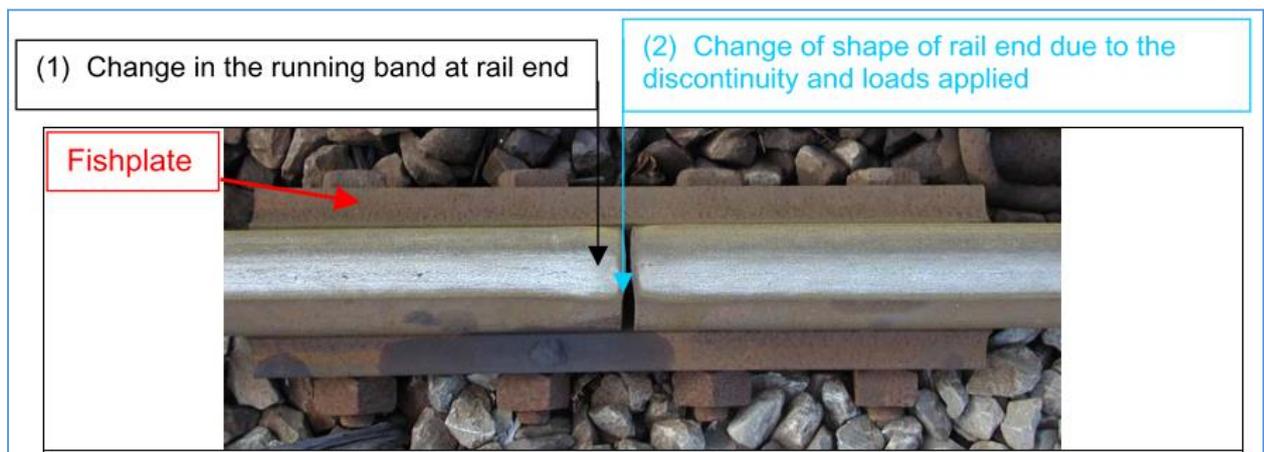
Fishplate

Figure 17: Aerial view of a two fishplates bolting two rails together. The deformation of the butt of the rail due to the gap between the rails is highlighted [34].

Figure 18: Bolt hole cracking [35].

Removing the gap by welding the rails together allows the wheel to be supported better over the gap and does not involve holes for cracks to initiate at. In order for the continuously welded rail (CWR) to also be able to expand in the heat too, the rails are hydraulically stretched during installation to minimise expansion on hot days. Expansion joints/ breather switches can also be implemented to allow relief from expansion.

Welding rails together in a factory environment can be done using flashbutt welding. This involves heating the butts of the rails using cycling electric currents to get them hot enough so that the rails butts can be forged together under high pressure creating an excellent weld, although it does not have the same material properties to the rails, making the weld prone to RCF developing [36] . Alternatively, aluminothermic welding thermite welding is an option. This involves creating an exothermic reaction and welding the two rails together with a filler material. As with the flashbutt weld, the material properties of the weld vary from the rail, making it a weak point for defect developing. An inherent issue with the thermal process of welding is that variations in the cooling rates, from the very hot centre of the weld to the cooler edges, leads to different microstructures. These microstructures have a different hardness and this variation is known as the Heat Affected Zone (HAZ).

2.3.2 High and low rail

The high rail is the outer rail of a curve and sits slightly higher above the equilibrium line than the low rail (Figure 19). This allows the train to corner without making flange contact with the rail. The ratio of the high rail to low rail (cant) depends on the speed of the trains expected on that route. Cant creates height differences between the high and low rail, up to 150mm in the UK, causing the train to lean into a curve. The high rail suffers from RCF. The low rail suffers RCF too, but also from wear and crushing, due to more of the train weight being over the low rail. An extreme example of the difference in damage between high and low rails can be seen in Figure 20.
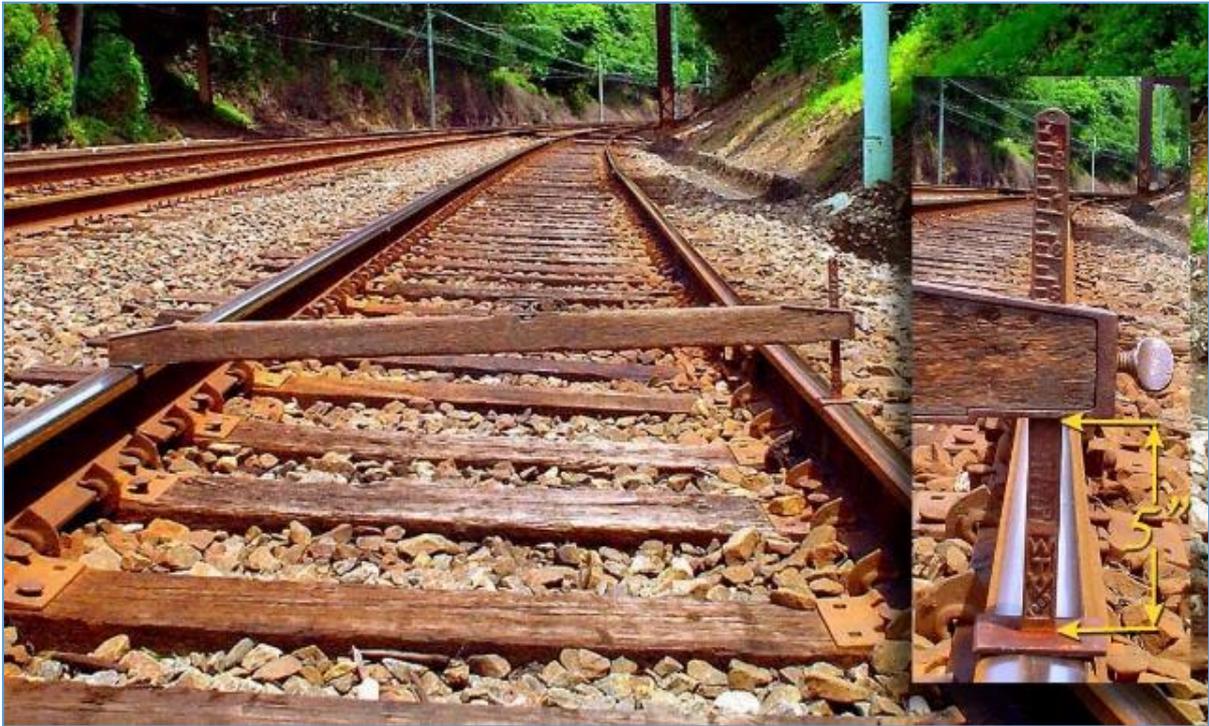
Figure 19: A level spirit gauge showing the difference in height of the high rail on the left to the low rail.
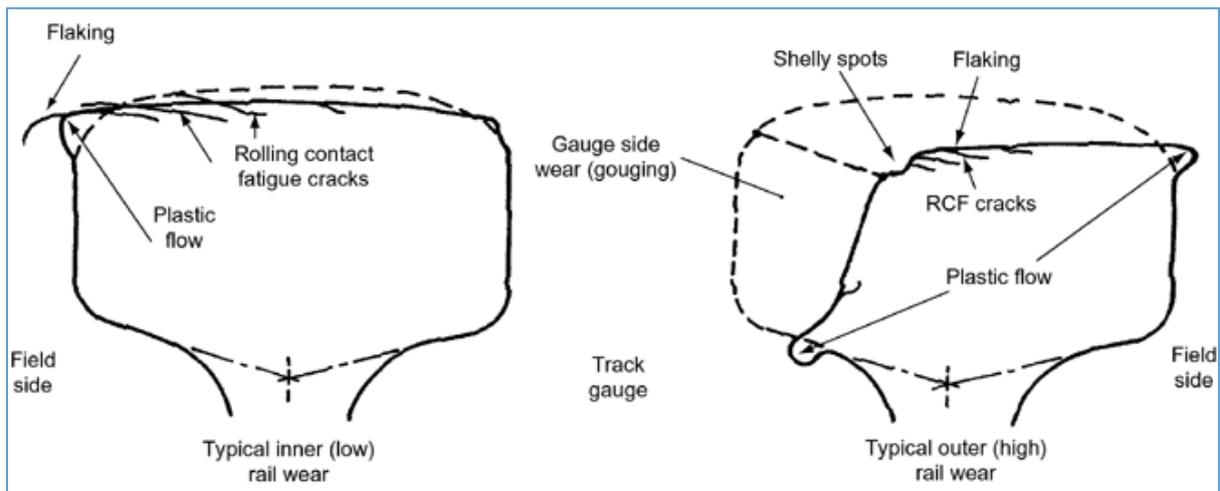
Figure 20: The effects of wear, RCF and plastic flow on a high and low rail profile. Grinding is needed to help to prevent this extreme degradation [37].

2.4 Rail Steel

2.4.1 Metallurgy

A key aspect of metallurgy is the control of the solidification of a metal as it cools from a liquid to a solid. The control of the solidification, or cooling of a metal partially determines its characteristics as a finished product.
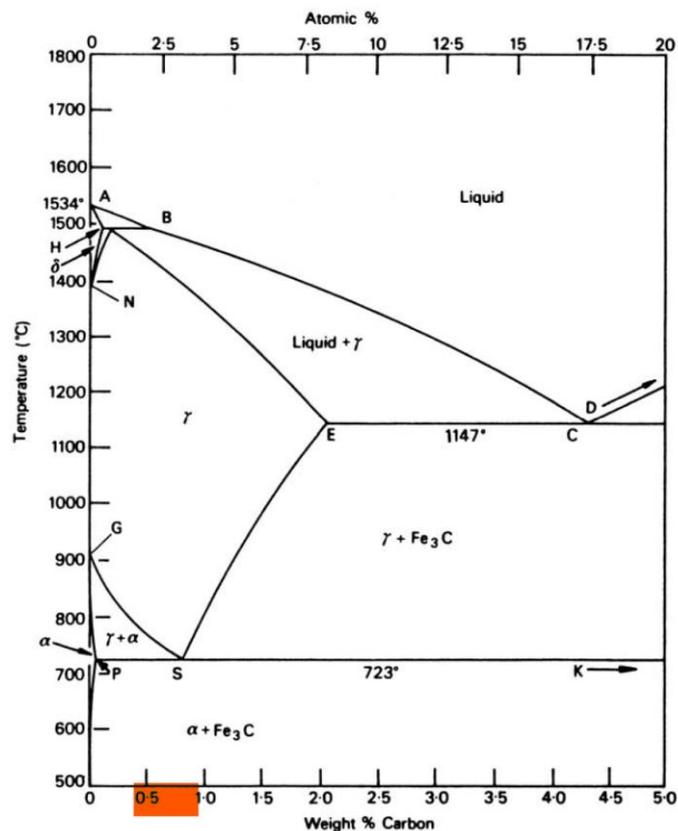
Figure 21: An iron-carbon phase diagram. Fe = iron and cementite is composed of iron and carbon [38]. *The red indicator shows where rail steels fall on the diagram.*

The iron-carbon phase diagram is shown in Figure 21. These diagrams are created experimentally using cooling curves [39] to determine the temperature at which phase changes occur. By knowing the carbon content of a steel and the temperature that it exists at, the combination of phases present can be calculated. Rail steels typically have 0.8%C, but the exact amount varies, as marked by the red bar on Figure 22. However, this is an equilibrium diagram and does not give any information about kinetics. The formation of bainite and martensite and

the retention of austenite are not evident from the diagram. Which phases are present and in what percentage is affected by the cooling rate of the steel. The phases present as a function of cooling conditions is shown by a Time Temperature Transformation (TTT) diagram (Figure 222). These diagrams are constructed for isothermal transformation, which is unusual industrially. Therefore, a continuous Cooling Transformation (CCT) diagram is often more useful, Figure 23.
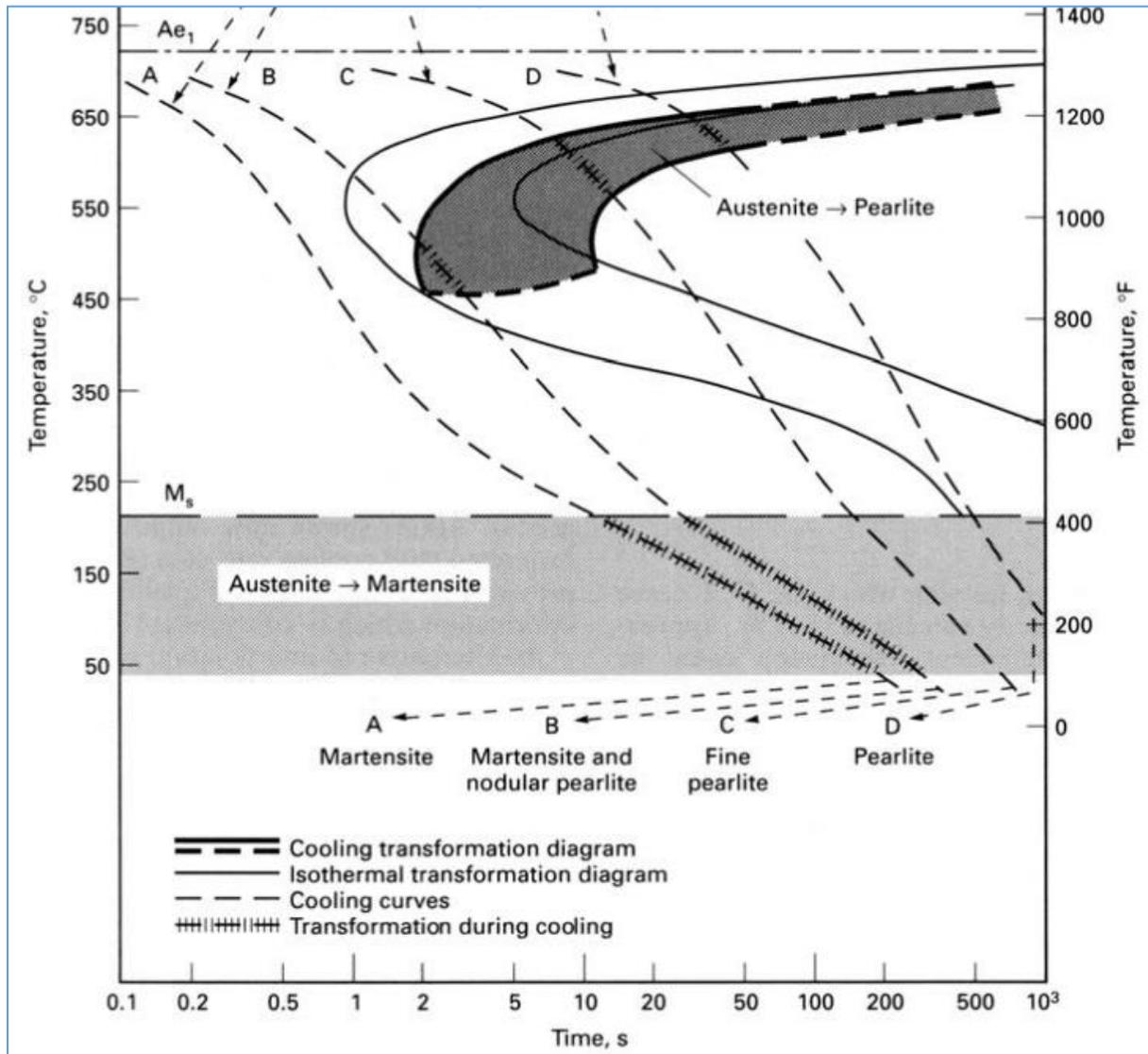
Figure 22: A TTT diagram for an 0.8% carbon steel showing 4 examples(A-D) of cooling regimes leading to different phase combinations [40].
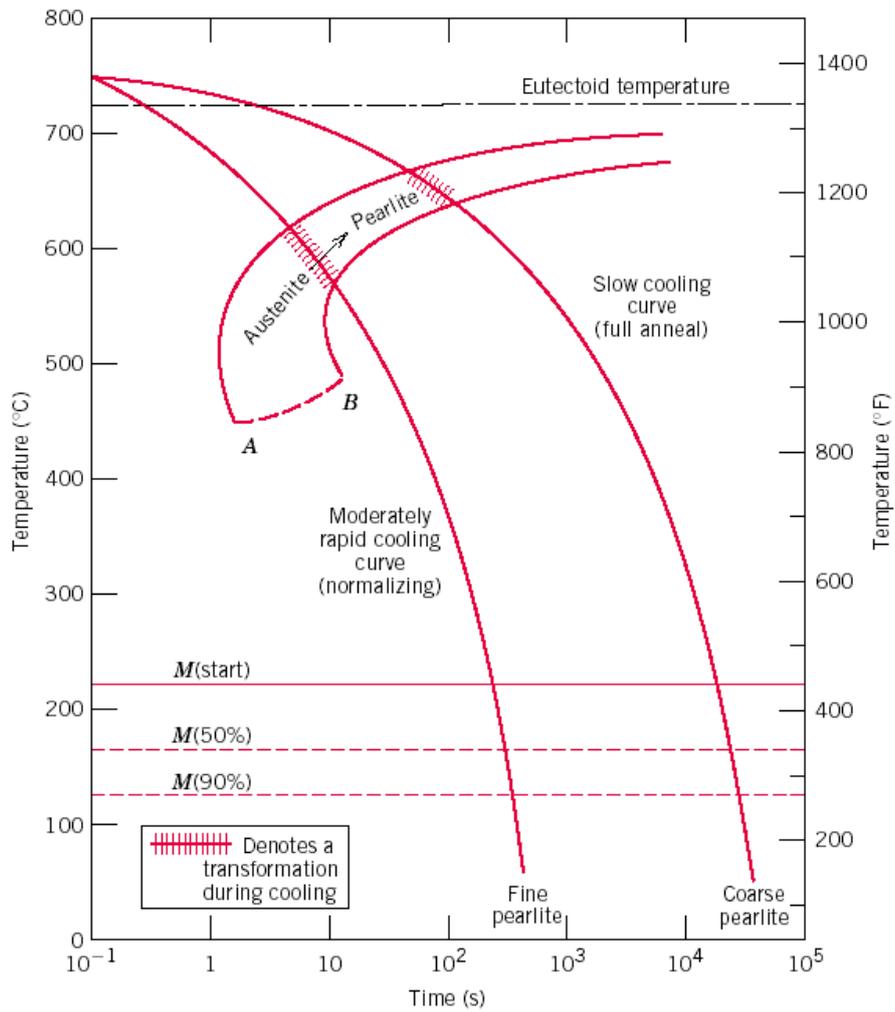
Figure 23: Continuous cooling transformation diagram for an 0.8% carbon steel [40].

Although a rail may be manufactured as 100% pearlite, if the surface of the rail is heated sufficiently through friction and then rapidly cooled due to the heat sink effect of the bulk of the material, martensite can form on the surface where pearlite used to be. The cooling of a manufactured rail will be relatively long due to the whole rail containing heat whereas surface heating will cool much faster even if the same temperature is reached.

Pearlite is named so because when viewed under a microscope in its etched form, it looks like 'mother of pearl'. The identifying feature of pearlite is its lamellar structure as seen in Figure 24, made up of alternating layers for ferrite and cementite. The darker contrast material is ferrite and the lighter contrast region is cementite (Fe_3C). This is the opposite of an optical image, where the ferrite exhibits lighter contrast than the cementite. It can be seen in the SEM image that the pearlite spacing varies from one region to another.

The lamellar spacing is inversely proportional to the temperature at which austenite transforms to pearlite. Therefore, a rapid cool produces a finer pearlite structure. A smaller inter-lamellar spacing (thinner lamellar) has been found to produce tougher and stronger steel [41]. Smaller grain size also increases the strength of the steel in accordance to the Hall-Petch equation.
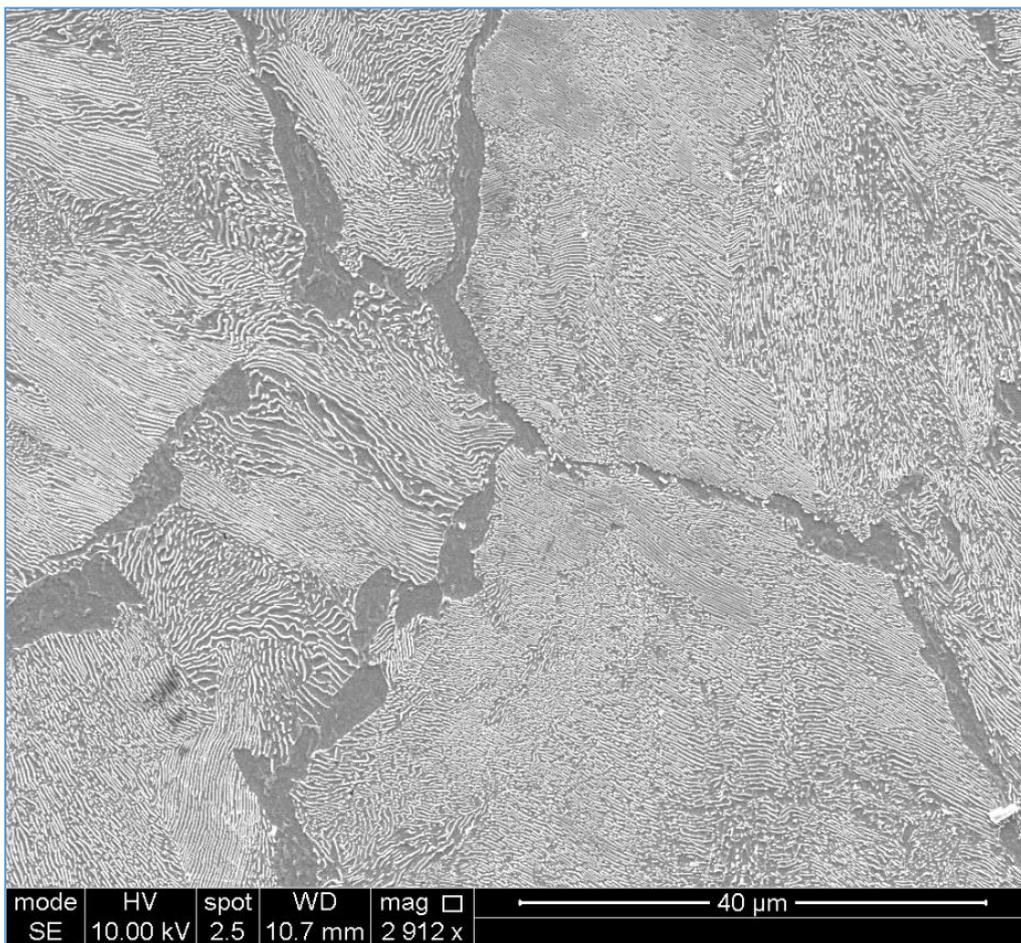
Figure 24: SEM image of pearlite. The prior austenite grain boundaries are decorated by ferrite.

2.4.1.2 Grain refinement

Mechanical work on the material can also refine the grain structure as well as deform it (ratchetting/ plastic flow). Grains can be broken down/ refined down to as small as 15nm [42] well below the resolution of the optical microscope and become classed as White Etching Layers (WEL), discussed more in section 2.5. Grains can also be refined by thermal events such as wheel slippage. The frictional heating reaustenitises the rail surface and as it cools again finer grains maybe produced or martensite maybe formed. In both cases, this can lead to residual stresses, which are detrimental to the fatigue resistance of the rail steel.

2.4.1.3 Martensite

As shown in Figure 22, a very rapid cooling of the steel from austenite results in the formation of martensite instead of the desired pearlite. The extremely rapid cooling during martensite formation does not allow time for diffusion of the carbon to form the thermodynamically predicted phases. Martensite is hard, brittle and contains residual stresses and cracks much more easily than pearlite (low toughness). Volume changes in the material due to transformation result in large residual stresses, a result of the transformation from austenite (FCC crystal lattice) to martensite (BCT crystal lattice) with the martensitic lattice having a larger lattice volume as a result of trapped carbon (Figure 25). The transformed martensite is constrained by the surrounding untransformed material, resulting in the large residual stresses.
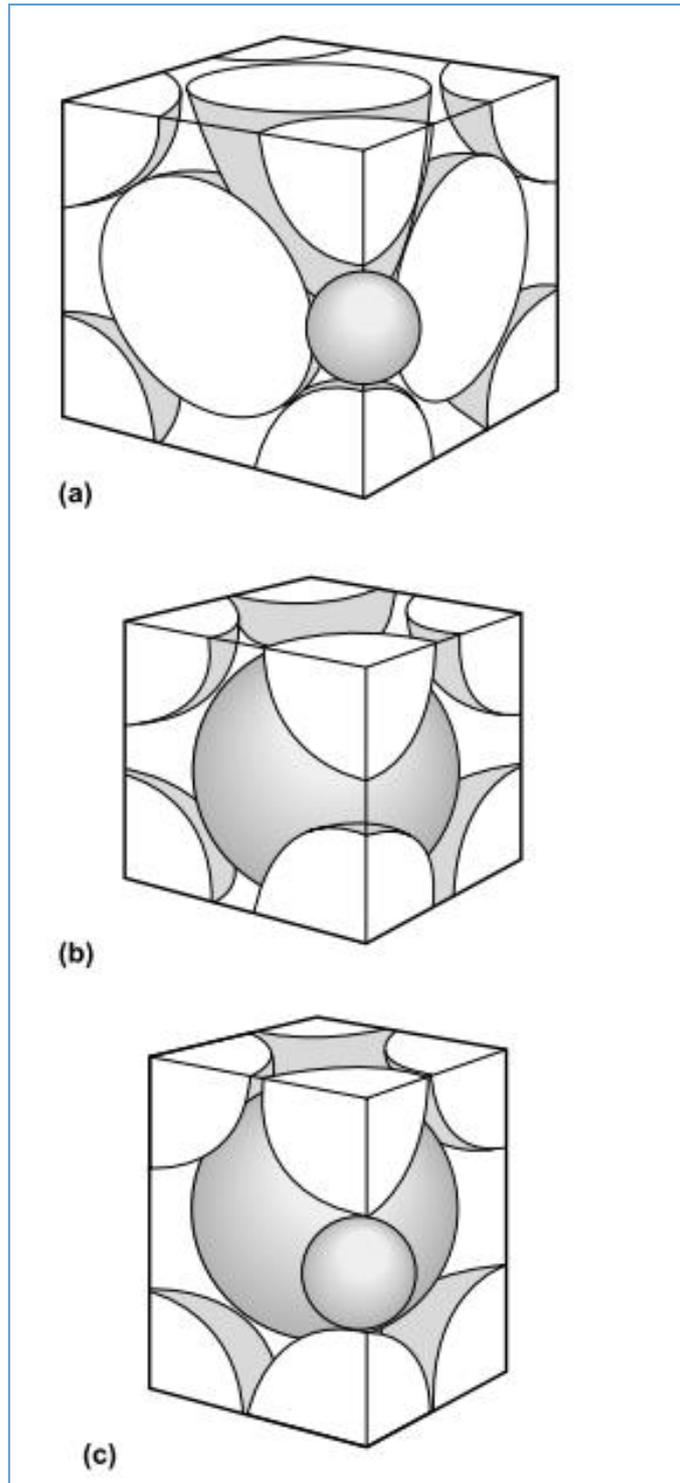
Figure 25: Crystal structures of a) Austenite (fcc), b) Ferrite/ iron (bcc) and c) Martensite (bct). (a) and (c) contain small carbon atoms but (a) is not being stressed like (c) because of the carbon slots interstitially between larger atoms [43].

2.4.2   Grades

Through advancements in metallurgy and the change in demand from the evolving railways, different grades of steel are available for rails. Each grade has characteristics that make it good for certain deployments in track but a poor choice for others. For example, a heavy freight line would need a harder, more wear resistant grade than a metro line. That same hard rail would suffer from fatigue due to its lower toughness and so increased crack growth if the traffic was very light (not enough wear).

2.4.2.1 Standard Grades

The standard grade of rail in the UK is now R260, where the 260 represents the minimum hardness of the running band in Hardness Brinell (HB). Prior to R260, R220 was the standard grade, but as the hardness measure suggests, this was a soft rail that had quite a high wear rate. Wear rates increased with the increase in axle loads of both freight and passenger trains. R260 is a 100% pearlitic rail that has very little ferrite on the prior austenite grain boundaries unlike its softer R220 predecessor.

R260 has 0.62-0.8wt% carbon, which means it sits just to the left of the eutectic point in the phase diagram (Figure 21). R220 has 0.4-0.62wt% carbon. Rails need enough carbon to make them hard enough to resist wear but still be easy to weld. High carbon steels are more difficult to weld. Rails also need to have good fracture toughness, fatigue resistance, tensile strength and other properties that a plain carbon steel alone cannot deliver, so rails typically have some other alloys to meet these requirements. Examples of some common rail grades and elemental compositions are in Table 1.

Table 1: Chemical compositions of various rail steels

BRITISH STEEL

Rail
Steel grades

Steel compositions and properties

High speed and mixed traffic

Specification	Grade	Chemical composition								
		% by mass								
		C	Si	Mn	P	S	Cr	Al	V	H2 (ppm)
UIC 860-O	700	0.40/0.60	0.05/0.35	0.80/1.25	≤ 0.050	≤ 0.050				
	900A	0.60/0.80	0.10/0.50	0.80/1.30	≤ 0.040	≤ 0.040				
	900B	0.55/0.75	0.10/0.50	1.30/1.70	≤ 0.040	≤ 0.040				
EN 13674-1	R200	0.40/0.60	0.15/0.58	0.70/1.20	≤ 0.035	0.008/0.035	≤ 0.15	≤ 0.004	≤ 0.03	≤ 3.0
	R220	0.50/0.60	0.20/0.60	1.00/1.25	≤ 0.025	0.008/0.025	≤ 0.15	≤ 0.004	≤ 0.03	≤ 3.0
	R260	0.62/0.80	0.15/0.58	0.70/1.20	≤ 0.025	0.008/0.025	≤ 0.15	≤ 0.004	≤ 0.03	≤ 2.5
	R260Mn	0.55/0.75	0.15/0.60	1.30/1.70	≤ 0.025	0.008/0.025	≤ 0.15	≤ 0.004	≤ 0.03	≤ 2.5
	R350HT	0.72/0.80	0.15/0.58	0.70/1.20	≤ 0.020	≤ 0.025	≤ 0.15	≤ 0.004	≤ 0.03	≤ 2.5
	R350LHT	0.72/0.80	0.15/0.58	0.70/1.20	≤ 0.020	≤ 0.025	≤ 0.30	≤ 0.004	≤ 0.03	≤ 2.5
	R370CrHT	0.70/0.82	0.40/1.00	0.70/1.10	≤ 0.020	≤ 0.020	0.40/0.60	≤ 0.004	≤ 0.03	≤ 1.5
EN 13674-2	R260Cr	0.40/0.60	020/0.45	1.20/1.60	≤ 0.025	≤ 0.025	0.40/0.60	≤ 0.004	≤ 0.06	≤ 2.5
IRS	880	0.60/0.80	0.10/0.50	0.80/1.30	≤ 0.030	≤ 0.030	-	≤ 0.015	-	≤ 1.6
	1080HH	0.60/0.80	0.10/0.50	0.80/1.30	≤ 0.030	≤ 0.030	-	≤ 0.015	-	≤ 1.6
British Steel	Premium grade: BLF320	0.10/0.30	0.80/1.80	1.20/1.80	≤ 0.025	≤ 0.025	0.30/0.80	≤ 0.004	-	≤ 2.0
	Premium grade: BLF360	0.20/0.40	0.80/1.80	1.20/1.80	≤ 0.025	≤ 0.025	0.30/0.80	≤ 0.004	-	≤ 2.0
	Premium grade: HP335	0.87/0.97	0.75/1.00	0.75/1.00	≤ 0.020	0.008/0.020	≤ 0.10	≤ 0.004	0.09/0.13	≤ 2.5
	Premium grade: SF350	0.72/0.80	0.15/0.58	0.70/1.20	≤ 0.020	≤ 0.020	≤ 0.15	0.004	≤ 0.03	≤ 2.5
	Premium grade: SFL350	0.72/0.80	0.15/0.58	0.70/1.20	≤ 0.020	≤ 0.020	≤ 0.30	≤ 0.004	≤ 0.03	≤ 2.5

A distinct difference between R220 and R260 under an optical microscope is the thicker alliotromorphic ferrite on the prior austenite grain boundaries in the lower carbon R220, as seen in Figure 24. This ferrite is soft and so easier for cracks to travel along. R220 also has ferrite on its grain boundaries and Figure 26 shows how a crack prefers this path rather than trying to cut across a grain through the pearlite.
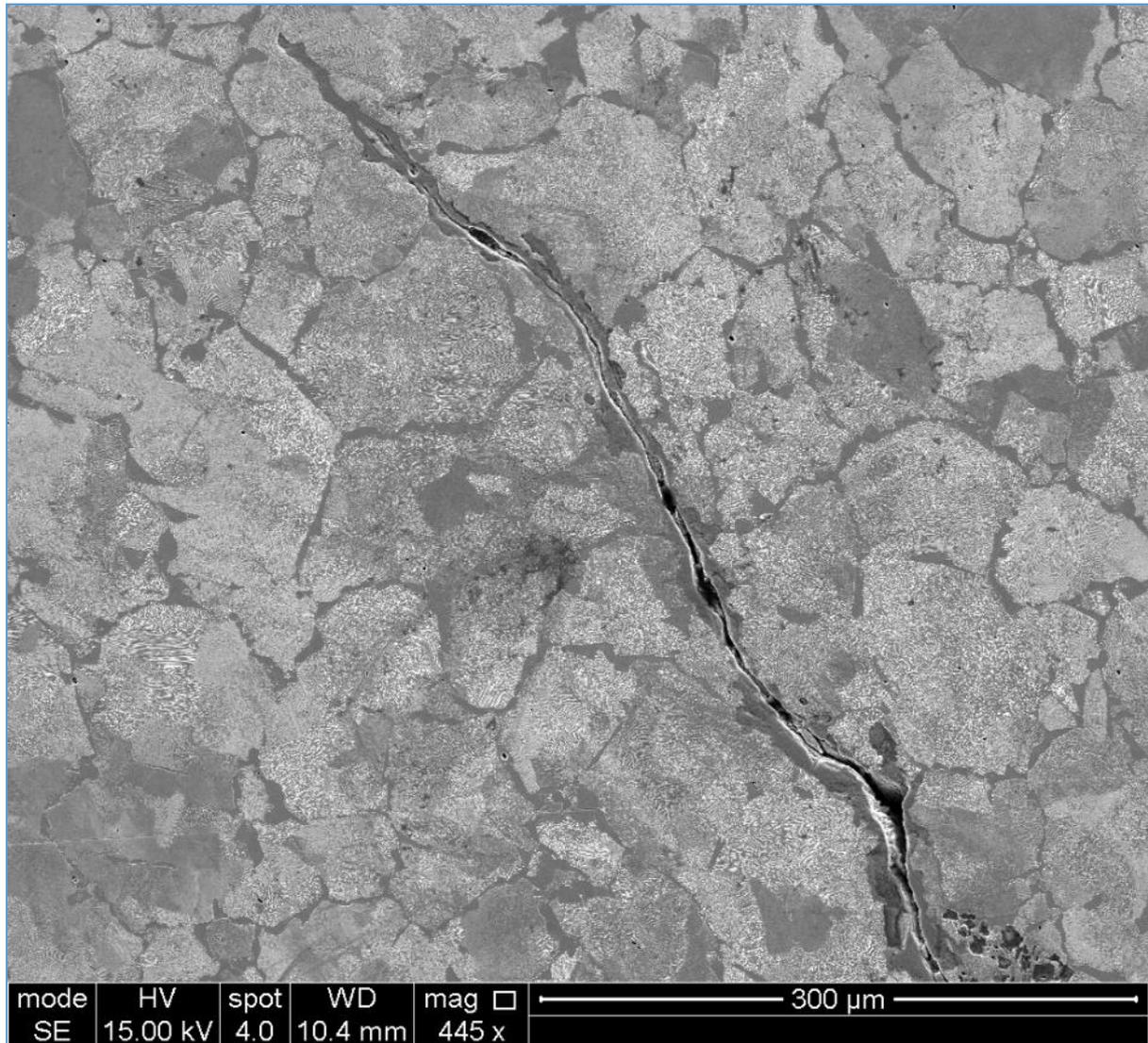
Figure 26: SEM image of a crack growing along a ferrite (dark) boundary in R220 grade steel.

2.4.2.2 Heat Treated Grades

R350HT is considered to be the standard heat-treated (HT) grade for rail track. It is not examined in this work but is included because Squats have been linked to heat treated grades in some research, as discussed more in section 2.7.

The refinement of grains to increase strength and toughness at a cost of ductility led to the production of heat treated grades of rail steel. This utilises the cooling and quenching mention in 2.6.1.1 to change the grain size at the surface of the rail to make it harder, therefore resisting wear, and tougher to resist crack initiations. Heat treated rails are considered to be a good investment for the reducing the overall life cycle cost for track operators [36], [44] and resisting

corrugation [45]. However, some research has found heat treated grades to be more susceptible to spalling than non-heat treated (standard) grades [46], [47]. This may be due to the smaller interlamellar distances making the rail more susceptible to cementite dissolution [48]. Heat treated grades provide a better wear rate than standard grade rails, but this shifts the risk of failure into the crack growth/ RCF domain when considering the magic wear rate principle [37]. Therefore, the use of heat treated rails comes with great benefits but the grinding regime (artificial wear) may need to be adapted to match the different natural wear rate of the rail.

A commonly used heat treated grade of rail is R350 HT. The European Innotrak project concluding technical report [44] advised the use of 350HT rail for:

- Lightly loaded tracks up to 700m to 1000m depending on boundary conditions
- Heavily loaded tracks with a radii between 3000m on 5000m (wide curves)

For medium and tight curves and after 370CRHT was recommended due to its higher wear rate. 370CrHT is a high chromium version of a 350 HT. More recently there has been an option of a R400HT rail, and as with other heat treated rails, this is considered to be a premium rail grade.

2.4.2.3 Premium grades

As mentioned above, R400HT is a relatively new rail. It was introduced in an ongoing effort to reduce wear, plastic deformation and the initiation and growth of cracks [49]. However, it is not necessary to heat treat rails for them to be considered premium. There is also another relatively new premium grade rail that has an increased carbon content in order to achieve a higher hardness. This rail, designated HP335, is a hypereutectoid steel and is also useful where a reduction in rail damage is required. Therefore, as the name suggests, a premium grade provides better damage resistance in high demand areas such as tight curves and heavy track loads. However, it is not suitable to use premium grades everywhere as a high wear resistance is not always suitable.

The reason there are so many grades available, many more than have been mentioned here, is because there are multitudes of track layouts that all require a different grade. Simply put: the

right grade is needed for the right situation. Premium rail is still fairly new so its performance against Squat type defects is still being investigated.

2.5 Rail degradation

This section will start to address the focus of this work, rail surface defects. It will begin by discussing the two main degradation mechanisms in rail steel, fatigue and wear, including factors that influence them. Then thermal damage will be discussed, as it can cause cracking without the need for cyclic fatiguing. This chapter will then conclude with a brief overview of some of the surface defects linked to Squats: the main topic that is covered in the next chapter.

2.5.1 Rolling Contact Fatigue (RCF)

RCF is high cycle fatigue of a steel rail by the passing of rolling wheels. This is not unique to railways, it occurs in systems when there is a rolling contact motion such as in bearings [50]. Unlike bearings, a railway contact is not a well-controlled environment. Intermittent lubrication causes slip of the wheels, high tractive forces, fluctuating friction and poor wheel-rail profile matching cause areas of high stress such as on the gauge corner. Surface damage such as RCF also increases at higher speed due to an increase in the dynamic wheel/rail forces [51]. Head checks (including GCC) and Squats are considered the two major forms of RCF on railways [52]. Increasing the hardness of the rail used can help to resist the development of RCF [53].

2.5.1.1 Crack Initiation

Damage to the surface of the rail from wheel passes results in changes to the surface microstructure, which leads to crack initiation sites. This damage can be thermal or mechanical. Thermal damage is discussed more in section 2.5.3. Mechanical damage can be seen in the form of ratchetting as discussed in section 2.2.4.

Ratchetting of the surface of a rail leads to strain accumulation, deformation and hardening of the surface material, re-orientating the surface microstructure. These changes leave the material prone to cracking, especially along the pro-eutectoid (PE) ferrite of the grain boundaries. However, this preferred path along the PE ferrite can lead to early stage crack blunting if the crack reaches the limit of the ferrite boundary, such as when the boundary changes direction

and is no longer in line with the maximum strain direction (Figure 27). Crack initiations are dominated by the microstructure of the material whereas the crack growth behaviour becomes dominated by the stress fields induced by the in-service environment [54]. This is important when trying to understand why not all RCF cracks develop into a more serious defect. Cracks can also initiate from hard and soft inclusions, although cleaner steel due to continuous casting methods has reduced this issue.
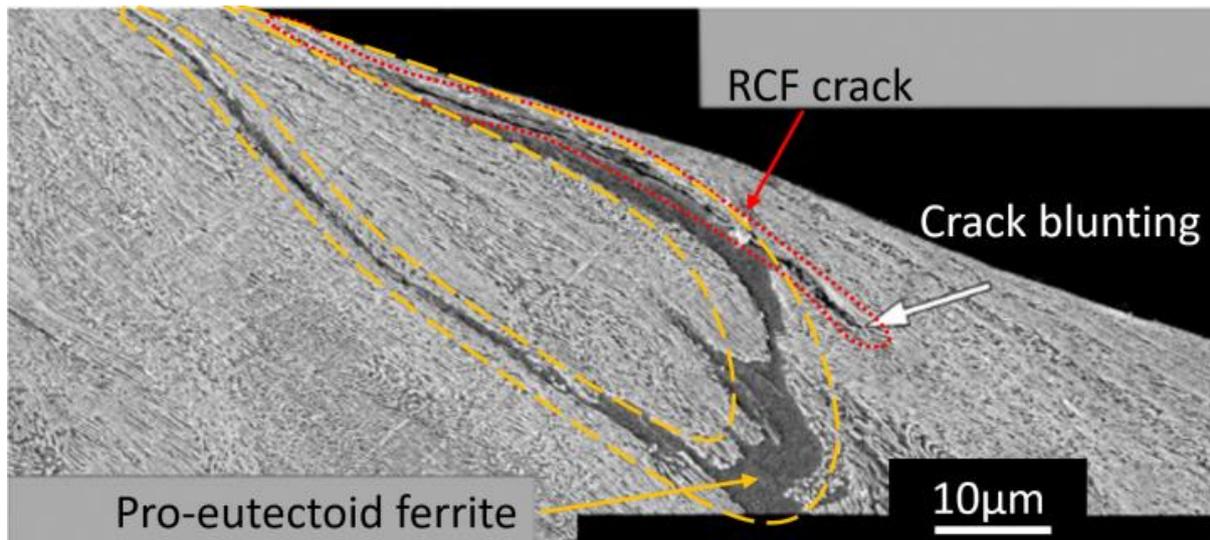
Figure 27: A crack tip blunting in pearlite after leaving the PE ferrite [54].

In the case of RCF in railway track, the two sets of wheels in a bogie have different roles in crack development in curves. On the low rail, the trailing wheelset initiates the cracks and then the leading wheels from the following bogie applies high stress to grow the cracks leading to spalling [55].

2.5.1.2 Dynarat 'Brick' model

The brick model was developed as part of simulations into RCF damage and ratchetting [11] and helps to explain crack initiations. It illustrates how a bulk material can separated into a grid system to produce layers or 'bricks'. When the ductility of a brick is exhausted through cycles of stress, that brick will fail. If that brick is on the surface then it will be lost as a discrete piece of wear debris, but if the brick is subsurface it will be retained and act as a microcrack that can interact with other failed bricks to cause subsurface damage. This means that a material does not have to fail from the surface downwards in sequence, nor does it have to fail a layer at a

time. It can fail in small discrete regions that link up as more bricks fail, meaning a brick failure is a potential initiation for a larger crack to develop. This is important when considering subsurface initiations and localised surface damage.
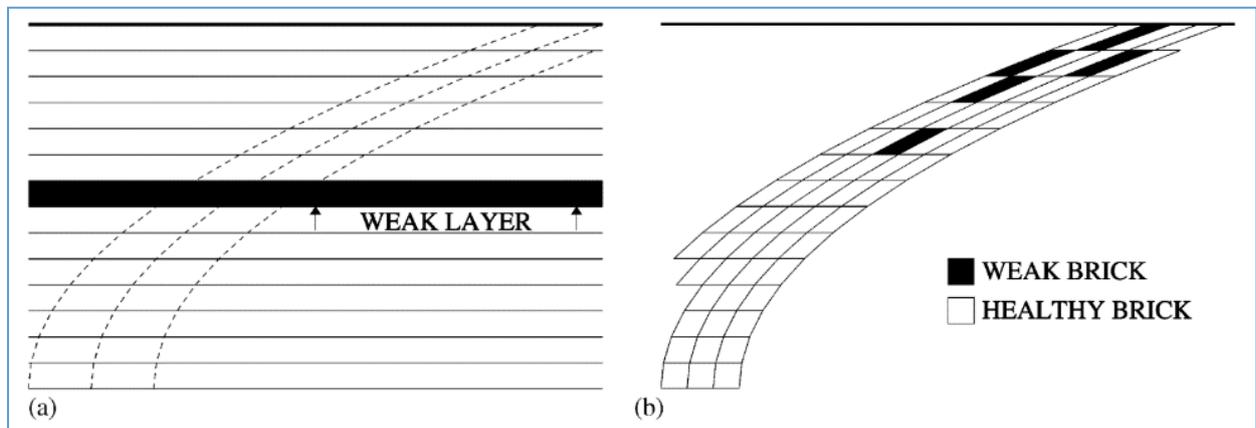
Figure 28: The brick model. a) A subsurface layer failure below intact material. b) Single brick failures that are fully contained [11].

2.5.1.3 Crack Propagation

Crack propagation immediately after initiation is quite slow in rail steels, but as the crack gets longer the propagation speed increases [56]. This is notable from the distance between beach marks on a fracture surface as shown in Figure 29. The beach marks look like the rings of a tree, and just like a trees rings, a larger gap between the beach marks indicates a faster crack growth rate.

Figure 29 also shows that there were at least two crack planes growing downwards with what appears to be the tearing of the material (oxidised region) between crack planes. The smoother region shows steady crack growth that increased as the crack grew deeper and longer. The mottled region shows where the rail failed suddenly as the rail was no longer able to support a passing train. Rubbing of the crack faces will have removed some of the details.
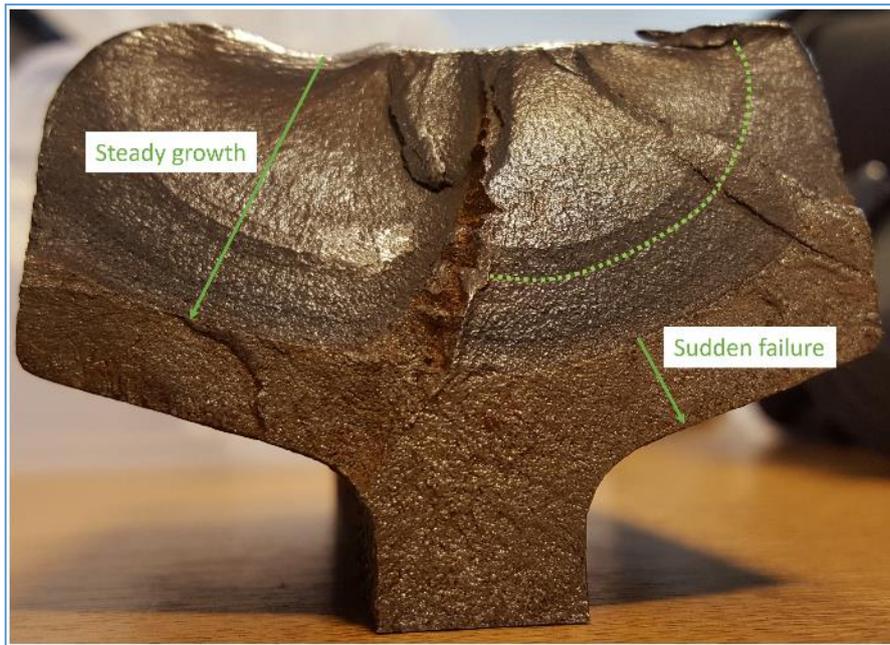
Figure 29: A transverse defect that grew from a Squat with the beach marks noticeable and indicated by the green curve.

Figure 30 shows how a crack initially grows rapidly due to ratchetting but then slows down into a varying crack propagation speed based on the contact stresses experienced and the length of the crack. Once the crack leaves the Hertzian contact area and enters the rail bending region it will grow rapidly and cause the rail to fail. Knowing how quickly a crack grows is important in order to adjust grinding regimes to truncate any cracks that grow.
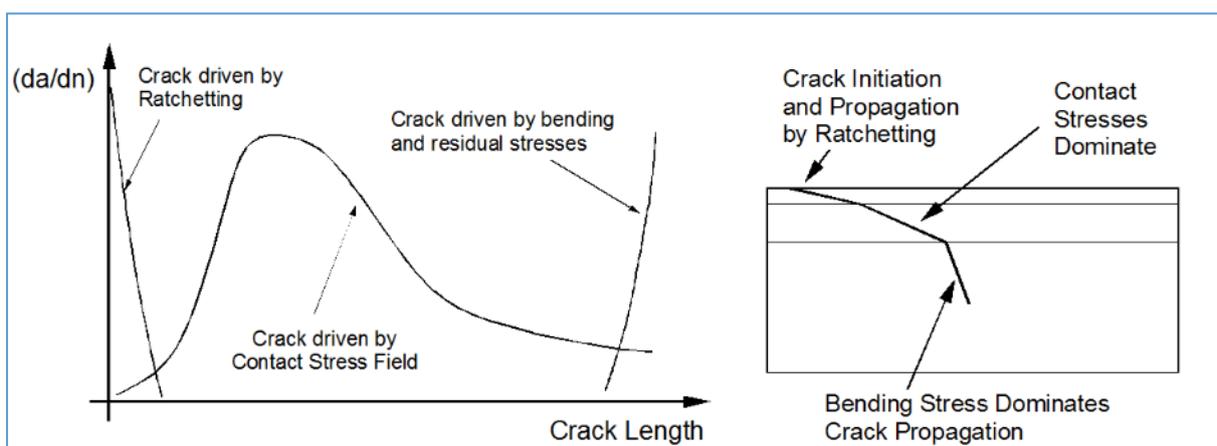
Figure 30: Drivers for crack growth. Ratchetting, leading to contact stress driven and finally rail bending [57].

Figure 30 illustrates how the crack growth rate can vary, even within the same regime. It also shows how the failure of a rail can be very rapid in the later stages, so a crack growing down into the rail may not seem very deep compared to the depth of the railhead, but it can grow very fast.

2.5.1.4 Fluid entrapment

Water has a low viscosity, especially when compared to oils used as lubricators for rail protection. It is known that water can penetrate cracks due to its low viscosity, but mineral oils are significantly less likely due to their much higher viscosity (~15x more viscous) [58], [59]. Water in a crack lubricates the crack faces, which allows shear crack growth (mode 2) to increase in intensity due to the lower friction between the crack faces (Figure 31a). In this scenario the water will increase crack propagation if the coefficient of friction between the walls is reduced to below 0.2 [59]. Furthermore, if the fluid is trapped in the crack, this can lead to pressurisation of the fluid which can pull the crack tip open, changing the crack growth to mode 1 (Figure 31c). However, once a surface cracks reaches ~15mm long, and 4-8mm deep, the fluid can no longer be confined by the contact patch, removing the effect of pressurisation [37].
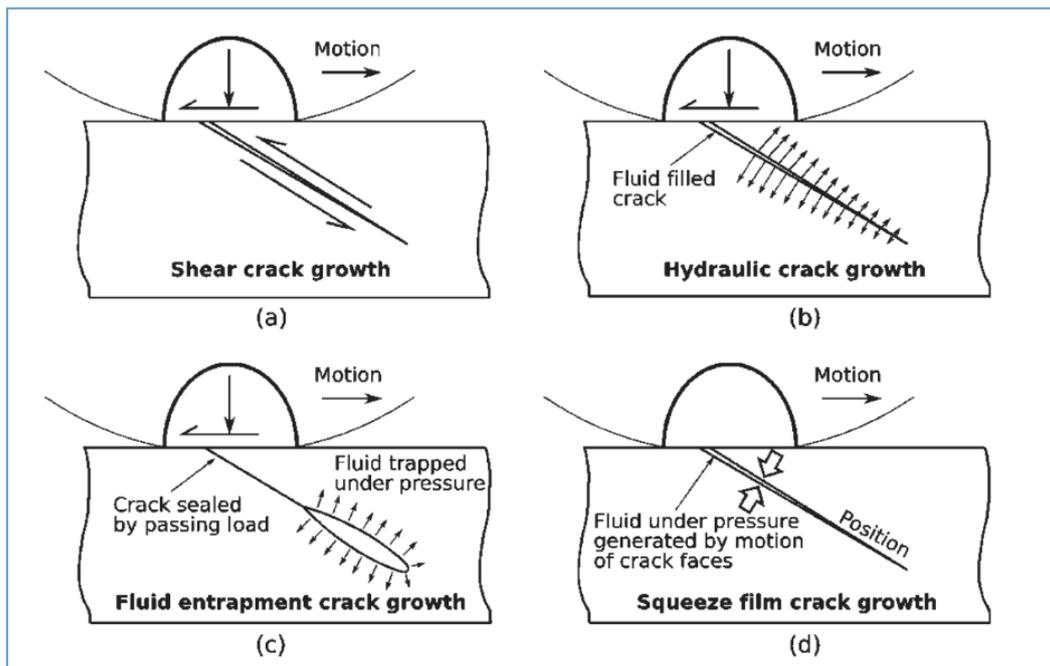
Figure 31: Interactions of fluids with cracks to produce different crack growth mechanisms.

2.5.2 Wear

As mentioned in section 2.4, wear is one of the areas of interest in tribology. When the friction and lubrication are not managed well, wear can become excessive and lead to premature failures in components. Wear in the wheel/rail interface (WRI) occurs on both the wheel and the rail. The hardness of the steel affects the wear rate as does the interlamellar spacing in the pearlite, although the hardness has a stronger correlation [60], [61].

2.5.2.1 Wear Types

Wear occurs due to the rolling and sliding of the wheel and the rail. The type of wear depends on the conditions. Due to the exposure of the WRI to the atmosphere, oxidative wear is common, where oxygen reacts with the materials in the sliding contact. Higher temperatures can increase the oxidisation and the resulting oxide type. Iron oxides have three phases, namely, Hematite, Wustite and Magnetite, which have different chemical compositions and lattice structures. Hematite is the brown 'rust' typically seen whereas Wustite is black and is produced at higher temperatures. Magnetite is typically found on meteorites due to the extremely high temperatures experienced entering the atmosphere, so is not relevant here. The expected oxide can be determined from a Fe-O phase diagram.

Adhesive wear is experienced during high load contacts and involves the transfer of material from one body to another. Abrasive wear is the gouging of a surface by trapped particles of protrusions in the contact, (e.g. using sandpaper on a surface). Fatigue wear/ delamination is the cracking of a surface due to cyclic fatigue that leads to material being released from the bulk leading to pits and voids. Fretting wear is very specific and contains features of multiple wear types. It occurs when two surfaces rub together cyclically. Adhesive wear pulls debris from each surface, which oxidises producing very hard abrasive particles that are trapped in the interface and so continue to wear away the contacting surfaces. Fretting wear in rails can be very aggressive, producing a very fine black oxide substance. Rail fretting is identified as a key aspect of this work.

Archard and Hirst showed that materials experience two degrees of wear, mild and severe, and that mild wear followed by severe wear can lead to heavy damage of the contacting surface [62]. Rails experience different types and degrees of wear due to the varying day to day

operating environment (temperature, moisture, debris). Friction management is needed to try to control this environment.

2.5.2.2 Profile Matching

Profile refers to the shape of the rail or the wheel. It is important that the wheel profile and the rail profile fits as closely as possible in order to distribute the enormous weight of the rolling stock over as large of an area as possible. In Section 2.2, Figure 2 shows an example of a wheel and rail profile coming together. Profiles are not discussed in any depth here because the profiles for the wheels and rails responsible for the defect are not known in this work. The main consideration is that an ideal profile would have as large a contact area as possible to distribute the load over. Any reduction in this contact area, due to poor profiles such as hollow wheels or worn railheads, would increase the contact force per unit area. Poor profile conforming can turn a normal contact pressure of 1500MPa into an extreme contact pressure of 4000MPa [63]

2.5.2.3 The Magic Wear Rate

Wear and RCF are two major problems in track maintenance and the fine balance between them is crucial. If a rail wears faster than RCF cracks can grow (Figure 32), then the crack will be truncated and wear will dominate. In this scenario, the rail will fail from wear, which is easier to monitor than RCF cracks. Fatigue failure can occur if the wear rate is not high enough to truncate these cracks. Magel discusses the magic wear rate in more depth [37].
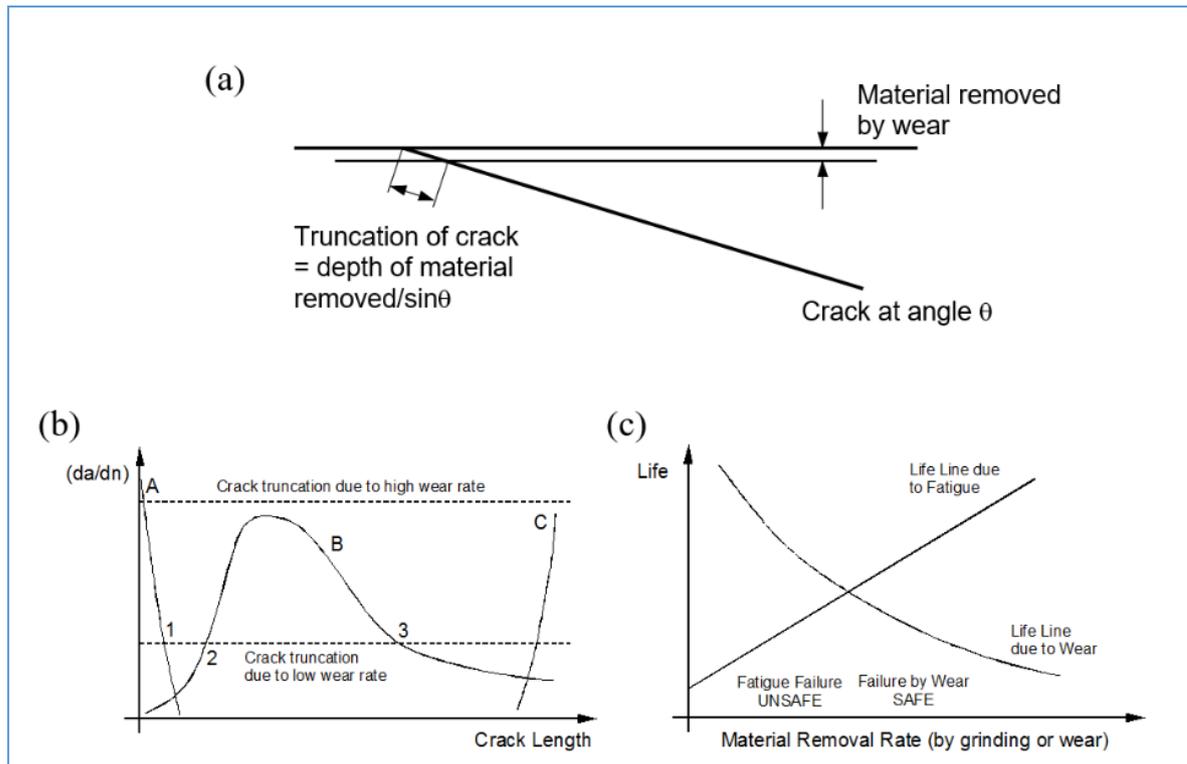
Figure 32: Interaction of wear and fatigue: (a) Crack truncation by wear; (b) Crack growth rate versus crack length; (c) Rail life versus material removal rate [57].

2.5.3 Thermal damage

The flash temperature rise in a contact leads to subsequent rapid cooling due to the large heat sink that the bulk rail provides to the small area affected. Rapid heating and cooling in turn leads to changes in microstructure and residual stresses in the rail. This can result in major microstructural changes with reaustenitisation of the pearlite with subsequent rapid cooling giving martensite, known as white etching layer (WEL). The expansion of the heated portion of steel causes residual stress because it is restrained in its expansion by the cooler bulk rail surrounding it and the wheel above pushing down on it. Thermal stress needs to be added to the mechanical stress that is being applied by the passing wheel so that:

$$\sigma_{mech} + \sigma_{th} = \sigma_{total}$$

2.5.3.1 White Etching Layer (WEL)

WEL was noted on rails in the early 80s [51]. The WEL does not etch and so appears white in a steel sample etched in nital, which gives it its name. It can be known by other terms, usually beginning with 'white etching', such as white etching matter (WEM) that appears subsurface in bearings [50]. This change in the phase of the steel is associated with thermomechanical damage and up to 50% of the rail surface in running lines are covered in this martensitic layer [51]. The structure of the WEL is generally considered to be martensitic, but WEL also contains carbides and some pro-eutectoid ferrite, and for a uniform WEL to form many cycles would be required [64], [65]. WEL could also contain some deformed pearlite, austenite and cementite [42]. Under a Scanning Electron Microscope (SEM) a little more is shown but not much. A Transmission Electron Microscope (TEM) is needed to see some detail in WEL and the reported compositions above are the results of these TEM investigations.

For a WEL to be formed thermally, its austenisation temperature would have to be reached. It has been suggested that when the rail is under multiaxial pressure it has a reduced austenisation temperature [65]–[67]. The exact reduction value varies fairly drastically from source to source, from as little as ~20°C to up to ~200°C reductions. The extent of WEL formation is increased by Mn content and reduced by Ni and Co [66]. Models of high temperatures producing WEL with good verification have been produced [18], but these conditions would not be considered to be normal for the whole rail. Simon [68] discusses the production of WEL through mechanical stress, which is more likely than high temperatures being produced across the entire rail section. Two WEL classes were presented by Al-Juboori [69]; fine martensite was considered mechanical by origin, whereas retained austenite and martensite was thermal. WEL can be a combination of the two.

Microstructural changes due to thermal input in the upper surface can cause a heterogeneous microstructure in the upper 500 microns of the railhead [70]. WEL has a deformed structural lattice and inserted carbon atoms that result in residual stresses that allow cracks to initiate in the rails surface more easily [71]. As well as cracking within the WEL, cracks have also been observed forming at the interface between the WEL and the softer pearlite [72], [73]. The increase in cracks with the WEL was noticed in the early 90s along with a difference between high and low rails and the effect of grinding on the presence of WEL (Figure 33) [47]. However, the boundary between the WEL and the parent pearlite can also hinder crack growth

[47]. In work by Carroll the boundary of the WEL can vary, from a sharp distinct boundary (sample RA34), possibly due to higher Cr content, to a more diffuse boundary (sample RA31). R350HT rail develops a thicker WEL than R260 [74], suggesting that the original microstructure/ chemistry of a rail affects the WEL formed.

Under the WEL there is a region known as the brown etching layer (BEL), possibly caused due to thermomechanical processes acting on the railhead. The border of the WEL and brown etching layer (BEL) is where the peak hardness occurs [42]. BEL is softer than the WEL but still brittle. It seems to be formed by mechanical deformation rather than thermal input, it may be a transitional phase between pearlite and the WEL and is a region where cracks propagate with different behaviour [75].
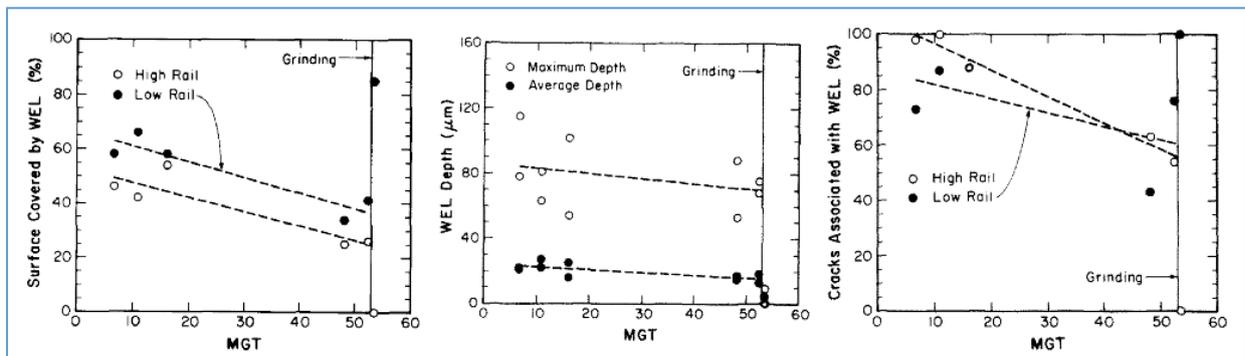
Figure 33: WEL as a function of traffic experienced in million gross tonnes (MGT) [47].

Figure 34 gives examples of a band of WEL with cracks in it, islands of WEL where cracks have appeared at the interface and the results of WEL even after the layer has been removed.
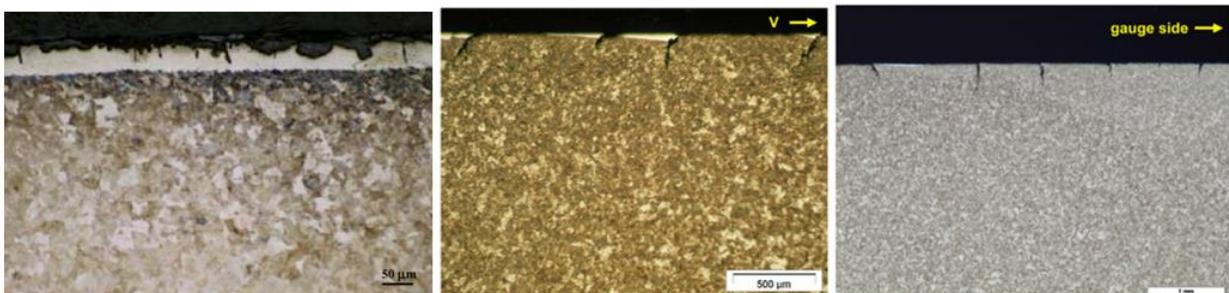
Figure 34: Images from left to right. WEL with cracks appearing [73]*, WEL where cracks have grown into the parent material and a sample that appears to have no WEL due to its removal via grinding* [76].

2.5.4   RCF related rail head surface defects

2.5.4.1 Corrugation

Corrugation in rails is waves appearing along the surface (Figure 35), resulting in peaks and troughs. This excites the suspension of the train causing it to oscillate up and down, increasing dynamic loading. Corrugation is grouped into six groups based on their frequency and is detailed in [77], along with their treatments, as they vary. Short pitch corrugation is not a problem on the Shinkansen (High Speed Lines) in Japan due to the use of soft rail pads [78]. Corrugation can be controlled using a tailored grinding regime that increases a rails life by nine times [37] or through the use of a harder rail. The interest in corrugation, which is a relatively long standing defect, is due to its link with Squat occurrences in the troughs of the waveforms [51].

Figure 35: Corrugation in heavy haul/ freight lines [77].

2.5.4.2 Shelling

Shelling is a crack that initiates ~6mm below the rail [79] resulting in the loss of a relatively large piece of material when the crack surfaces. "Deep seated shells" are shelling that occurs due to excessive loading, which leads to yielding and failure along slip lines, without the need for inclusions in the steel [16]. Spalling is found with very similar features to Squats including their presence amongst corrugation, WEL, V-shaped cracks and longitudinal cracks along the edge of the running band [79], [80]. Squats are known to spall in their latter stage and they

share features to the extent that a Squat can be mistaken for spalling or vice versa [46]. Some Japanese papers refer to "Shinkansen shelling". These features are called Squats in other countries and are not the same as shelling [78]. Black spots can indicate shelling as the surface deforms due to the subsurface crack, so black spots are not exclusive to Squats.

Figure 36: An example of shelling of the railhead [6].

2.5.4.3 Flaking and spalling

Flaking defects share some influencing factors with Squats, such as their prevalence on high rails, association with deep grinding marks and links with harder grades of rail. Although they are not shelling, they do involve the loss of material from the rail. They are not related to RCF, are found on the gauge corner of the rail in high curves and are also known as "gauge corner collapse" [79]. Canada recognise "gauge corner collapse" as shelling but do not refer to it as flaking as it is associated with inclusions [16], unlike flaking. The term "gauge corner collapse" may refer to slightly different defects on different continents. The US refer to flaking as an alternative term for mild spalling.

Grinding (covered in 2.8.1) roughness is directly linked to flaking: grinding marks are visible on the rail in Figure 37. Rough grinding leaves facets on the rail that are stress intensifiers (Figure 38), leading to plastic deformation and eventually cracks due to high subsurface stresses [10], [81]. Flaking occurs where there is excessive lubrication, showing the importance

of maintaining the correct friction coefficient. Lubricant is known to increase shear crack growth by lubricating the crack faces [28], [59] and fluid in cracks may be able to pressurise the crack, increasing its growth rate (Figure 39).

Figure 37: A discrete flaking defect (left) and a continuous flaking defect in R350HT rail [79].
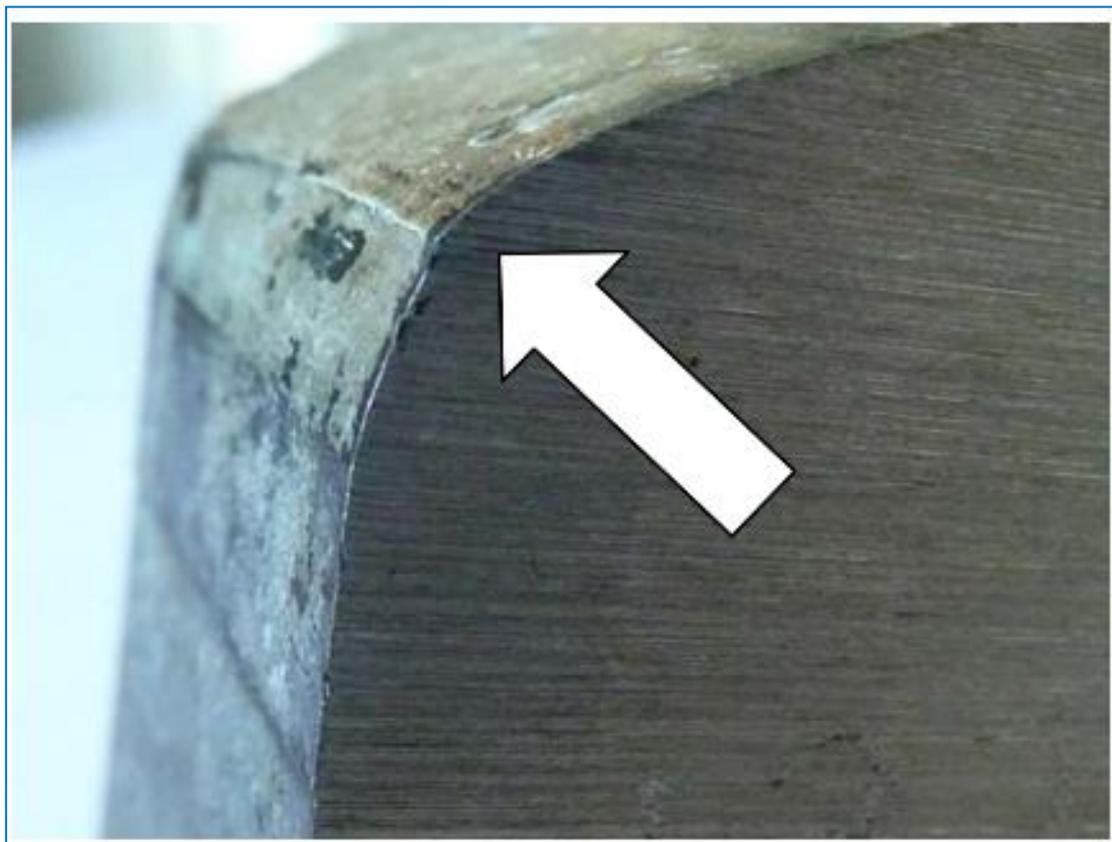
Figure 38: Flaking occurring at a ledge (stress intensifier) indicated by the arrow. The ledge is at the intersection of two facets [79].
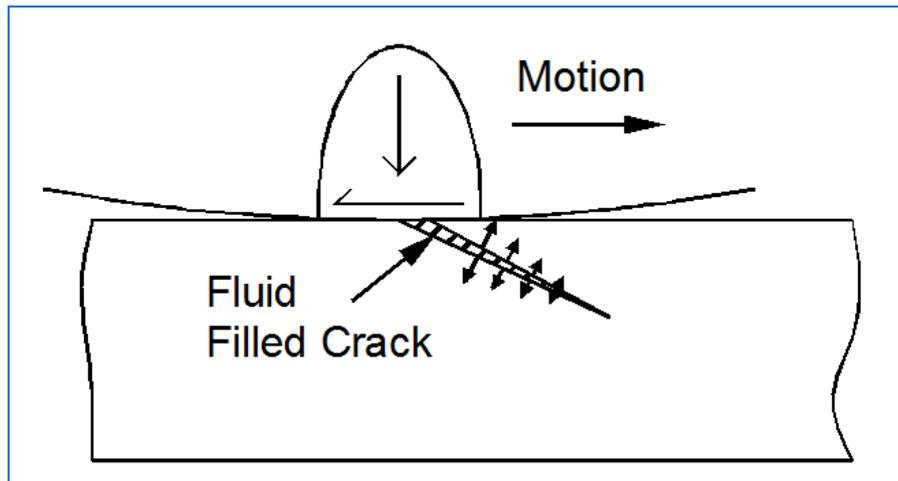
Figure 39: Crack in a rail opening by fluid pressurisation as a wheel passes over it.

Spalling is also a material loss defect like flaking and shelling except that it is localised and tends to be much shallower material loss than shelling (Figure 40). High contact stress and martensite transformations that break away from the bulk material are associated with spalling.

Figure 40: Moderate spalling of the running band [16].

2.5.4.4 Gauge Corner Cracking (GCC)

GCC and Squats also have some strong similarities but are not the same defect. They are linked in their similarities [82] and GCC has been seen to coexist as a developing defect (Figure 41) with Squats [76] and Studs [83]. GCC and Squats are grouped together in Japan [84] although there is acceptance that there are differences.

Sometimes GCC is referred to as "head checks". Head checks are still cracking of the steel surface just like GCC. The main difference is that head checks occur on the centre of the railhead as opposed to the gauge corner.
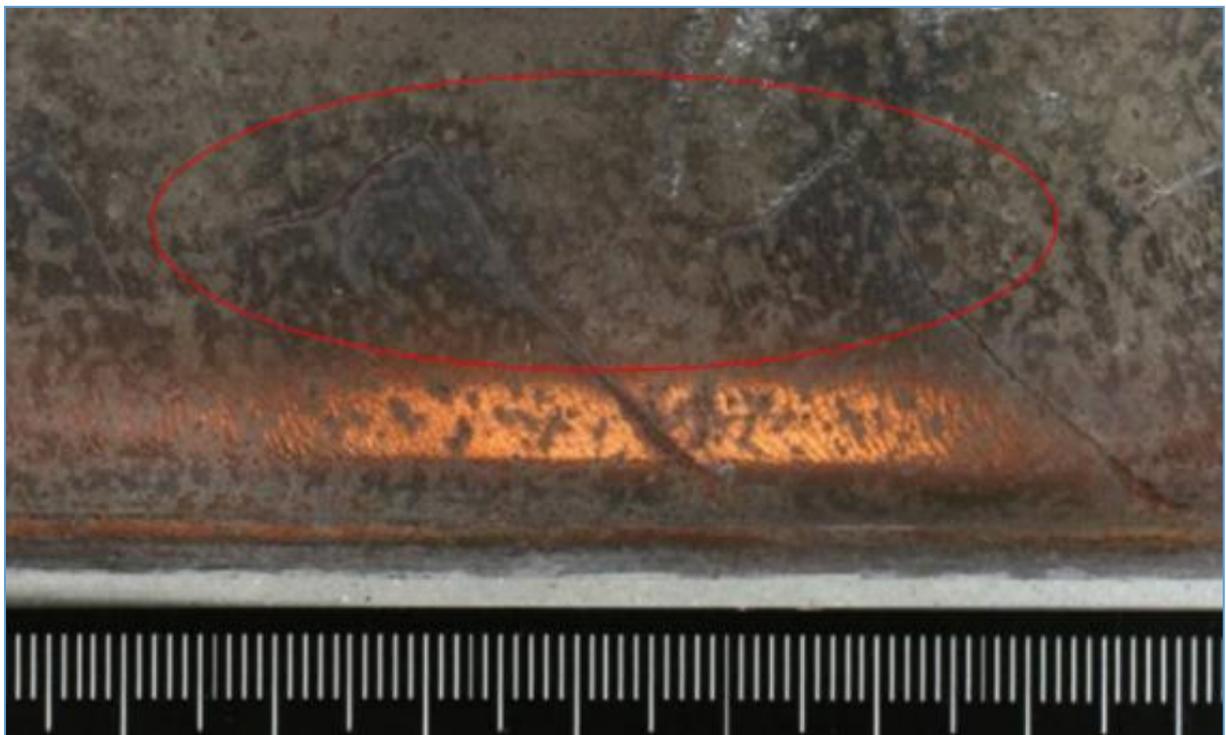
Figure 41: A GCC trailing into a Squat in a 370LHT rail. From a shallow curve in the Netherlands. The Squat is identifiable by the v- shaped cracks that are forming [76]. The GCC cracks are ~30mm apart, which is different from the cracks found in head hardened rail in Japan (Figure 42).

Figure 42: GCC in head hardened rail [84].

GCC appear in parallel cracks on the gauge corner, usually in tighter curves and are due to static contact between the flange and the gauge corner. If allowed to grow unmanaged they can develop transversely and break the rail, which is what caused the derailment in 2000 at Hatfield, UK. They are controlled using grinding, lubrication of curves of <1200m radii, using the correct rail grade (usually harder than standard) and using a wheel profile that provides stress relief to the gauge corner [85]. Despite their similarities, GCC and Squats are different, shown by the inability to use the same correction methods for GCC on Squats (e.g. harder rails) as well as the difference in their location on the rail.

2.5.4.5 The issue with defects

The introduction of a defect usually involves an irregularity in the smooth running surface along the rail. This roughening of the surface increases dynamic loading on the already damaged rail and a rough surface increases subsurface stress [10]. Defects can either break a rail by growing down into the rail until it weakens its integrity. It can also cause derailment through material loss from the rail, resulting in the wheelset no longer being constrained to the track, leading to wheel climb and derailment. This severe consequence, from defects that start as microcracks, is the reason that defects need to be understood, monitored and ideally eliminated though maintenance strategies and technological advances.

Even within the few defects shown above, confusion can occur. The correct identification of a defect is usually based on how they initiate, propagate and fail as well as how they look. For

countries to "compare notes" on the statistics of the occurrence of these defects, a common and agreed upon standard for defect identification and terminology needs to be strongly established.

2.6 Squat type defects

Squats and Squat Type Unidentified Defects (Studs) are rail running surface defects that are difficult to separate due to their similarities. This section looks at these defects as a group to establish the differences between them, focusing on cause and features.

In the literature there is not always a distinction between the two defects. Firstly, this is partly due to the developing understanding of what is a fairly new problem (compared to corrugation for example). Secondly, countries have different experiences due to a different infrastructure and rolling stock. For this reason, some defects within the literature will be referred to as Squats, but there is a possibility that it is either a Squat or Stud. Although the samples examined in this work are mainly Studs, a background on the knowledge of Squats is covered. This is because historically Studs have been identified as Squats and possibly vice-versa, some of the established background on Squats may be relevant to Studs.

Squat type defects are a significant contributor to RCF damage on railways [86]. They cost Network Rail in the UK ~£4m in repair costs as well as reduced network performance, with thousands of new Squats appearing every year [1]. As of 2015, the number of Squats appearing per year had been on a steady decline since a high of over 10,000 new Squats in 2012. This could be related to the amount of research into Squats since around this time, based on the publication dates of Squat related papers. Squats appear to develop at 'hotspots'. Australian railways are ~80% Squat free, but they are still considered a big issue [87].

Although from the UK, the author attempts to consider the experience of Australian, Dutch, French, Japanese and UK railways regarding Squats, in order to try to establish common perspectives. Different countries sometimes have different opinions, which is expected as the use of railways for freight and passengers varies from country to country.

2.6.1 The cause of Squats

In 1950s Japan, there were defects named dark cracks that displayed a dark circle as part of the defect. These have since been called rail surface shelling [84] or black spots [88], [89]. This was to distinguish them from gauge corner cracking (GCC), although GCC and Squats are considered the same defect in Japan [84]. These 'shells' are what the UK call Squats and France call Tache Noir (black/ dark spot) although sometimes in Japan, surface shelling and Squats are mentioned as though being different [90]. Figure 43 shows an example of a Squat as a visual introduction, but the features of the defect will be discussed later in 2.6.1.2.

Squats were unknown in the UK pre 1972 and first became a problem in the UK when rails began to last longer due to construction improvements, leading to "heavily worked" routes showing Squats, which reduced their lifespan from their expected 25 years [51]. Squats were first identified by British Rail on the West Coast Main Line and were often mistaken for Tache Ovales [86], but once the cleanliness of steel improved they were noted as being a new, although long standing defect.
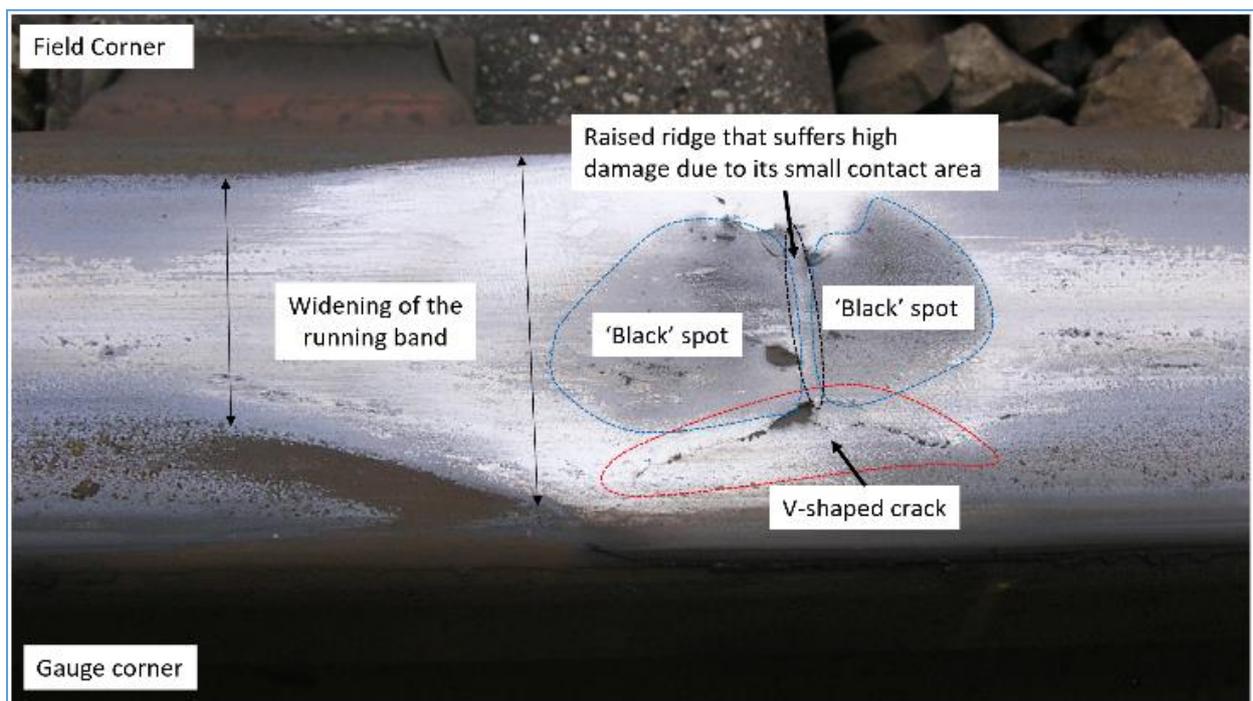
Figure 43: Typical features of a Squat. Original image courtesy of British Steel.

The exact cause of Squats is a key concept and has been a focus of Squat research. It is difficult to determine as the evidence of initiation often wears away by the time a Squat becomes detectable. A Squat in very early development would provide more clues as it may still hold clues about the initiation. Defects that may look like Squats, but are not, need to be identified and eliminated from the big data, as they can create anomalous results if they have a different cause.

2.6.1.1 Early stages

Identifying a Squat early is still a challenge because there is a lack of knowledge about the early stages of Squat initiation and growth [68]. Eddy current trains can detect small cracks in the rail surface but it is not currently able to identify a Squat from another defect. Squats take approximately 40MGTs of traffic to "seed" and 100MGTs of traffic to become detectable by ultrasound [82], which typically takes around 20 years to accumulate. Recent research has focussed on non-destructive automated methods of detecting Squats, providing a repeatable process [88], [91], [92].

Very early stage Squats are believed to stem from dents in rail surfaces, which provide an irregularity that under resonance from passing trains leads to waves of deformation forming on the surface, further increasing dynamic loading [51], [93]. The dents themselves are probably caused by a wheel picking up a harder material and indenting the rail surface, especially in the case of periodic Squats [86]. Therefore a dent could be considered a very early stage Squat. RCF due to dents have been coupled together in previous research too [51]; the cross sections of the RCF defect looked exactly like a Squat although the term RCF was used instead. Dents are considered to be a vertical irregularity along with corrugations and welds, which are also known to influence the presence of Squats, with 10-17% of Squats appearing on welds [93].

The Australian report on Squats R3-105 [94] refers to a "baby Squat" that is very early stage surface cracking but with no sign of a dent. This surface crack initiation is attributed to high traction forces: particularly lateral ones that are experienced in curves. So perhaps these slightly different initiations need to be sub-categorised to distinguish between them. As one appears to be more related to material exhaustion of the surface layers ductility, whereas the other is overloading due to dynamic forces. This difference in initiation could also contribute to the probability of a subsurface initiation versus that of surface initiation. An internal presentation

by Tata Steel showed EBSD results that suggested that an incipient Squat most likely initiated subsurface as the surface crack appeared above the deepest region of deformation [95]. White Etching Layer (WEL) is a factor for consideration regarding surface initiation mechanisms.

2.6.1.2 Links to WEL

Early stage Squats can share similar features to wheelburn due to the presence of WEL, making Squats harder to identify. Squats have developed from wheel burns [96] as seen in Figure 44. So perhaps a wheel burn could also be considered a precursor to a Squat developing? Alternatively, perhaps this is a developing Stud instead? A review of Squat research suggests a continuum from Squats, through Studs and into wheelburns and finds thinner WEL with lower carbon contents, so low carbon bainitic steels could be a solution [48].
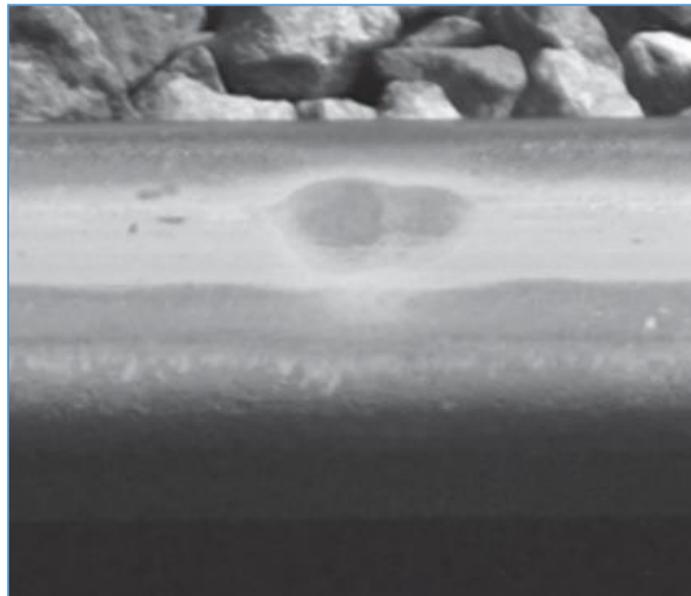
Figure 44: A Squat developed from a wheelburn [96].

WEL provides easy initiation points for cracks where the WEL meets the parent material [47]. If the crack grows into the parent rail, then even after the WEL shells the initiation is still present and able to grow further. WEL is covered in detail in 2.5.3, this section deals with the relationship between Squat defects and WEL.

Even where there is no WEL in a Squat sample, there has been WEL fragments in the Squat cracks [94], suggesting that the WEL has a role in the initiation of the Squat even if the WEL

does not last. A major distinction between head checks/ GCC and Squats is that Squat initiation is only seen where a WEL is found. Some papers show Squats with no WEL but the images of the railhead show that it has been ground recently, most likely removing the evident WEL.

A harder rail wears slower and so WEL has longer to form before wearing away. WEL can form mechanically through the severe plastic deformation of the pearlite which breaks down into a solution which forms martensite upon recrystallisation [72]. This can be referred to as dissolution of cementite. It can also form thermally due to high temperatures in the WRI such as in the formation of Studs [30]. In summary, although harder rails are generally better for RCF resistance, the initiation of Squats through WEL build-up means that harder rails suffer from Squats more than softer rail due to the crack propagation being faster through brittle structures. [81].

Grinding of the rails was introduced in the 90s and is an effective way of controlling RCF and corrugation [37], [44], [97]. However grinding leaves WEL/ friction-induced martensite (FIM) on the surface of a freshly ground rail [74], [80], [98], as well as a rough surface finish, especially in the longitudinal direction (Figure 45). This creates more stress in the rail from a reduced contact patch and folded over WEL becoming embedded in the rail surface. Embedded WEL increases residual stress and provides crack initiation sites. Initiations sites that develop into a defect can appear periodically, roughly the same periodicity as the distance between deep grinding marks [80], [94]. This strong connection between deep grinding marks has resulted in a different notation for Squats that arise due to grinding: grinding induced Squats (GIS) [74].
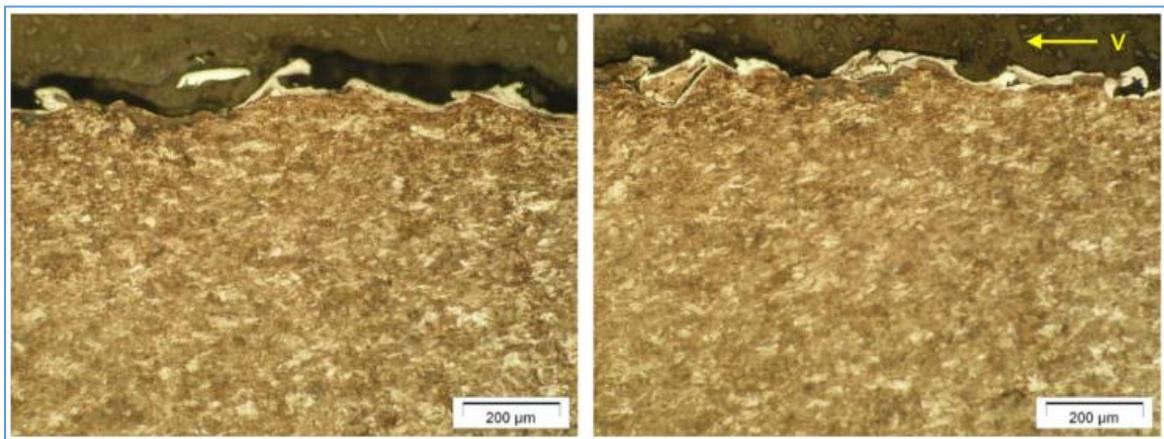
Figure 45: WEL immediately after grinding on a R260Mn standard rail. Images are of the longitudinal orientation [80].

2.6.1.3 Wheel-rail profile matching

When considering Squat initiations, the contact bands that different wheel profiles leave on a rail can lead to incipient Squat cracks appearing on the interface between two bands. Different longitudinal bands on the surface of the rail (Figure 46) experience different surface tribological transformations, so they have slightly different material responses. The borders of these bands then provide an ideal location for a crack to initiate [71]. Profile matching is discussed broadly in 2.5.2.2.
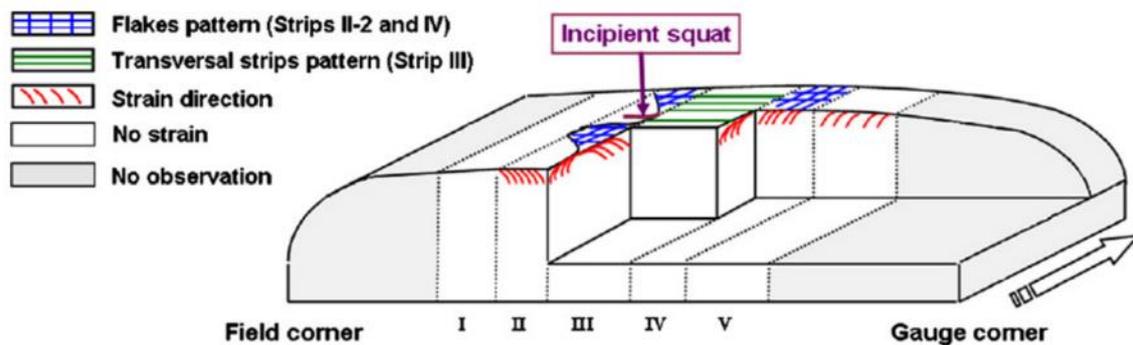
Figure 46: Five bands across the rail that have experienced different contact conditions, leading to a difference in material damage characteristics in each band, that cracks can initiate on the boundary of [71].

2.6.1.4 Track parameters

As mentioned in section 2.2.2, the tractive forces present in a curve causes peak stresses to appear closer to the surface, whereas a well lubricated surface would experience peak stresses deeper into the rail. The tractive forces experienced in a tangent/ straight track would be different to those experienced as a curve, especially when differentiating between lateral and longitudinal forces. This is pertinent to the observation that Squats can appear on the gauge corner in light curves (600m<R<1000m) or on the running band in straighter track (>1000m radius) [94].

High rails can suffer more from RCF due to higher lateral forces as the train is guided into the curve, whereas low rails suffer more from wear and plastic deformation as a larger portion of the weight of the train sits on the low rail. Therefore, Squat occurrence is almost twice as likely

on a high rail as on a low rail and the appearance of a Squat is 40-45% when there is a Squat on the other rail, compared to an 11% chance of Squats appearing on a random rail [87].

The effect of track curvature in Squat development is illustrated by the Australian distinction between running surface Squats and gauge corner Squats, the two being from different curvatures [87]. Perhaps these two Squat types always need distinguishing between, as well as if they are high or low rail, to aid in clarifying what conditions lead to what Squat type defect.

A finite element model produced in Australia [99] predicted Squats appearing on high rails in curvatures of less than 600m radii due to high lateral forces and on softer foundations. One possible reason for expecting Squats on tight curves is that the model does not consider "rail type". If "rail type" means rail grade then it needs to be considered, as the wear rate will be different for harder rails (see 2.5.2.3 for clarification). Squats being more likely in tight curves contradicts the European experience. The same research also showed that younger rails were almost four times more likely to develop Squats than older rails. Perhaps these younger rails were replacements for rails that previously had Squats and the track support damage that had developed during the previous Squatting led to rapid development of new Squats. Alternatively, the younger rails may have been harder than their predecessors, as rail grade was not part of the research. A harder rail would resist wear more effectively, but crack truncation would suffer. The softer foundation makes sense, as more rail bending will be possible.

Rail bending is influenced by sleeper separation, so sleepers play a role in the location of Squats alongside track curvature. The train passing over the track leads to cycles of compressive and tensile stresses in the rail. Rail-bending influences developing cracks once they leave the Hertzian contact zone, where the wheels influence is dominant, so adequate support is important. This has led to investigation into the support of the rail as a controlling factor in Squat growth. Research into the location of Squats in relation to sleepers and found that 74% of Squats occur on top of sleepers [89], [93]. This is linked to the tensile forces the surface of the supported rail experiences as the train wheels pass due to the unsupported rail bending down slightly (Figure 47).

There is also the possibility of a "pumping sleeper", which can exacerbate crack growth. A "pumping sleeper" is an unsupported sleeper that acts like a weight driving the oscillation of the rail rather than supporting the rail. The sleeper separation effectively doubles if the

substructure is not supporting a sleeper. The introduction of slab track where, instead of sleepers a full concrete bed continuously supports the rail, should effectively demonstrate the effect of track support.
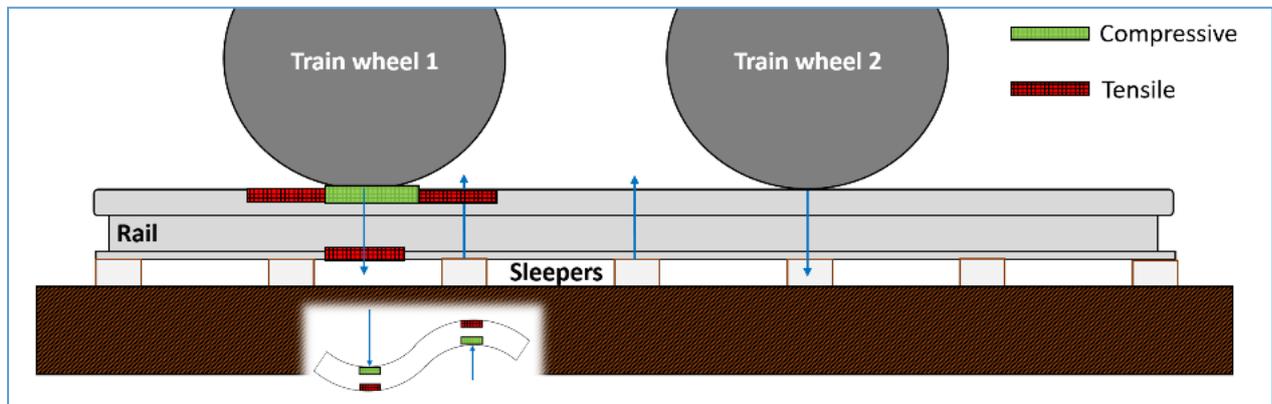
Figure 47: Compressive and tensile forces experienced by a rail during a train pass. Inset: An exaggerated bending of a rail to illustrate why the different forces are applied based on the downward force of the rail against the upward force of the supporting sleeper.

2.6.1.5 Head hardened rails

Heat-treated/ head hardened rails have a non-uniform hardness distribution in the rail due to the heat treatment and the ratcheting behaviour of the material. This non-uniform hardness could provide a preferable location for Squats to form [81]. If the wear rate is not high enough, cracks propagate too quickly for wear to contain them. Japanese railways implemented harder rails to try to combat Squats, but it was ineffective [48]. Australian railways have noticed that a head hardened rail develops Squats faster than a standard carbon rail and French railways have started trialling softer rails to combat Squats [94]. This is where the "magic wear rate" mentioned in section 2.5.2.3 becomes important. Harder rails generally wear slower, so if cracks are an issue then a softer grade should be considered. In contrast, if a rail is wearing too fast, but there is no issue with crack growth, a harder rail could be better.

2.6.1.6 Damp conditions

Damp conditions (repeated wet and dry cycles) are a component in the production of a Squat [26]. They lead to lubrication of the crack face, possible pressurisation of the crack [28] and this supports the observation that Squats are not found in tunnels unless the tunnel is leaky.

Furthermore, Australia appears to suffer fewer rail breaks from Squats due to its warmer and drier weather [94]. An alternative theory for the lack of Squats in tunnels is that the general climate and the tribology of the rail the tunnel is more constant, so the chances of starting wheel slip is reduced [94].

A successful production of Squats in a twin disc machine demonstrated the powerful effect that wet and dry cycles have on Squat initiation and growth, with no cracks appearing without the introduction of wet conditions [27]. The presence of lubricants in a crack change the friction and therefore the crack tip stress, and possibly even the crack mode as discussed in 2.5.1.4. As mentioned in section 2.5.2.1, mild wear followed by severe wear leads to heavy damage of the contacting surfaces. Wet cycles producing mild wear, followed by dry cycles that cause severe wear match this observation.

Lubricants on the running surface affect traction and traction is directly linked to Squat development [27], [100], though there is not always a strong link between Squat locations and high traction sites (accelerating and braking) such as at signals and station. This suggests that although a factor, high traction is only one of the conditions that need to be fulfilled for a Squat to develop.

Experience in Sydney, Australia suggests that tracks near trackside lubricators have a similar number of Squats to unlubricated track [94]. Perhaps the lubricators are effectively lubricating the track evenly between lubricators, preventing initiations. Alternatively, maybe lubrication that is viscous enough has minimal influence on the initiation of Squats, unlike water. This is shown in the reduced chances of viscous lubricant oils entering cracks compared to low viscosity water [58], [59]. Water is often encompassed by the term "lubricator". Given that water influences Squats but viscous lubricators do not seem to, perhaps water should be named as such rather than as a lubricant. Therefore reducing the confusion and possibility the misinterpretation of what lubricants role is in the development of Squats.

In summary, attributing factors to Squat initiation are:

- Dynamic loading due to a vertical irregularities such as dents, welds and corrugations,
- High friction/ adhesion events such as braking, accelerating and high slip/ creep,
- Damp conditions, or wet and dry cycles,
- Light curves where lateral forces increase but wear is not sufficient to truncate cracks.

Many factors have been mentioned and they are easier to view in the compiled flowchart in Figure 48.
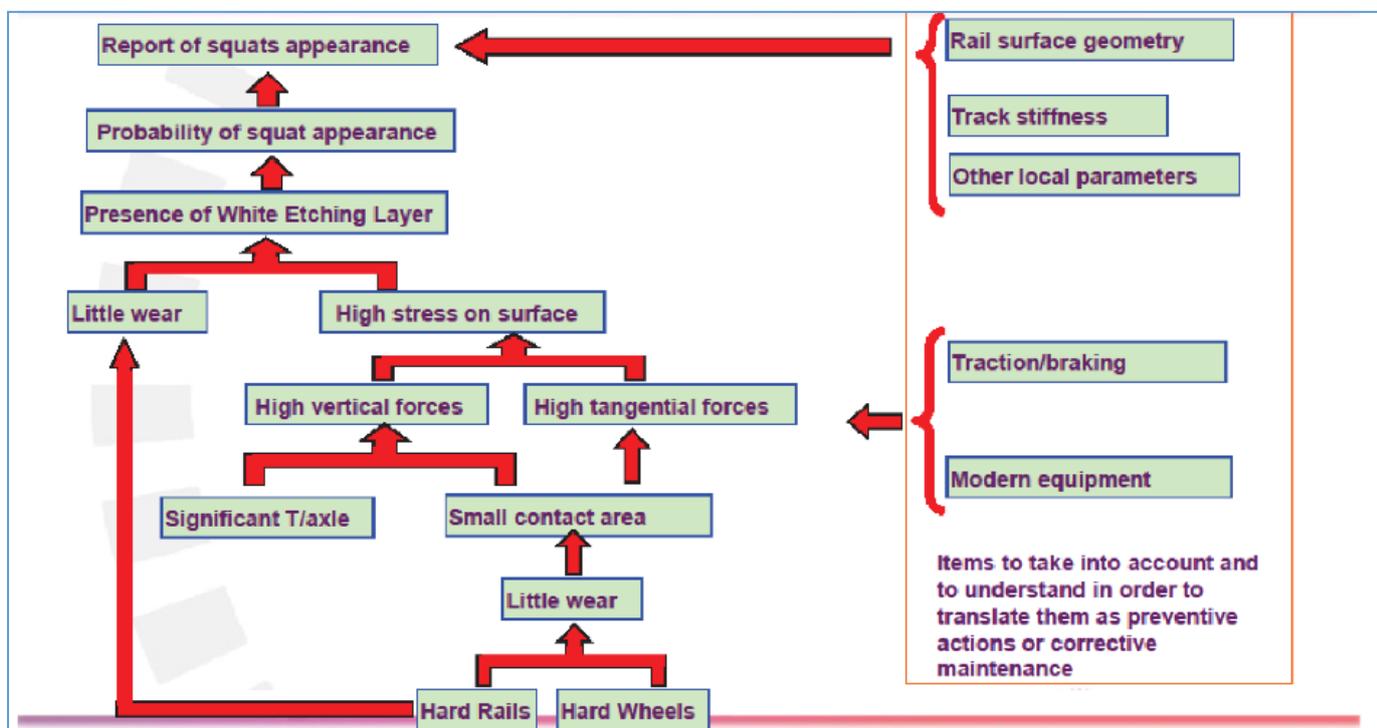
Figure 48: Flowchart of factors that influence Squat occurrences [94]**.**

2.6.2   Features

The typical features of Squats are a result of their growth behaviour. The Japanese names rail shelling and black spots (mentioned in 2.6.1) highlight two key features of a Squat. Firstly, it has a tendency to shell the rail head when it develops to a critical stage. Secondly, dark black spots on the railhead, within the plastically deformed region, can identify a mature Squat. A Squat is obvious once the dark spots and crack networks develop, but as mentioned in 2.5.4.2, black spots can indicate shelling as well as Squats. Two black spots very close together is more

likely to be a Squat, especially if there is widening of the running band and a surface crack near them (Figure 43). Sectioning the rail to reveal the crack paths can clarify this. A non-destructive technique (NDT) is also possible by CT scanning the initiation site [101], [102]. This allows a 3D inspection of the crack structure without making irreversible cuts into the defect.

Once a Squat has initiated, it grows two cracks at acute angles to the surface. Firstly as a leading crack generated by RCF [76] that follows the direction of travel. Then a trailing crack develops in the opposite direction due to the leading crack weakening the structure. The growth of these cracks are illustrated in Figure 49, which shows a schematic of the cracks on the surface, a photo of the surface and a longitudinal cross section of the defect. Figure 50 shows a finite element model that demonstrates how the forces experienced at the leading crack tip are higher than those at the trailing crack tip. These cracks can grow 25-50 mm deep into the railhead [103]. The two cracks allow the detached surface material to flow and deform resulting in a widening of the running band and dark spots over where the subsurface cracks are. This is due to a drop in the material that allows corrosion/ contaminants to build up and not be cleaned away by wheel contact. This drop in the running surface causes vertical excitation of the wheelset, just like dipped welds, misaligned joints and rail corrugation. Squats result in an increase in the noise created by wheels passing over them and Kerr et al. showed that the noise increases as the Squat develops [87]. This vertical excitation leads to further damage to occur, to both the rail and the vehicle. [45], [63]. The dip in in the track also results in a redistribution of the contact pressure to just the edges of the contact as shown in Figure 51.
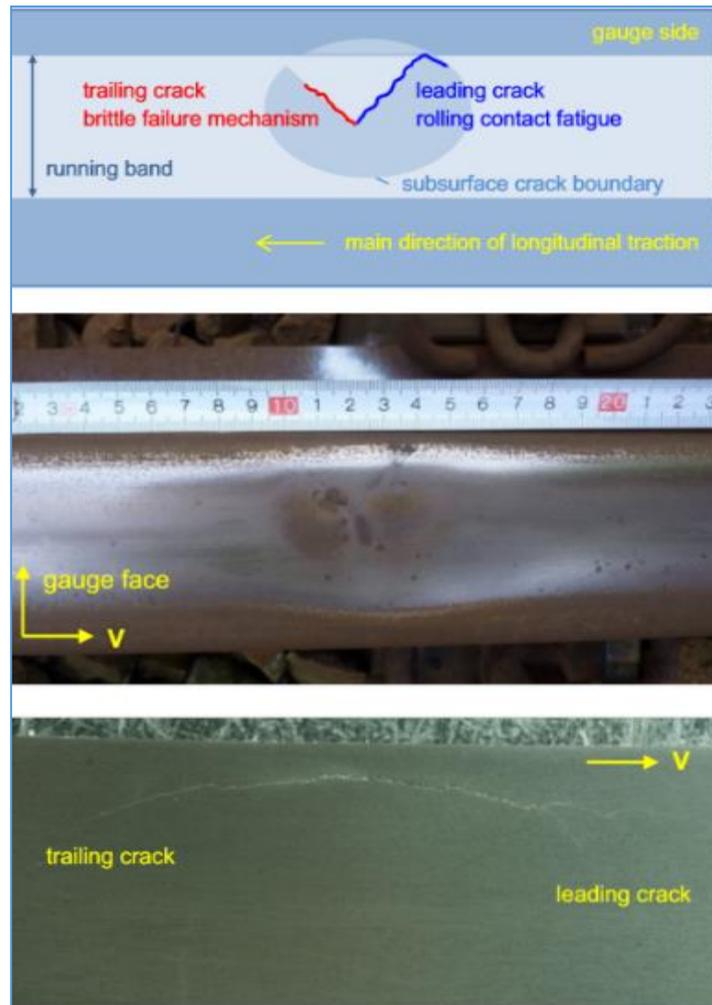
Figure 49: Views of a Squat from a schematic, an aerial view and a longitudinal cross section [76].
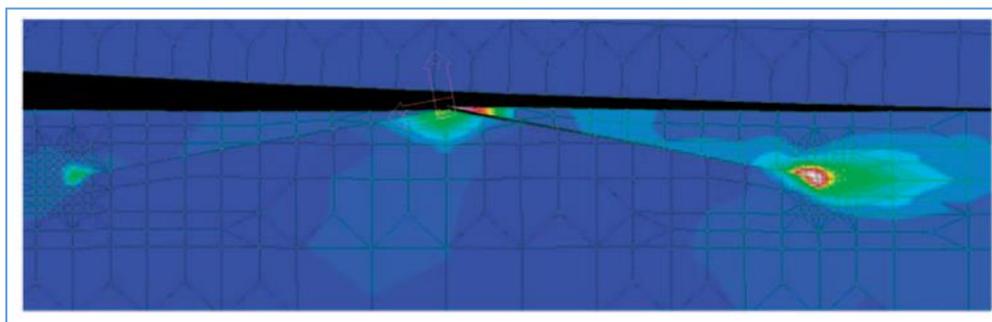
Figure 50: FE model of plastic strain in leading (right) and trailing (left) crack tips [7].
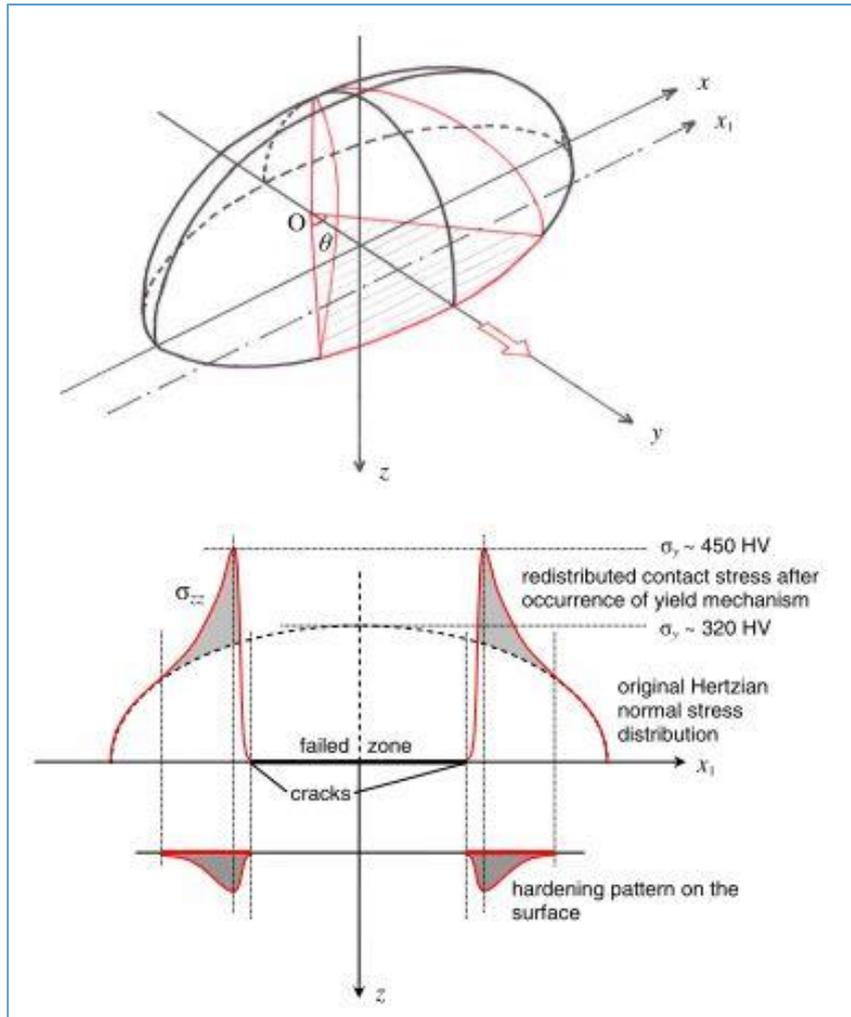
Figure 51: Schematic showing the original Hertzian normal stress distribution and the redistributed contact stress after the dip appears in the rail due to Squat growth. The new distribution results in a hardening of the surface [104].

The v-shaped crack seen in Figure 49 is also a tell-tale sign of a Squat, although when it is less developed this crack can be less obvious. This V-shape is the result of the leading and trailing cracks that have been mentioned previously. Squats can appear as periodic defects for reasons covered in 2.5. These periodic Squats will eventually link up and shell [46], [87]. The V-shape can be a tell-tale sign when trying to determine if a shelled region of a rail was due to a Squat or other reason. Figure 52 shows what appears to be a shell, but the remnants of the V-cracks for periodic Squats are visible, showing that it was a Squat before it shelled.

The slip damage associated with wheelburns is also associated with Squats, but with less damage (0.03mm of thermal damage depth for Squats, 4-6mm for wheelburns) [45]. This is

due to the WEL present in both. WEL can often be seen around a Squat defect as shiny patches. Eventually a Squat can develop enough that it can develop a transverse defect (TD), which means that one of the growing cracks grows downwards into the railhead until it compromises the railhead integrity, causing the rail to break [29], [82], [87].
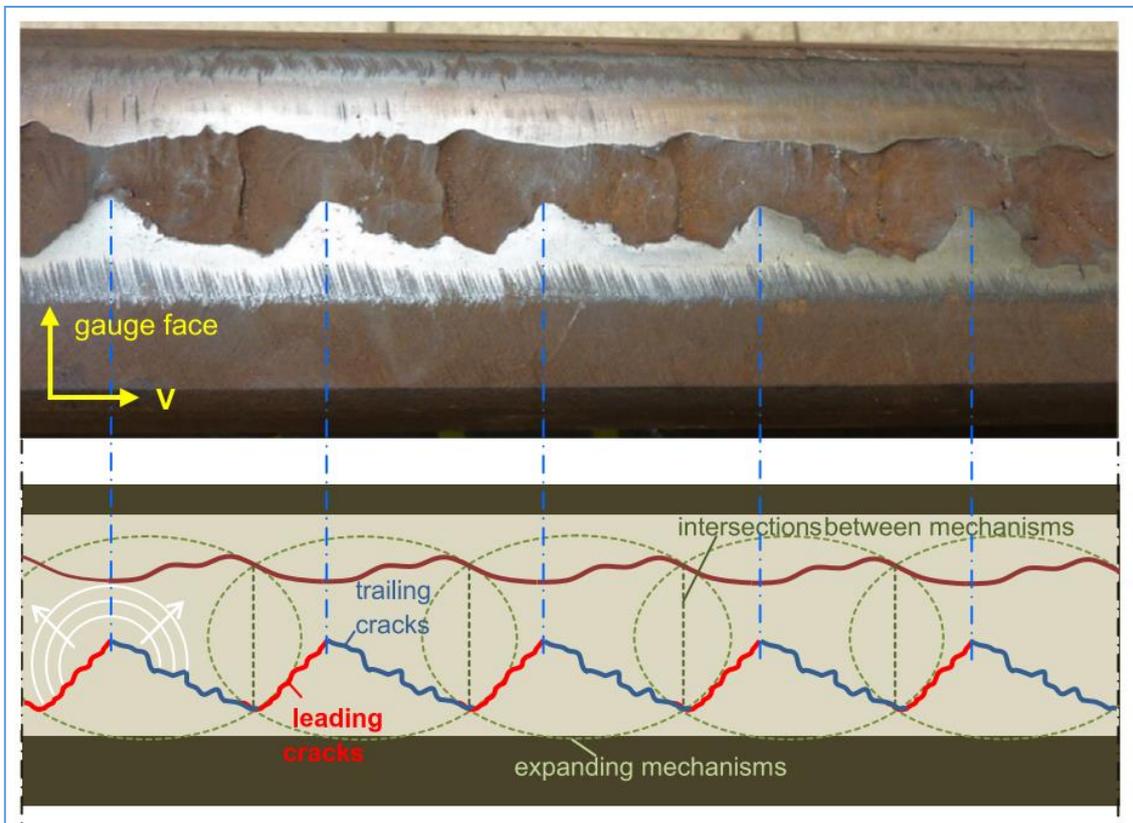
Figure 52: Periodic Squats that have fully developed and shelled. Although this looks like a case of severe shelling, the V-shape cracks made up of leading and trailing cracks make these Squats [8].

2.6.3   Studs (Squat Type Unidentified Defects)

Studs are a recently named defect [30], distinguishing between RCF initiated Squats (classical Squats) and thermally initiated Squats (Studs). Studs are still a relatively new research area compared to the research conducted on Squats. They share a very similar surface features (Figure 53) and subsurface crack morphology to Squats, but they have some key differences as tabulated in the literature [30]:

- They do not require RCF to develop, so there is comparatively little plastic deformation on the surface

- They develop much faster than Squats (detectable at ~10MGTs rather than ~40MGTs)

- WEL exists in all locations across a Stud and the grain boundaries can be made out within the WEL

- Squats that arise from wheelburns were probably Studs [30], [96]

- Initiates in the running band rather than the gauge corner

- Cracks wander through grains, they do not necessarily follow the soft ferrite around grain boundaries

- No evidence of fluid entrapment as a growth mechanism
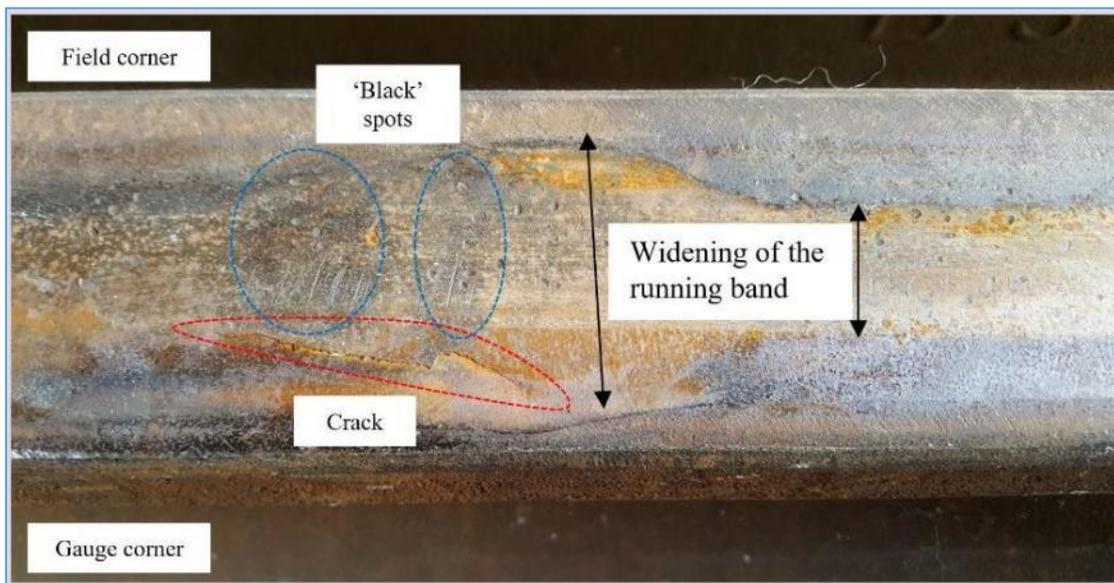
Figure 53: A Stud from a metro line with the widened band, large crack on the gauge corner (bottom of image) and a smaller crack on the field side. There is not much obvious surface plastic flow and the rail has little wear.

2.6.3.1 Causes

Studs are believed to be caused by thermal damage to the rail surface [30]. This thermal damage was proposed due to the WEL noticeable along the surface of the rail and the lack of plastic deformation/ ratchetting that usually accompanies Squats. If the rail temperature reaches the austenisation temperature (~720°C) there is a change in the lattice volume for pearlite transforming to austenite, then transforming into martensite on subsequent cooling, leading to

to residual stress within the material following severe thermal input [105]. Studs have been linked to the use of modern AC traction control systems often used on metro rolling stock [106] operating where the first Studs (by name, not necessarily first ever) were found [30]. Without traction control, wear is typically higher [94] because traction control systems limit slip so that gross slip does not occur. Gross slip of a wheel would generate a WEL and would lead to a wheel burn defect, which can look like a Stud, but would not have the Squat-like crack structure beneath. An example of the damage caused by slip events that heat the rail is seen in Figure 54.

Factors that cause Squats could still be relevant to Studs, as it may be that Studs that have been studied in the past have been named Squats. In some Australian literature Studs are referred to as thermal Squats as a distinction, although more recent papers started using the Stud term as a clearer distinction. If a Squat and a Stud were taken from the same environment (traffic type, track layout, climate etc.), they could be compared to determine the differences in their initiation.
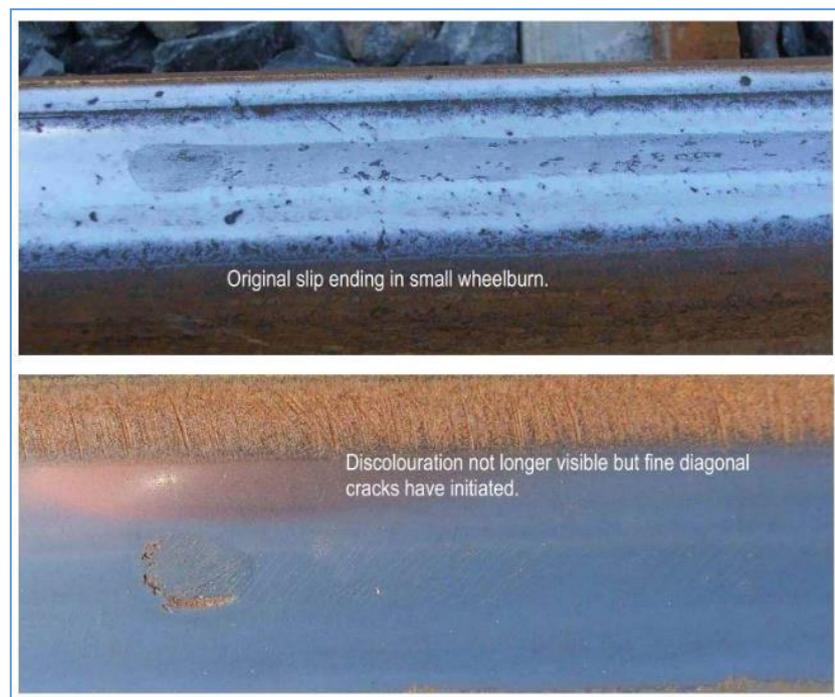
Figure 54: Damage/ WEL caused by slip that resulted in cracking of the surface in the future. Above) Just after the event, depicting the contact band for the wheel. Below) 11 months after the event [107].

Looking at the surface in Figure 53, Studs also have a widening of the band just like Squats. The surface is smooth in both directions due to the lack of ratchetting [30], as opposed to rough in one direction. They can have surface breaking cracks just like Squats and the subsurface crack structure is very similar to a Squat as it branches down in an inverted V-shape. There are noticeable black spots on the surface due to the subsurface cracks, but without the plastic flow. High contact stresses on the thermally damaged surface area leads to cracks on the surface, either on both the gauge corner and field corner sides of the defect or just on one side, depending on the contact patch [94].

Both Squats and Studs develop a saddle shaped subsurface crack network as shown in Figure 55. The change in the crack growth changes as the contact patch changes during the crack development, with the crack tip reaching the edge of the deformed microstructure at 2-4mm deep [37], [104].
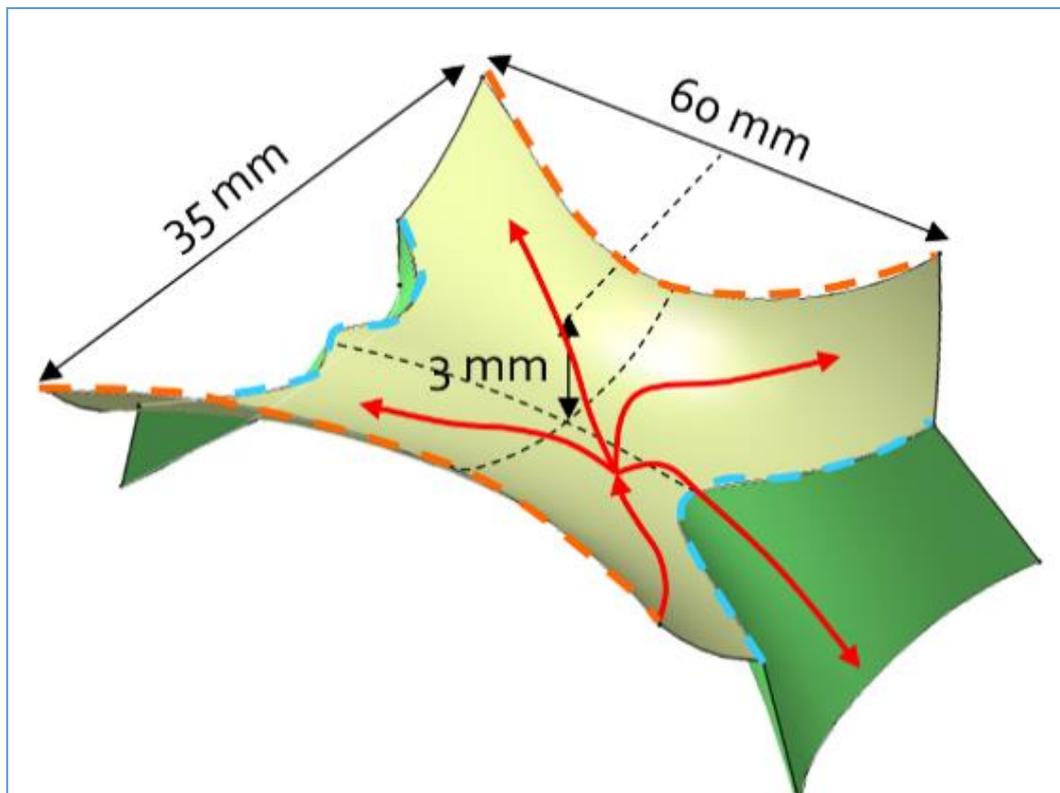
Figure 55: A computer drawn representation of the Squats crack planes including arrows to show their propagation directions [68].

The literature does not show evidence of Studs that have broken a rail due to a transverse defect (TD), although the presence of a Stud interacting with GCC has been noted to result in a TD [83]. Again, if a Stud that broke a rail was misidentified as a Squat, this may not be the case. Especially considering the surface damage/ shelling that would occur during an in-service rail break: making surface plastic flow hard to observe.

Being thermally initiated and linked to wheelburns makes Studs similar in mechanism, if not in degree, to engine burns that are experienced in the US and Australia [35], [108], [109]. In Australia, engine burns are used as an alternative term for wheel burns. Multiple engine burns can be found close together and they are initiated by wheel slip that causes thermal damage. This thermal damage leads to cracks forming between the thermally damaged region and the parent rail, which can become a TD as shown in Figure 56, also known as an Engine Burn Fracture [109]. Engine burns closely resemble Studs. Being found in the US and Australia, where railways are used more for heavy freight, perhaps engine burns are almost the same as Studs but are separated by the traffic type. Engine burns being related to heavy axle loads (freight) whereas Studs are related to lighter axle loads (passenger).

Figure 56: Engine burn fracture [109].

2.6.4   Possible solutions

2.6.4.1 Grinding

Squats are currently only managed through grinding [78], [110], which is an introduction of artificial wear on the rails to increase the natural wear in order to truncate any cracks that are growing. The ideal grinding regime can be determined by applying the principle of the 'magic wear rate' [37]. However, Squats have been linked to grinding in Australia, Denmark [87] and the Netherlands [111]. These Squats could be considered to be grinding induced Squats (GIS). Therefore, the quality of the finish when grinding is an important factor. Rail milling uses carbide teeth to remove flakes of steel rather than grinding stones, resulting in a smoother finish but at a higher cost.

2.6.4.2 Profile matching

Profile matching is much easier said than done. If a profile conformed as closely as possible, high peak stresses on localised regions could be avoided. Choosing the right rail and wheel profiles would achieve this, however, freight trains may need different wheel profiles to passenger trains as they navigate curves differently. Having a rail profile that closely matches all wheel profiles experienced on that line would be ideal, if possible. Profile matching is also improved through good rail and wheel surface management, i.e. smooth surfaces.

2.6.4.3 Softer rail

Softer rail would increase the wear rate to stay ahead of the cracks but at the expense of more regular rail replacements. However, wear is easier to monitor than crack growth.

Harder rail is capable of resisting initiation of cracks for longer but once the crack initiates it becomes a race against time to replace the rail before the crack grows to a critical length.

The ideal solution could be a hard surface to resist initiation and a softer subsurface to increase wear once the crack grows into the rail. This would be a heat-treated grade, but these grades have been criticised in the literature (see section 2.4.2.2 for details). This suggests that there is more to the solution than just the hardness of the rail. Microstructure could provide a crack initiation and propagation retardation as well as ideal wear rates. This is one of the reasons why

bainitic rail steels are being explored as a better option, although bainitic steels are not completely resistant to Squats.

2.6.4.4 Friction management

A CoF of 0.3 or above is detrimental to the rail because it raises the peak stress to the surface, making plastic deformation more probable. Therefore, controlling the WRI friction, without compromising the trains ability to change speed, could reduce the chances of fatigue-induced initiations. Controlling the friction can also reduce wheel slip or spin events, minimising thermal damage. Traction gels can improve traction without providing fluids that may exacerbate existing RCF. Effective lubrication application in curves reduces high lateral forces that are associated with defects such as Squats.

2.6.5   Squat defects summary

Being the main subject of this work and having such a large body of information presented, Squats are summarised here before being condensed further in the final literature summary.

The term Squat type defects covers a group of defects; Squats, Studs and GIS. There is also a category of Squats seen in mainland Europe called Belgrospi [112], though this category is not discussed here. Wheel burns are linked to Squats and have been suggested to be part of a continuum with Squats and Studs. If this is so then perhaps engine burn fractures are also in that continuum. Shells share features with Squats and Studs, such as black spots and the eventual loss of material. It is unclear if a shell that contains apex cracks, such as in Figure 52, should still be called Squats. If not then when does a Squat become a shell. If it is when it shells, what does a partially shelled Squat get classified as? For big data analysis to effectively yield clues as to where and why Squat type defects appear, these categories would need to be clear and agreed upon. Different causes would require different grouping.

The depth of deformation of a rail depends on the loading conditions and history. Mixed traffic complicates the matter further. The issue is further compounded by the different wheel profiles used, on changing rail profiles and the depth to which a crack grows through the weakened deformed layer of a rail can vary. A railhead may have multiple contact bands from different wheel profiles and the layout of the contact bands can vary. Each band can be slightly different

microstructures due to different amounts of deformation, created by the different profile matches. This creates subtle boundaries between these contact bands, between which Squats may initiate.

Traction control is key to reducing the occurrence of various problems including RCF and corrugation. Preventing high slip events keeps thermal and mechanical damage minimised. However, traction control systems have been linked to Stud development, so these systems need upgrading, coupling with the appropriate TOR product, or both. Tunnels might suffer less Squats because the traction is more consistent due to less water being present. A comparison between issues in older tunnels that are leaky and damp with modern dryer tunnels could also shed some light on this concept.

Damp conditions are a significant factor in crack growth, shown by twin disc tests. Keeping a rail dry at all times is not plausible, especially in the UK. A softer rail would allow the higher wear rate to compete with the accelerated crack growth rate caused by pressurisation. The introduction of softer rails (R200) in France should illustrate how effective this could be. Grinding or milling to a sufficient depth would have the same effect but provide more control at a cost.

Grinding needs to be to a high quality finish as GIS have been re-designated from Squats due to their strong connection with grinding marks and the periodicity of grinding. Some of the samples examined in this work are also GIS. Perhaps having another grinding pass with smoother stones along Squat hotspots will aid in reducing their reoccurrence. If rail replacement is needed due to defects becoming too numerous and/or deep, the supporting structure below the Squatted rails would also need repair. This is because the increased dynamic loading caused by Squats can disrupt the supporting structures. If they are left disrupted, the reduced track support can contribute to the Squats reappearing.

If Squats are appearing as rails get harder and rails get harder to combat other defects, then controlling Squats will become a balancing act, as it is between wear and RCF. Understanding the cause of the Squats and designating sub-categories if needed is key to controlling their occurrence. Direct evidence of initiation is difficult to obtain as the rail wears away, so perhaps the key to understanding the causes is the comparison of different Squat type defects to find common factors and variables. The variables may be what creates the sub-categories. To do

this to full effect, requires as much track information to be gathered as possible for identified defects.

2.7 Track Maintenance

Rail running surface defects are a problem because they cause vertical irregularities that increase dynamic forces and reduce the contact with the rail to a much smaller contact, increasing the probability of plastic flow or/ and crack initiations. Enough defects on a rail can cause a train to derail, either through rail climb, where the wheel flange is not held by the rail, or by a break of the rail itself. There are also residual tensile stresses in the rail head from manufacturing which can pull open defects as cracks start to initiate [21], [63]. Because cracks initiate and propagate slowly at the start and then propagation speed increases as the crack becomes longer [56], detecting and managing a defect early is important for effective track maintenance and safety.

Defects are monitored by industry using employees on track walks, ultrasound detectors for cracks (both train mounted and hand held), train mounted eddy current devices for surface defects and aerial video footage of the network. Aerial images can show white patches in the ballast where crushing has occurred due to high dynamic forces caused by an issue in the area. Industry corrects defects using maintenance and corrective grinding, rail milling, weld repairs and rail replacement. Rail replacement is undesirable as it is expensive, time consuming and interrupts services. This chapter will briefly discuss the different approaches currently employed in the maintenance of rails. The inclusion of these techniques gives an idea of where improvements in the detection and management of Squats can be made.

2.7.1   Non-destructive Inspection Techniques

Ultrasonic inspection is used in the final stages of the manufacturing process to inspect the rail head, web and foot for defects. This same technology can be used to monitor rails in service to search for defects that develop after manufacturing. The ultrasonic probes are mounted within a wheel on either a train or a manually pushed trolley. The ultrasonic signal bounces back when it reaches an interface, such as the air gap within a crack, and is detected by transducers. The probes are orientated at a variety of angles in order to maximise the probability of detecting a variety of defects that can have different orientations.

The location and depth of a crack can be determined using this technique, allowing monitoring of a growing defect. A defect can then be removed from the track before the crack grows to a critical length. 3D ultrasound mapping can be used as a NDT method to map the development of a Squats crack plane [88]. In Squats, sometimes a transverse crack develops rather than the rail shelling, though the reason for this is unclear. Transverse cracks are important to detect as they grow downwards and can break the rail. These cracks are difficult to detect by ultrasound because a shallow crack can shield them (Figure 57). This is especially the case in Squats as the leading and trailing cracks grow first, then the transverse crack grows from the leading crack.

Ultrasound cannot penetrate through this crack.....

....so it will not detect a deeper crack here
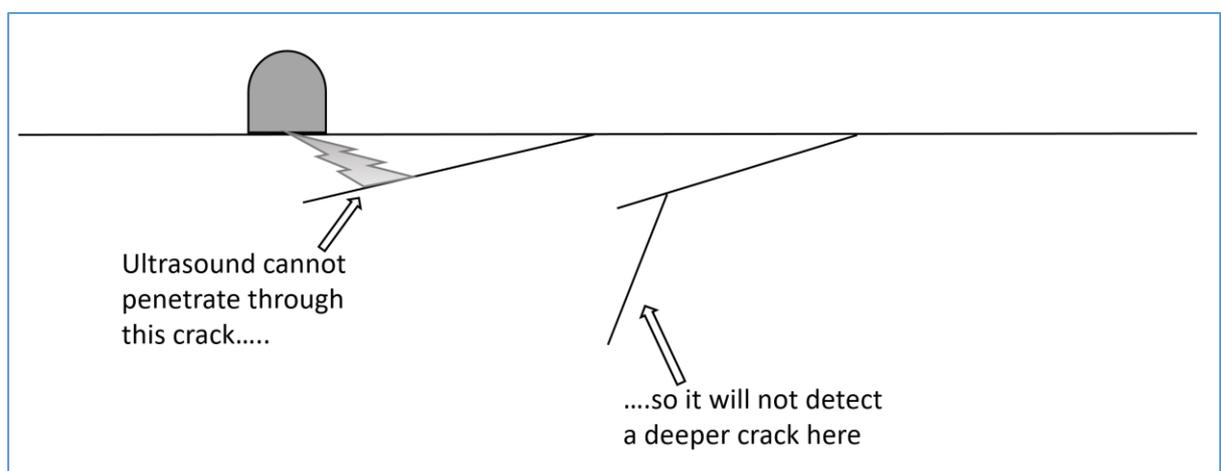
Figure 57: A crack shielding a deeper crack from ultrasound detection. Figure adapted from [63].

Ultrasound is effective for inspecting the bulk of the rail for defects, but for detection of flaws very close to the running surface, such as RCF cracking, eddy current testing is used. Eddy current testing trains use the change in the conductivity and permeability of the rail material due to defects. There are other methods for detecting Squats that have shown some success such as visual inspection using video recognition [113], thermography [114], axle box acceleration measurements [115] and rail magnetisation [91].

2.7.2 Grinding and Milling

After the Hatfield derailment in the UK, grinding was re-introduced as part of the strategy to combat RCF. Implemented in 2003/4, grinding provided an increase in rail life from 4 years to

10 years [116]. Grinding was first introduced to remove corrugation, which plagued rails in the 1960s and 1970s [37]. Grinding involves the use of large grinding stones spinning as a group at various angles to remove the upper layer of the rail. This allows the removal of the damaged upper layer of a rail as well as the re-profiling of the rail back to a standard. If a rail is not re-profiled, it can become deformed (Figure 20) and provide very poor profile matching.

High speed grinding was developed in Germany and approved for trials in 2006, using unpowered grinding stones to artificially wear away the surface, removing short wavelength corrugation and RCF. High speed grinding can remove up to 0.05 mm of material per pass compared to standard grinding removal of 1.5mm per pass [117].

Maintenance grinding to remove a small depth regularly is more economical and effective than using corrective grinding after a defect has developed. However, there are concerns that grinding can exacerbate RCF development due to changes in the contact area and WEL being folded into the parent material [80], [81]. In December 2017, Network Rail announced the procurement of three new grinding trains [118], so the surface finish quality of the grinding may change. For an excellent surface finish, milling is an alternative to grinding.

Milling uses rotating wheels fitted with carbide teeth to dig into the rail, and then shear flakes of steel from the rail. Squats need up to 8mm of material removal to completely eliminate them, as any remnant of the Squat crack network will result in it returning at the same site.

2.7.3   Track repairs

For the UK, aluminothermic welding was proposed as a repair method for a severe Squat and involved removing a 10mm depth of rail material around the defect and welding in replacement material [119]. Weld repair consistency is operator dependant and can introduce defects into the rail. The mechanical properties of the weld needs to be as close as possible to the original rail. Automated weld repairs are more consistent but the technology is still in development. If a rail develops defects that are too numerous or deep, full or partial rail replacement becomes a final option to return the rail to a defect free and reliable state. The expense is not just the cost of the new rail, but the closure of the track includes fines from train operators and there is the labour cost. There are enormous automated rail replacement vehicles that can replace large

stretches of track, but a small section of replacement probably does not warrant the logistics of getting such a machine involved. After a rail has been replaced, tamping is needed.

Tamping is the agitation of the ballast under the track to resettle it into its original position. This is needed because the high dynamic forces caused by the defects affects the ballast causing it to break down and shift, reducing track support. Unfortunately the act of tamping also degrades the ballast [32], turning the rough interlocking stones into smoother stones that will move past each other. Therefore, the ballast eventually needs replacing too. Severe ballast degradation can be seen easily from an aerial view using a UAV. It shows as a white area around the rails where white dust has been created from the crushing of the ballast by high dynamic forces.

The Hatfield derailment highlighted the need for constant monitoring of the integrity of the rails, which extends to the supporting structures. Extensive rail network monitoring uses testing trains fitted with ultrasonic, eddy current and visual detection systems and more techniques are being explored to add to the maintenance tool belt. Surface defects are removed through grinding, milling, weld repair or rail replacement. In some cases, defects are so severe that the track substructure requires repair too.

2.8 Summary of entire literature review

2.8.1　Wheel-Rail Interface

When a train sits and moves on a track, all forces that influence the movement of the train have to pass through the very small interface, the wheel-rail interface (WRI). The complex group of factors determine how this interaction between train and track are understood through contact mechanics and tribology. The wheel requires traction, via friction to accelerate and decelerate, operating on a contact patch that varies depending on the profiles and integrity of the wheel and the rail.

The conditions of this contact patch in terms of traction and forces acting are influenced by: track curvature, where lateral forces become larger in tight curves; loading that varies on the high and low rails due to cant and the shift in position of the conical wheels; third bodies such as lubricants, traction enhancers, and contaminants such as leaves.

During speed and direction change, this patch often contains regions of slip and stick simultaneously, shifting the balance of traction and slip. When slip dominates, the wheels can rotate faster than the trains relative speed to the track. This is somewhat analogous to a car wheels spinning when the clutch is released too quickly. As a car leaves damage on the road in the form of rubber transfer, the wheel and rail can both suffer damage during slip. This damage can be both thermal and mechanical in nature.

The conical wheels shifting the wheels position in curving combined with different profile matches mean that this damage does not always occur in the exact same place. Especially in curves where the profile is related to the load of the train, which also influence the severity of the damage. Damage is more common where slip is more likely, where trains start and stop, mainly at signals and stations. The depth at which damage occurs depends on the friction and roughness of the wheel-rail surfaces.

2.8.2　RCF

The damage referred to is usually grouped within the category of rolling contact fatigue (RCF). The type of RCF seen depends on the factors mentioned above, so some defects are more

common in tight curves than gentle, high rails than low rails, heavy haul lines rather than metro etc. Slip of the wheels can be controlled through traction control systems and friction management, such as top of rail modifiers that change the friction in the WRI, reducing the damage in advance. After the damage occurs, wear is a naturally occurring mechanism that can be artificially enhanced (grinding) to remove the damaged regions. Grinding is employed as a major maintenance strategy, with high speed grinding now being introduced to meet the modern demands of a railway system. This artificial wear needs to be of a good quality, as a rough finish can change the damage behaviour. Milling is proving to be an effective tool to this end. Though it comes with a higher initial cost, if it retards the reinitiating of defects, the cost may be worthwhile.

Wear can be an issue in itself though and the need for rails to last longer, coupled with the need to resist crack initiations, led to the more common use of harder rails. The standard grade for rails even increased from 220HV to 260HV. However, this is not a complete solution because hard rails wear less, so crack growth can become the cause of rail failure. The use of harder rails, sometimes called premium rails, can be very beneficial to extending a tracks life. However they must be used appropriately, where wear is a bigger problem than fatigue, otherwise the magic wear rate is harder to achieve. Wear is easier to monitor than crack growth.

The surface hardness of an installed rail will not remain the same though. Phase changes due to incremental mechanical damage (ratchetting) and thermal inputs (wheel slip and slide) create very hard white etching layers (WEL), which are prone to cracking and spalling, roughening the surface. WEL has been connected to Squat type defect development.

2.8.3   Squat type defects

Squats have been an issue for decades now, but have become highlighted because they are now the biggest cause of early rail replacement in the UK and other countries. Understanding why Squats initiate and why some pose a safety risk, as transverse defects, would lead to an increase in rail life, providing significant savings and safer railways.

The field of Squats has evolved recently though, with two new classifications, perhaps subcategories of Squats being documented. Squat Type unidentified Defects (Studs) and Grinding Induced Squats (GIS) have been given different designations, due to their more

dominant initiation aspect being thermal or their initiation being linked to grinding marks respectively. Hence, the grouped term used here of Squat type defects.

Various factors in the cause of Squats have been identified, though the magnitude of each factor is not clear, and may vary depending on the location. WELs, compromised track support, irregularities in the continuity of the running band (dynamic loading), surface roughness, wheel slips, wheel slides (cause 'flat tyres' that hammer the rail) and even traction control systems probably all factor in different magnitudes. There are also other factors such as increasing traffic loads and frequency, new train types (suspensions) and constantly changing tracks through innovation, degradation and replacement. These era relevant factors make information validity dependant on its age. Observing networks that use limited or only one type of rolling stock are one way of reducing the variables to try to hone in on the strongest influencing factors.

Squats are tackled with different approaches. A computational modelling perspective is a powerful tool that allows the inner development of the defect to be simulated in a way it could not be monitored using traditional track observations. Track observations document a defects growth and its track parameters before removing the defect for inspection. Well-established metallurgical methods are used to understand the material response behind the defects growth. Big data is making use of all of the information gathered by industry to understand what influences a defects appearance.

A computational modelling approach is limited by the known data that can be used as an input, as the exact cause of Squats is not clear. Metallurgical inspection is limited to observing one state of the defects growth and limited cuts can be made to observe the substructure. Big data relies on the track parameters and defect types being classified correctly for the patterns to emerge. This is an issue if the newer categories of Studs and GIS are more common than initially thought, because this will make data from before their recognition deceptive.

The combination of all three methods should help to clear the big picture. The metallurgical evidence should be able to verify the input data in the computational models, which should provide information to clarify the big data analysis. There are other cross overs but the point is that many methods need to be employed to understand Squat type defects. Even new methods such as the recent use of computed tomography X-ray scanning to map the initiation sites of Squat defects.

2.8.4   Gaps in research

The main unknowns within Squat research are the causes for initiation, particularly the significance of the factors, the reasons why some defects develop transversely whilst others shell and how defects vary from each other depending on their track parameters and traffic history. There is also developing work into the detection of Squats, especially with the aim of detecting them automatically with train mounted equipment, but that is outside the scope of this work.

All three of the main unknowns are tackled within this work along with an extension to the CT scanning approach. Defects will be compared to each other to try to identify reasons behind any variations as well as establish the presence of any of the newer categories of Squat type defect. If any transverse defects are found, their cause/ origin will be explored. Although it is difficult due to the 'missing evidence' associated with a constantly wearing system, the cause of the defects investigated will also be addressed.

2.9 References

[1]   Network   Rail,   "Rail   Head   Squats,"   2015.   [Online].   Available: https://cdn.networkrail.co.uk/wp-content/uploads/2017/03/Challenge-Statement-Track-Plain-line-rail-head-Squats.pdf. [Accessed: 13-Jun-2018].

[2]   K. L. Johnson, *Contact Mechanics*. Cambridge University Press, 1985.

[3]   H. Wu and N. Wilson, "Flange climb derailments: causes and prevention," *Int. Railw. J.*, vol. 52, no. 3, p. 47, 2012.

[4]   D. I. Fletcher and J. Beynon, "Development of a Machine for Closely Controlled Rolling Contact Fatigue and Wear Testing.," *J. Test. Eval.*, vol. 28, no. 4, p. 267, 2000.

[5]   R. S. Dwyer-Joyce, C. Yao, and R. Lewis, "A Comparison of Three Methods for the Measurement of Wheel Rail Contact," *STLE/ASME 2008 Int. Jt. Tribol. Conf.*, no. November 2015, pp. 541–543, 2008.

[6]   J. J. Kalker, *Three-dimensional elastic bodies in rolling contact*. Kluwer Academic Publishers, 1990.

[7]   J. J. Kalker, "A fast algorithm for the simplified theory of rolling contact," *Veh. Syst. Dyn.*, vol. 11, no. 1, 1982.

[8]   J. Srivastava, P. Sarkar, and V. Ranjan, "A numerical study on effects of friction-induced thermal load for rail under varied wheel slip conditions," *Simul. Trans. ofthe Soc. Model. Simul. Int.*, pp. 1–12, July 2018.

[9]   T. Telliskivi and U. Olofsson, "Contact mechanics analysis of measured wheel-rail profiles using the finite element method," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 215, no. 2, pp. 65–72, 2001.

[10]   A. Kapoor, F. J. Franklin, S. K. Wong, and M. Ishida, "Surface roughness and plastic flow in rail wheel contact," vol. 253, pp. 257–264, 2002.

[11] F. J. Franklin, I. Widiyarta, and A. Kapoor, "Computer simulation of wear and rolling contact fatigue," *Wear*, vol. 250-251, pp. 949–955, 2001.

[12] O. Arias-Cuevas, Z. Li, R. Lewis, and E. A. Gallardo-Hernández, "Rolling-sliding laboratory tests of friction modifiers in dry and wet wheel-rail contacts," *Wear*, vol. 268, no. 3–4, pp. 543–551, 2010.

[13] J. H. Al-Bedhany and H. Long, "Microscopic investigation of subsurface initiated damage of wind turbine gearbox bearings," *J. Phys. Conf. Ser.*, vol. 1106, no. 1, p.12029, 2018.

[14] E. Magel and J. Kalousek, "Martensite and Contact Fatigue Initiated Wheel Defects," in *12th International Wheelset Congress*, 1998, pp. 100–111.

[15] J. A. Williams, "The influence of repeated loading , residual stresses and shakedown on the behaviour of tribological contacts," vol. 38, pp. 786–797, 2005.

[16] E. Magel, "Rolling Contact Fatigue: A Comprehensive Review," *ECS Trans.*, vol. 28, no. 32, pp. 35–44, 2010.

[17] S. R. Lewis, R. Lewis, and D. I. Fletcher, "Assessment of laser cladding as an option for repairing/enhancing rails," *Wear*, vol. 330–331, no. July, pp. 581–591, 2015.

[18] C. Bernsteiner, G. Muller, A. Meierhofer, K. Six, D. Kunstner, and P. Dietmaier, "Development of white etching layers on rails: simulations and experiments," *Wear*, vol. 366–367, pp. 116–122, 2016.

[19] F. . Carter, "On the Action of a Locomotive Driving Wheel," *Proc. R. Soc.*, vol. A112, pp. 151–157, 1926.

[20] J. J. Kalker, "Wheel — rail rolling contact theory," vol. 144, pp. 243–261, 1991.

[21] L. Smith, "Rolling Contact Fatigue in Wheel-Rail Contact," University of Newcastle Upon Tyne, 2007.

[22] R. Lewis and U. Olofsson, "Chapter 5: Tribology of the Wheel–Rail Contact," in *Handbook of Railway Vehicle Dynamics*, S. Iwnicki, M. Spiryagin, C. Cole, and T. McSweeney, Eds. Taylor & Francis Group, LLC, 2006, p. 131.

[23] R. Stock, L. Stanlake, C. Hardwick, M. Yu, D. Eadie, and R. Lewis, "Material concepts for top of rail friction management – Classification, characterisation and application," *Wear*, vol. 366–367, pp. 225–232, 2016.

[24] R. Stock, D. T. Eadie, D. Elvidge, and K. Oldknow, "Influencing rolling contact fatigue through top of rail friction modifier application - A full scale wheel-rail test rig Study," *Wear*, vol. 271, no. 1–2, pp. 134–142, 2011.

[25] J. Williams, *Rolling contacts and rolling-element bearings*. Cambridge University Press.

[26] D. I. Fletcher and J. Beynon, "The effect of intermittent lubrication on the fatigue life of pearlitic rail steel in rolling- sliding contact," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 214, pp. 145–158, 2000.

[27] M. Kaneta, K. Matsudo, K. Murakami, and H. Nishikawa, "1998 ASME Kaneta A Possible Mechanism.pdf," *ASME*, vol. 120, pp. 304–309, 1998.

[28] S. Bogda, P. Lewicki, and M. Szymaniak, "Experimental and theoretical investigation of the phenomenon of filling the RCF crack with liquid," vol. 258, pp. 1280–1287, 2005.

[29] N. Gao, R. S. Dwyer-Joyce, and D. G. G. Grieve, "Disc machine testing to assess the life of surface-damaged railway track," *Proc. Inst. Mech. Eng. Part F-Journal Rail Rapid Transit*, vol. 215, no. 4, pp. 261–275, 2001.

[30] S. L. Grassie, D. I. Fletcher, E. A. Gallardo Hernandez, and P. Summers, "Studs: a Squat-type defect in rails," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 226, no. 3, pp. 243–256, 2011.

[31] C. Hardwick, R. Lewis, and R. Stock, "The effects of friction management materials on rail with pre existing rcf surface damage," *Wear*, vol. 384–385, no. November 2016, pp.

50–60, 2017.

[32] M. Sol-Sánchez, F. Moreno-Navarro, and M. C. Rubio-Gámez, "Analysis of ballast tamping and stone-blowing processes on railway track behaviour: the influence of using USPs," *Géotechnique*, vol. 66, no. 6, pp. 481–489, 2016.

[33] E. T. (Ernest T. Selig and J. M. Waters, *Track geotechnology and substructure management*. Thomas Telford, 1994.

[34] J. Miss, "Running Surface Features and Defect Catalogue," pp. 1-71, 2010.

[35] NSW Transport RailCorp, "Rail Defects Handbook: TMC 226," pp. 1–83, 2012.

[36] P. Mutton, J. Cookson, C. Qiu, and D. Welsby, "Microstructural characterisation of rolling contact fatigue damage in flashbutt welds," *Wear*, vol. 366–367, pp. 368–377, 2016.

[37] E. Magel, M. Roney, J. Kalousek, and P. Sroba, "The blending of theory and practice in modern rail grinding," *Fatigue Fract. Eng. Mater. Struct.*, vol. 26, no. 10, pp. 921–929, 2003.

[38] Bhadeshia, Harshad K. D. H., et al. Steels: Microstructure and Properties, Elsevier Science & Technology, 2006.

[39] D. Aliya and A. Analytical, "Physical Metallurgy Concepts in Interpretation of Microstructures," *Metallogr. Microstruct.*, vol. 9, pp. 44–70, 2018.

[40] University of Cambridge, "Interpretation of cooling curves." [Online]. Available: https://www.doitpoms.ac.uk/tlplib/phase-diagrams/cooling.php. [Accessed: 29-Apr-2019].

[41] F. P. L. Kavishe and T. J. Baker, "Effect of prior austenite grain size and pearlite interlamellar spacing on strength and fracture toughness of a eutectoid rail steel," *Mater. Sci. Technol.*, vol. 2, no. 8, pp. 816–822, 2014.

[42] H. W. Zhang, S. Ohsaki, S. Mitao, M. Ohnuma, and K. Hono, "Microstructural investigation of white etching layer on pearlite steel rail," *Mater. Sci. Eng. A*, vol. 421, no. 1–2, pp. 191–199, 2006.

[43] G. E. Totten and G. E. Totten, "Failures Related to Heat Treating Operations," *Fail. Anal. Prev.*, vol. 11, pp. 192–223, 2018.

[44] International Union of Railways, "INNOTRACK concluding technical report," Paris, France, 2010.

[45] ARTC, "Rail Defects Handbook Some Rail Defects , their Characteristics , Causes and Control RC 2400," no. March, 2006.

[46] M. Steenbergen, "Rolling contact fatigue: Spalling versus transverse fracture of rails," *Wear*, vol. 380–381, pp. 96–105, 2017.

[47] V. Dikshit, P. Clayton, and D. Christensen, "Investigation of rolling contact fatigue in a head-hardened rail," *Wear*, vol. 144, no. 1–2, pp. 89–102, 1991.

[48] TataSteel, "Review of anti Squat research," 2013.

[49] R. Stock and R. Pippan, "Rail grade dependent damage behaviour – Characteristics and damage formation hypothesis," *Wear*, vol. 314, no. 1–2, pp. 44–50, 2014.

[50] W. Solano-Alvarez and H. K. D. H. Bhadeshia, "White-etching matter in bearing steel. Part I: Controlled cracking of 52100 steel," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 45, no. 11, pp. 4907–4915, 2014.

[51] P. Clayton and M. B. P. Allery, "Metallurgical Aspects of Surface Damage Problems in Rails," *Can. Metall. Q.*, vol. 21, no. 1, pp. 31–46, 1982.

[52] UIC, "Recommendations for the use of rail steel grades," pp. 4, 2015.

[53] E. Magel, P. Sroba, K. Sawley, and J. Kalousek, "Control of Rolling Contact Fatigue of Rails,", pp. 8, 2005.

[54] J. E. Garnham and C. L. Davis, "Very early stage rolling contact fatigue crack growth in pearlitic rail steels," *Wear*, vol. 271, no. 1–2, pp. 100–112, 2011.

[55] M. Burstow, "Experience of premium grade rail steels to resist rolling contact fatigue (RCF) on GB network," *Ironmak. Steelmak.*, vol. 40, no. 2, pp. 103–107, 2013.

[56] A. D. Hearle and K. L. Johnson, "Mode II stress intensity factors for a crack parallel to the surface of an elastic half-space subjected to a moving point load," *J. Mech. Phys. Solids*, vol. 33, no. 1, pp. 61–81, Jan. 1985.

[57] A. Kapoor, D. I. Fletcher, and F. J. Franklin, "The role of wear in enhancing rail life," *Tribol. Ser.*, no. 41, pp. 331–340, 2002.

[58] J. E. Garnham and J. Beynon, "The early detection of rolling-sliding contact fatigue cracks," *Wear*, vol. 144, no. 1–2, pp. 103–116, 1991.

[59] D. I. Fletcher, P. Hyde, and A. Kapoor, "Investigating fluid penetration of rolling contact fatigue cracks in rails using a newly developed full-scale test facility," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 221, pp. 35–44, 2007.

[60] A. J. Perez-Unzueta and J. H. Beynon, "Microstructure and wear resistance of pearlitic rail steels.pdf," *Wear*, vol. 162–164, pp. 173–82, 1993.

[61] P. Clayton, "Tribological aspects of wheel-rail contact: a review of recent experimental research," *Wear*, vol. 191, pp. 170–183, 1996.

[62] J. F. Archard and W. Hirst, "The wear of metals under unlubricated conditions," *Proc. R. Soc. London.*, vol. 236, no. Ser. A. Math. Phys. Sci., pp. 397–410, 1956.

[63] D. F. Cannon, K. O. Edel, S. L. Grassie, and K. Sawley, "Rail defects: An overview," *Fatigue Fract. Eng. Mater. Struct.*, vol. 26, no. 10, pp. 865–886, 2003.

[64] S. B. Newcomb and W. M. Stobbs, "A transmission electron microscope Study of the white etching layer on a rail head," *Mater. Sci. Eng. A*, vol. 66, pp. 195–204, 1984.

[65] J. Wu, R. H. Petrov, M. Naeimi, Z. Li, R. Dollevoet, and J. Sietsma, "Laboratory simulation of martensite formation of white etching layer in rail steel," *Int. J. Fatigue*, vol. 91, pp. 11–20, 2016.

[66] M. Steenbergen and R. Dollevoet, "On the mechanism of Squat formation on train rails – Part I : Origination," vol. 47, pp. 361–372, 2013.

[67] C. Bernsteiner, G. Muller, A. Meierhofer, K. Six, D. Kunstner, and P. Dietmaier, "Development of white etching layers on rails: simulations and experiments," *Wear*, vol. 366–367, pp. 116–122, 2016.

[68] S. Simon. De la dynamique ferroviaire `a l'accommodation microstructurale du rail : Contribution des TTS `a la r´eponse tribologique des aciers : Cas du d´efaut de squat. Mechanics of the solides. INSA de Lyon, 2014. French.

[69] A. Al-juboori *et al.*, "Squat formation and the occurrence of two distinct classes of white etching layer on the surface of rail steel," vol. 104, pp. 52–60, 2017.

[70] W. Lojkowski, M. Djahanbakhsh, G. Bürkle, S. Gierlotka, W. Zielinski, and H. J. Fecht, "Nanostructure formation on the surface of railway tracks," *Mater. Sci. Eng. A*, vol. 303, no. 1–2, pp. 197–208, 2001.

[71] S. Simon, A. Saulot, C. Dayot, X. Quost, and Y. Berthier, "Tribological characterization of rail Squat defects," *Wear*, vol. 297, no. 1–2, pp. 926–942, 2013.

[72] R. I. Carroll, "Surface Metallurgy and Rolling Contact Fatigue of Rail," University of Sheffield, 2005.

[73] F. Alwahdi, "Wear and Rolling Contact Fatigue of Ductile Materials," University of Sheffield, 2004.

[74] A. Ekberg, "D3.1- Enhanced track structure- Status, key influencing parameters and prioritised areas of improvement v.6," pp. 11-12, 2018.

[75] S. Li, J. Wu, R. H. Petrov, Z. Li, R. Dollevoet, and J. Sietsma, "'Brown etching layer':

A possible new insight into the crack initiation of rolling contact fatigue in rail steels?," *Eng. Fail. Anal.*, vol. 66, pp. 8–18, 2016.

[76]   M. Steenbergen, "Squat formation and rolling contact fatigue in curved rail track," *Eng. Fract. Mech.*, vol. 143, pp. 80–96, 2015.

[77]   S. L. Grassie, "Rail corrugation: characteristics, causes, and treatments," *Proc. IMechE Vol. 223 Part F J. Rail Rapid Transit*, pp. 581–596, 2009.

[78]   K. Kondo, K. Yoroizaka, and Y. Sato, "Cause, increase, diagnosis, countermeasures and elimination of Shinkansen shelling," *Wear*, vol. 191, no. 1–2, pp. 199–203, 1996.

[79]   F. Fau, H. Smith, S. Fretwell-Smith, L. Deng, "Effect of grinding quality, lubri- cation quality and rail hardness on flaking defect initiation on high rails" *Proceedings of the 10th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2015), Colorado, USA, 30 August–3 September.*

[80]   M. Steenbergen, "Rolling contact fatigue in relation to rail grinding," *Wear*, vol. 356–357, pp. 110–121, 2016.

[81]   A. C. Athukorala, "Investigation of Wear and Ratcheting Response of Head-Hardened Australian Rail Steel under Conditions of Rolling Contact Fatigue," Queensland University of Technology SCIENCE, 2016.

[82]   S. L. Grassie, "Squats and Squat-type defects in rails: the understanding to date," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 226, no. 3, pp. 235–242, 2012.

[83]   S. L. Grassie, "Studs and Squats: The evolving story," *Wear*, vol. 366–367, pp. 194–199, 2016.

[84]   M. Ishida, "Rolling contact fatigue (RCF) defects of rails in Japanese railways and its mitigation strategies," *Electron. J. Struct. Eng.*, vol. 13, no. 1, pp. 67–74, 2013.

[85]   S. L. Grassie, "Rolling contact fatigue on the British railway system : treatment," vol. 258, no. June 2003, pp. 1310–1318, 2005.

[86]   Railtrack plc, *Rolling contact fatigue in rails: a guide to current understanding and practice*. Heron Press, Birmingham, UK, 2001.

[87]   M. Kerr, A. Wilson, and S. Marich, "The epidemiology of Squats and related rail defects," *CORE 2008 Conf. Railw. Eng. rail- the-core Integr. Transp.*, no. September, pp. 83–96, 2008.

[88]   S. Kaewunruen, M. Ishida. In Situ Monitoring of Rail Squats in Three Dimensions Using Ultrasonic Technique. *Exp Tech* 40**,** 1179–1185 (2016).

[89]   X. Zhao, *Dynamic Wheel/Rail Rolling Contact at Singular Defects with Application on Squats*. Southwest Jiaotong University, 2012.

[90]   M. Ishida, M. Akama, K. Kashiwaya, and A. Kapoor, "The current status of theory and practice on rail integrity in Japanese railways-rolling contact fatigue and corrugations," *Fatigue Fract. Eng. Mater. Struct.*, vol. 26, no. 10, pp. 909–919, 2003.

[91]   S. Kaewunruen, "Identification and prioritization of rail Squat defects in the field using rail magnetisation technology," *Struct. Heal. Monit. Insp. Adv. Mater. Aerospace, Civ. Infrastruct. 2015*, vol. 9437, no. April 2015, p. 94371H, 2015.

[92]   S. Kaewunruen and M. Ishida, "Rail Squats: understand its causes, severity , and non-destructive evaluation techniques," *Proceedings of the 20th National Convention on Civil Engineering* Vol. 1 Iss. 1, 2015.

[93]   Z. Li, X. Zhao, C. Esveld, R. Dollevoet, and M. Molodova, "An investigation into the causes of Squats-Correlation analysis and numerical modeling," *Wear*, vol. 265, no. 9–10, pp. 1349–1355, 2008.

[94]   M. Kerr and W. Daniel, "Final Report on the Rail Squat Project R3-105," CRC for Rail Innovation, Brisbane, 2011.

[95]   L. Deng, "Understanding of rail Squats from the point of view of metallurgy, Tata Steel internal presentation." .

[96]   Z. Li, "Squats on Railway Rails," in *Wheel-Rail Interface Handbook*, R. Lewis and U. Olofsson, Eds. Woodhead Publishing Limited, 2009.

[97]   E. Magel, P. Mutton, A. Ekberg, and A. Kapoor, "Rolling contact fatigue, wear and broken rail derailments," *Wear*, vol. 366–367, pp. 249–257, 2016.

[98]   C. J. Rasmussen, X. Zhang, H. K. Danielsen, and S. Fæster, "Grinding induced martensite on the surface of rails," *Proc. 35th Risø Int. Symp. Mater. Sci. New Front. Nanometals*, vol. 35, pp. 439–446, 2014.

[99]   M. Farjoo, W. Daniel, P. Bellette, and P. A. Meehan, "Field statistical and finite element analysis of rail Squats," *Eng. Fract. Mech.*, vol. 109, pp. 117–129, 2013.

[100] Z. Li, R. Dollevoet, M. Molodova, and X. Zhao, "Squat growth — Some observations and the validation of numerical predictions," *Wear*, vol. 271, no. 1–2, pp. 148–157, 2011.

[101] M. Naeimi *et al.*, "Reconstruction of the rolling contact fatigue cracks in rails using X-ray computed tomography," *NDT E Int.*, vol. 92, no. April, pp. 199–212, 2017.

[102] C. Jessop, J. Ahlström, L. Hammar, S. Fæster, and H. K. Danielsen, "3D characterization of rolling contact fatigue crack networks," *Wear*, vol. 366–367, pp. 392–400, 2016.

[103] N. Gao, R. S. Dwyer-Joyce, and D. G. G. Grieve, "Disc machine testing to assess the life of surface-damaged railway track," *Proc. Inst. Mech. Eng. Part F-Journal Rail Rapid Transit*, vol. 215, no. 4, pp. 261–275, 2001.

[104] M. Steenbergen and R. Dollevoet, "On the mechanism of Squat formation on train rails – Part II : Growth," vol. 47, pp. 373–381, 2013.

[105] D. I. Fletcher and S. H. Sanusi, "the Potential for Supressing Rail Defect Growth Through Tailoring Rail Thermo-Mechanical Properties," in *10th International Conference on Contact Mechanics and Wear of Rail/ Wheel System*, 2015.

[106] D. Scott, D. I. Fletcher, and B. J. Cardwell, "Simulation Study of Thermally Initiated

Rail Defects," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 228, no. 2, pp. 113–127, 2012.

[107] M. Kerr and A. Wilson, "Rail Squat Defects," *Perm. W. Inst. Inc*, pp. 13–25, 2000.

[108] Sperry Rail Service, "Rail Defect Manual," p. 43, 1968.

[109] Federal Railroad Administration, "Track Inspector Rail Defect Reference Manual," 2015.

[110] H. M. Smith, "An examination of Squat defects in a 60E1 R260 rail removed from the Bussy Saint Georges test site in the Paris suburbs (RER line)," Rotherham, UK, 2011.

[111] M. Itabashi and K. Kawata, "Carbon content effect on high-strain-rate tensile properties for carbon steels," *Int. J. Impact Eng.*, vol. 24, no. 2, pp. 117–131, 2000.

[112] Z. Popovic Radovic, V., "Analysis of cracking on running surface of rails," *Gradevinar*, vol. 65, no. 3, pp. 251–259, 2013.

[113] T. Szugs *et al.*, "Combination of Ultrasonic and Eddy Current Testing with Imaging for Characterization of Rolling Contact Fatigue," *19th World Conference on Non-Destructive Testing (WCNDT 2016), 13-17 June 2016 in Munich, Germany (WCNDT)*, 2016.

[114] D. Peng and R. Jones, "Lock-in thermographic inspection of Squats on rail steel head," *Infrared Phys. Technol.*, vol. 57, pp. 89–95, 2013.

[115] M. Molodova, Z. Li, A. Nunez, and R. Dollevoet, "Automatic Detection of Squats in Railway Infrastructure," *IEEE Traansactions Intell. Transp. Syst.*, vol. 15, no. 1, pp. 1980–1990, 2014.

[116] Network Rail, "Re-Profiling Rail to Remove Defects & Extend Rail Life." [Online]. Available: https://www.networkrail.co.uk/wp-content/uploads/2019/06/Challenge-statement-track-re-profiling-rail-to-remove-defects-and-extend-rail-life.pdf. [Accessed: 19-Aug-2019].

[117] Taubert, M. "High-speed grinding." *International Railway Journal, 45*(12), p.40-41, 2005.

[118] Network Rail, "Innovative, new grinding trains will help passenger and freight trains run smoothly and safely for years to come." [Online]. Available: https://www.networkrailmediacentre.co.uk/news/innovative-new-grinding-trains-will-help-passenger-and-freight-trains-run-smoothly-and-safely-for-years-to-come. [Accessed: 09-Aug-2019].

[119] A. Faley, "Repair of a 'Squat' defect by thermit welding," no. 919, pp. 95–97, 1979.

Chapter 3 : Methods and techniques

This chapter describes all of the methods and techniques used for this work into one location. In later chapters, when techniques are referred to, they will only be mentioned and then cross-referenced with this chapter. Coverage includes the use of an infinite focus microscope for the 3D mapping of surfaces, µ-CT scanning, which is described in reasonable detail, sample preparation/ metallography, hardness mapping for material characterisation and optical and electron microscopy. The flowchart in Figure 58 shows the process that the project followed, with the techniques from this chapter employed at the relevant stages.
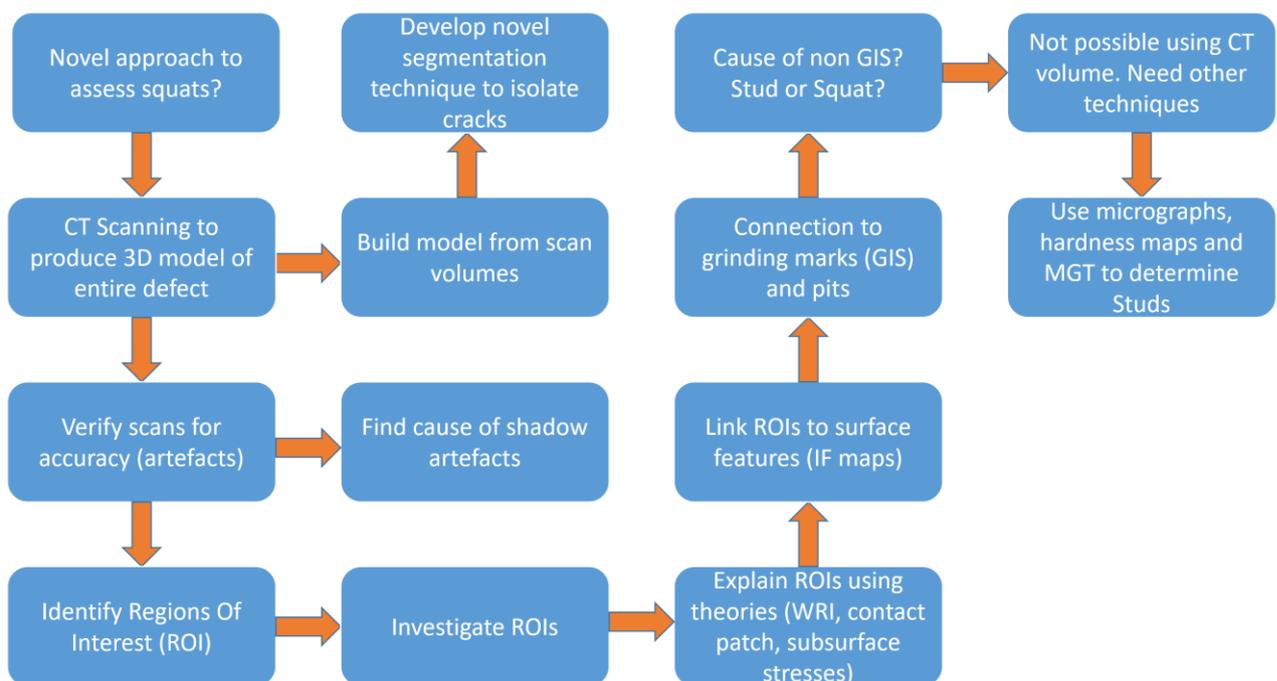
Figure 58: Flowchart showing the process of the project and where techniques where needed.

3.1 Alicona Infinite Focus SL 3D Microscope

The details of the surfaces of the samples were mapped using an Alicona Infinite Focus SL 3D Microscope, with attention focused around surface breaking cracks and vertical irregularities such as grinding marks and pits. The infinite focus (IF) microscope, coupled with IF-MeasureSuite Version 5.1 software, generated a 3D model of the surface that allowed roughness measurements to be extracted. Specifying a 3D space using two reference points that

encompass the surface allows the software to take multiple images within that space, select the focussed images and then stitch them together to create a 3D map. This solves the issues of trying to view a rough surface in its entirety on an optical microscope. The IF microscope can measure surfaces up to 87°, so some features cannot be captured and appear as white regions within the bulk image (areas of no data). Below is an example of the x, y and z coordinates and settings used to scan a sample:

Table 2: Settings and coordinates used for one of the scans using the IF microscope

Co-ordinate	Reference point 1 (mm)	Reference point 2 (mm)
X	-17.353	16.799
Y	13.905	-19.809
Z	-1.556	2
Setting	**Value**	
Lateral resolution	16 microns	
Vertical resolution	115 microns	
Estimated scan time	3-4hrs	

For this work, the co-ordinates are always a positive and negative pair because the 0,0,0 position was set as being the surface of the sample (in focus) and the centre of the region of interest. The reference points would be opposite corners of the cube (i.e the top left corner zoom in too far and the bottom right corner zoomed out too far). If the file size that the scan produced was too large, shown by a software pop up warning, the file size was reduced by eliminating scan points. This is achieved using the 'decimate' function, which was always used once, twice for relatively large areas but never three times. If an area has gridded points labelled 1,2,3,4,5 etc. decimate would leave points 1,3,5 etc. and another use of decimate would produce 1,5 etc. So decimate reduces the resolution each time. Comparison of two scans of the same area, one using decimate once, and one using decimate twice, showed no detectable quality change to the human eye, though it may have an effect on post processing if very subtle surface changes are sought.

3.2 X-ray Micro-Computed Tomography (μ-CT) Scanning

The use of X-rays to see inside an object is fairly well known. The limitation of 2D X-rays is that it only shows one plane (orthoslice) of a 3D object. Taking many X-rays from different angles and using computer software to merge them together to create a coherent 3D model provides a much more detailed volume. This is CT scanning. X-ray imaging relies on the interaction of X-rays with materials. To understand X-ray images, X-ray scattering needs to be understood. Then this Chapter covers how a CT scanner is configured before discussing the scanning itself, construction of 3D volumes and the problems associated with X-ray detection (artefacts).

3.2.1   X-ray scattering

Just as a piece of paper held in the sunlight becomes a source of scattered light, so a piece of paraffin placed in the path of a beam of X-rays becomes itself a source of scattered X-rays.

Arthur H. Compton [121]

When X-rays interact with the electrons in matter, scattering occurs, resulting in a primary beam signal and a secondary scattered beam signal. The wavelength of the scattered beam becomes larger than the primary beam, reducing intensity, as the scattering angle (angle of deflection) increases [121]. This reduction in energy/ intensity is known as the Compton Effect. This effect applies to gamma ray photons as well as X-ray photons, but this work only deals with X-rays. The scattering of an X-ray is due to a single electron (or other charged particle), so the scattering behaviour of light elements and heavy elements are distinguishable due to their difference in their electron numbers. By knowing the wavelength of the primary X-ray and detecting the change of wavelength in the secondary X-ray allows the calculation of the scattering angle (Figure 59). This is one of the main contributors in photon attenuation.

The photoelectric effect also contributes to a majority of the attenuation of an X-ray. This is where an X-ray photon ejects an electron from its atom, ionising the atom. The ionised atom returns to neutral resulting in the creation of another photon that does not contribute to or hinder the imaging process. This is the main cause of photon absorption below 500keV [122].

Figure 60 shows that for the energy range of general industrial X-ray use (grey box), the Compton and Photoelectric effects (C and PE) are the two main contributors to X-ray photon attenuation. Raleigh scattering (R) and pair production have minimal to no effect in this range of 90- 1100KeV. The sum of these mechanisms determines a materials absorption coefficient (μ).
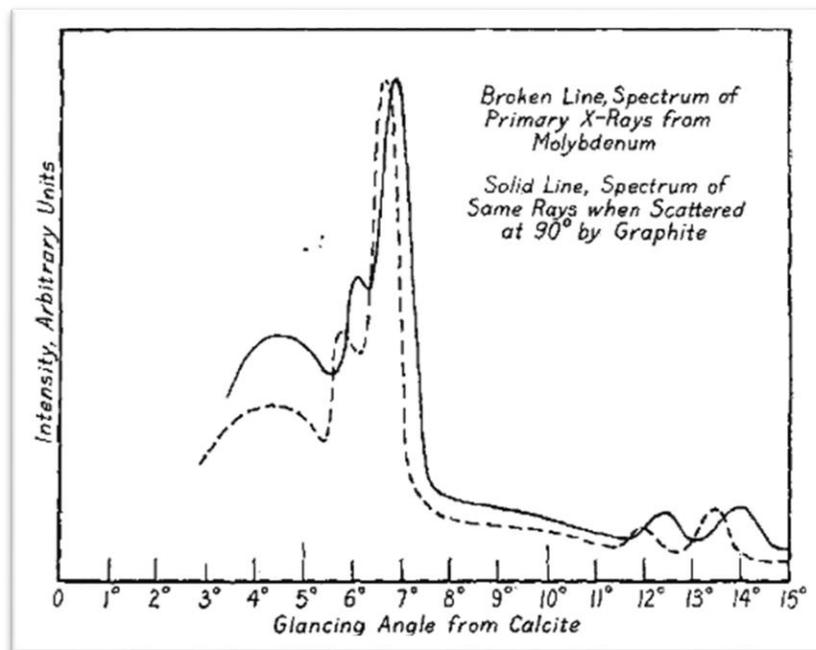
Figure 59: Glancing angles using a molybdenum target and a carbon radiator to scatter the beam [121]. *The solid line is the primary rays and the dotted line is the scattered rays*
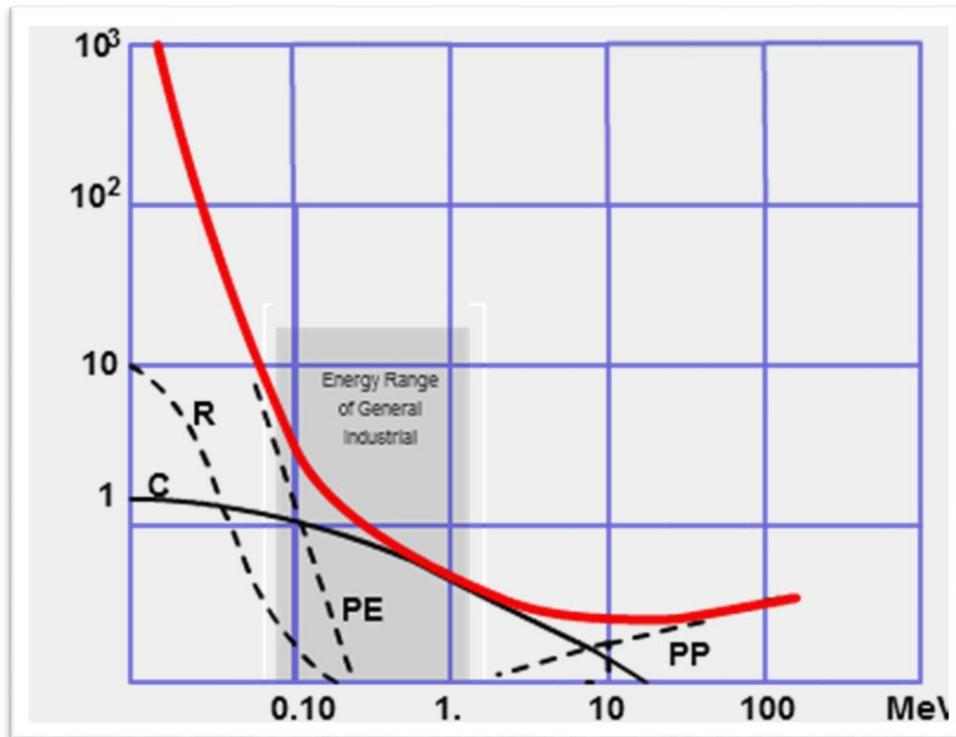
Figure 60: Contributions of mechanisms that cause attenuation of an X-ray and the power of the X-ray source (MeV). The red line is the summary of the contributors [122]

Various X-ray imaging techniques exist that each rely on different results of X-ray interactions (Figure 61). Luminescence is useful for detecting X-rays using scintillators, while fluorescence allows elemental analysis of an object and CT scanning uses attenuated X-rays. A drawback to using attenuated X-rays is a lack of sensitivity due to noise on the transmitted X-ray signal, meaning that small decreases in the attenuation of the X-ray are hard to detect [123].
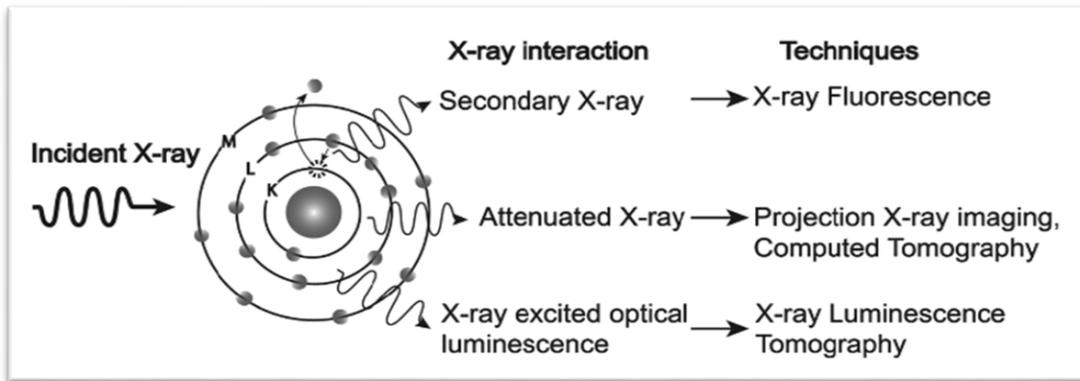
Figure 61: Possible X-ray interactions that allow different imaging and analysis techniques
[123].

3.2.2   μ -CT setup

There are two ways to setup a CT scanner, the scanner primarily consisting of the X-ray emitter and the detector. Either the scanner revolves around the object, or the objects rotates within the scanner. For medical use, the patient probably would not appreciate spinning around on a stage, so medicine tends towards the rotating scanner. This work used a scanner where the object rotates within the scanner. Regardless of whether the object or the scanner moves, the basic setup is the same, as shown in Figure 62.
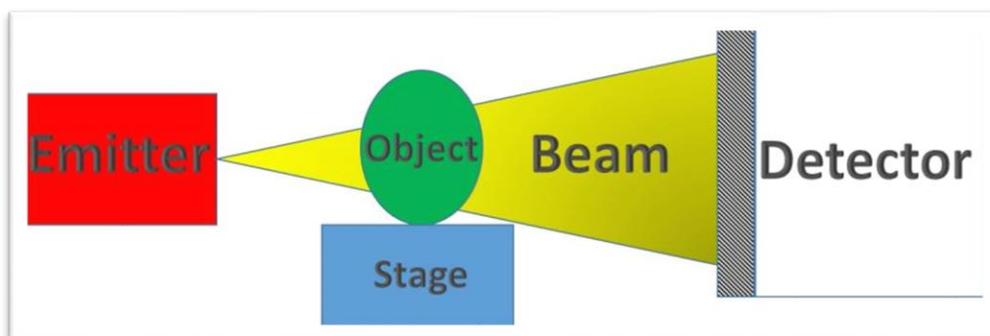
Figure 62: Schematic of a CT scanner setup

X-ray emitters comprise a cathode and an anode. Heating the metal cathode produces a thermionic reaction, releasing electrons in the direction of the anode. The electrons colliding with the metal atoms in the anode lose energy, and for the conservation of energy to remain

true, photons are produced. These type of X-rays are known as "Bremsstrahlung" or breaking radiation. Only ~1% of the energy produced by the collisions in the anode produce X-rays. The rest is mainly converted to heat, which means the anode heats up very fast. The anode and cathode need to be in a vacuum, which makes cooling the anode difficult, so it is common to use a rotating circular anode. The anode rotates so that an area has a full rotation of the anode in which to cool down before exposure to electrons again. Bremsstrahlung X-rays can collide with other metal atoms in the anode, causing electrons to be knocked out of their orbits. This leads to electrons in higher, less energetically favourable orbits, to fall into the now vacant lower electron shells. This releases more photons due to the energy saved by transitioning to a lower orbit. The value of the photons will be of discrete values based on the shells involved in the exchange. These X-rays are known as characteristic X-rays [124].

The stage sits between the emitter and the detector. Various adaptors are available for the stage depending on what is being scanned. The stage is capable of moving in four directions which allows the object being scanned to be moved as needed. The only direction the stage does not move is up and down to change the height of the object, but the emitter and detector have this capability. The detector can also move away from the emitter to allow larger objects. The stage is also capable of rotation. The combination of all of these movements allows various scanning options, all of which is automatic once the system is set up, so a scan can be left to run unmonitored overnight.

The detector was a flat panel detector (FPD) of 2000x2000 pixels that uses an X-ray scintillator to convert the X-ray into light, or semiconductor material to convert the X-rays into charge [125].

3.2.3 Scanning of the samples

The samples referred to in this work are all rail defects removed from various locations in different countries, detailed more in Chapter 4. There were 5 samples in total. The CT scans were conducted on samples 1, 2 and 4 and 5 at the µ-VIS X-ray Imaging Centre (University of Southampton, UK) using a custom built, dual source 225/450 kV walk in room (Nikon Metrology, UK). The scans were acquired using a micro-focus 450kV source fitted with a tungsten reflection target together with a Perkin Elmer XRD 1621 CN03 HS detector. A high

kV X-ray source was needed due to the density and thickness of the steel that needed penetrating. Figure 63 shows the scanner used for this work.

The samples were made as small as possible without cutting into the defect. The smaller the sample could be made, the better the images could be. The sizes of the samples varied slightly depending on the size of the discernible defect. Sample 4 was reduced in size slightly more after an initial quick scan allowed a more accurate idea of where the crack network was to be found. At its full size, sample 4 was approximately 8 cm wide, 10.5cm long and 3cm deep and the scan was able to identify the crack network clearly enough to trim the sample. Sample 2 was scanned a second time, as a smaller specimen, for a higher resolution of an area of interest. The width was taken along the bottom of the sample rather than over the curved surface. The final size of the samples scanned are shown in Table 3.
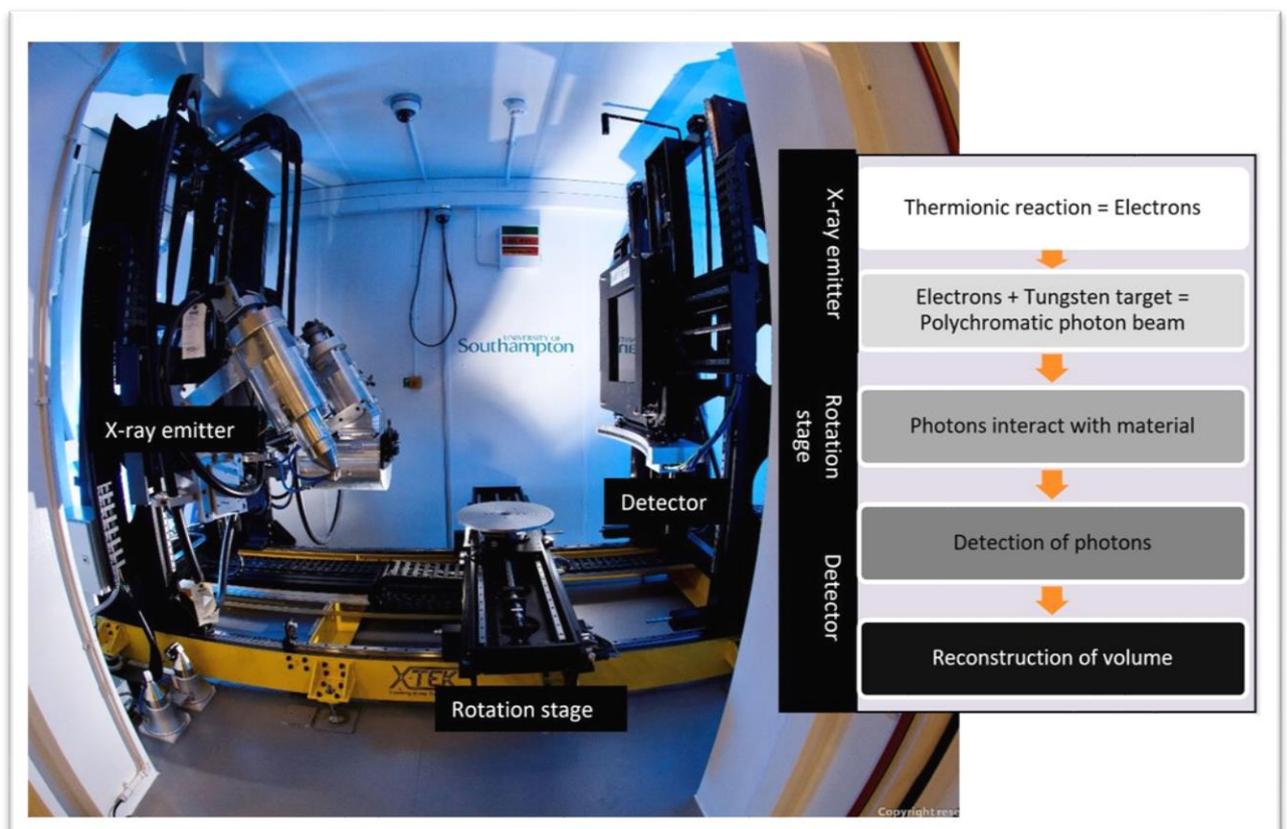
Figure 63: A photograph of the 450kV scanner showing the emitter on the left and the detector on the right. (225/450kV Hutch CT scanner image courtesy of Sharif Ahmed, μ-VIS X-ray Imaging Centre, University of Southampton).

Table 3: Dimensions of samples that were µ-CT scanned

Sample	Width (mm)	Length (mm)	Depth (mm)
1	70	55	15
2	74	85	15
2 (high res)	25	25	15
4	66	65	26
5	40	40	20

Each sample was mounted within a 3mm thick Perspex tube in a vertical orientation, so as to minimise X-ray photon penetration path length. The source to object distance was set to 235 mm, and the source to detector distance was set to 800 mm, achieving a reconstructed voxel (cubic pixel) resolution of 50 µm. 4 mm of copper pre-filtration was used in addition to the aluminium window that forms part of the target housing on the X-ray gun. Each µ-CT scan was performed at 400 kVp (peak voltage) and 248 µA, using a 177ms image exposure time; 2801, 3142 and 2601 projection images were acquired during a full 360º rotation, with 32, 16 and 32 frames averaged per projection, for samples 1, 2 and 4, respectively. The projection images were reconstructed into 2000x2000x2000 voxel 32 bit volumes using the FBP algorithms implemented within X-TEK CTPro 3D and CTAgent software packages (Nikon Metrology, UK). The higher resolution scan of sample 2 used the same emitter but with a Perkin Elmer XRD 1621 CN03 HS detector.

3.2.4 Volume reconstruction

Micro-focus cone beam X-ray CT (µ-CT) is a non-destructive volumetric imaging method, which works by acquiring a series (thousands) of 2D projection radiograph images based on X-ray photon absorption and deflection, as the specimen is rotated around a single axis, usually through 360º. The 2D projection images are then reconstructed into a 3D volumetric dataset using mathematical tomographic reconstruction algorithms, commonly based upon Filtered Back Projection (FBP) [1]. The voxel intensity (grey scale value) in the reconstructed volume

slice images reflects a combined function of the variation in X-ray absorption, (which is a function of the specimen's physical and radio-density), and CT artefacts from the acquisition and reconstruction process [1], [126]. Therefore, it can be inferred for the reconstructed images in this work, that the brighter voxels represent the dense metallic material, and the darker pixels represent less dense materials, i.e. air, cracks or voids in the sample. Care must be taken when interpreting images as artefacts can provide distorted images [126].

Two scans were used to cover the full height of sample 2, and the reconstructed volumes were concatenated using Fiji/ImageJ. Each sample volume was converted to 8 bit in Fiji/ImageJ to reduce the computation time for analysis, and saved as a raw volume for review in VGStudio MAX. The results were viewed as three windows of orthogonal slices through the rail, providing a longitudinal, cross section and aerial view of the rail in orthoslices. The data was also investigated using Avizo, mentioned in section 4.3, building a 3D crack model. Cracks were highlighted every five orthoslices and added to the model, then the planar interpolation feature was used to interpret where the crack was most likely to be between the five slices. The built in crack detection software struggled to identify the cracks accurately if at all, hence the manual segmentation method. This was done for both the longitudinal and cross sectional slices.

3.2.5 Artefacts

The use of a polychromatic beam has an effect on the reconstructed scan volumes leading to artefacts appearing within the volume. The stated kV of a source does not mean that all X-rays produced are of that energy. They are created as energy wave packets, so some will be lower than the peak stated energy. The lower energy X-rays will scatter more easily than higher energy X-rays, which can lead to a less clear image. This can be improved using a filter. Usually made of Al, Cu, or Pb, the filter will block lower energy X-rays to make the imaging clearer. The thickness of the filter depends on the material used and the power of the X-ray source.

Scattering being a primary beam and a scattered secondary beam provides a basic concept. However, many sources of scattering occur and the software that reconstructs the volume has to interpret all of these sources.

3.2.6 Data analysis

Once a raw volume had been created from the scan data, software packages were used to view, post-process and display the data. Fiji/ImageJ is an open source image processing software (National Institutes for Health, USA) used to initially process and view the raw volume. It was used to create videos of the volume and resliced the volume into a different orientation if needed.

Avizo 9.3 (FEI SAS, Thermo Fisher Scientific, USA) is a 3D materials characterisation software that allows the visualisation and analysis of large data sets. The user can segment volume data to form colour coded models, i.e. grey scale voxels, are labelled with colours corresponding to different structures identified within the specimen, such as different materials and features such as cracks and pores.

For this work, Avizo was used to initially map out the planar crack network of the defects within the rail samples. In the orthoslices of the scan the image was changed from greyscale to green and black to make the cracks more visible by eye. A touch screen pad and stylus connected to the computer were used to manually draw in the cracks for each 2D slide, first longitudinally then cross sectional. The model is then improved from a mesh structure into a smoother planar structure using linear interpolation between two known points. The two known points are the lines that have been drawn into the orthoslice manually. The models created gave a crude idea of the crack network morphology. However, for use with cracks in steel, much of the processing was manual and so was not very accurate (crack planes had a "woolly jumper" texture due to the lines that had to be drawn in). Ideally, a process that was more repeatable and less user dependant is desirable.

VGStudio MAX (Volume Graphics GmbH, Germany) is a voxel-based software that was used to view the orthoslices, take measurements of features within the μ-CT volume data and eventually build a 3D colour model of the crack networks in a more repeatable method. The software proved easier to build and animate the models. It also suffered less frequently from technical issues and was available as a free viewer file (myVGL) for viewing the model. The viewer, myVGL, would run on a typical laptop without the need for the high computing power needed to run either Avizo or VGStudios MAX. The viewer did not allow processing of the volume though. The models were more accurate due to the techniques that were developed,

such that the crack planes were of a representative thickness and contained a texture typical of a crack plane.

3.2.7 Building the models in VGStudiosMax from the raw volumes

In order to build a model of the crack network, the cracks needed to be isolated from the bulk volume. The cracks generally appeared as darker voxels than the steel so highlighting all voxel within a certain range can pick out most of the cracks. There is a module called 'crack segmentation' that can do this automatically. However, with the whole sample being dense steel, there were issues with scatter and the production of surface hardening and other artefacts within the volume data [1]. Sectioning the bulk into regions (Figure 64) reduced the range of the voxels allowing for more accuracy when detecting the cracks. This resulted in the associated histogram filtering from a double peak, background noise and the object, to just a single peak (Figure 65). The object can be automatically isolated from background noise by isolating the surface using a software module, then cropping out the background. This produced less desirable results for a steel sample. The diffuse boundary between the object and the background (surface effects) meant that part of the object would be cropped or some background left in depending on the settings.
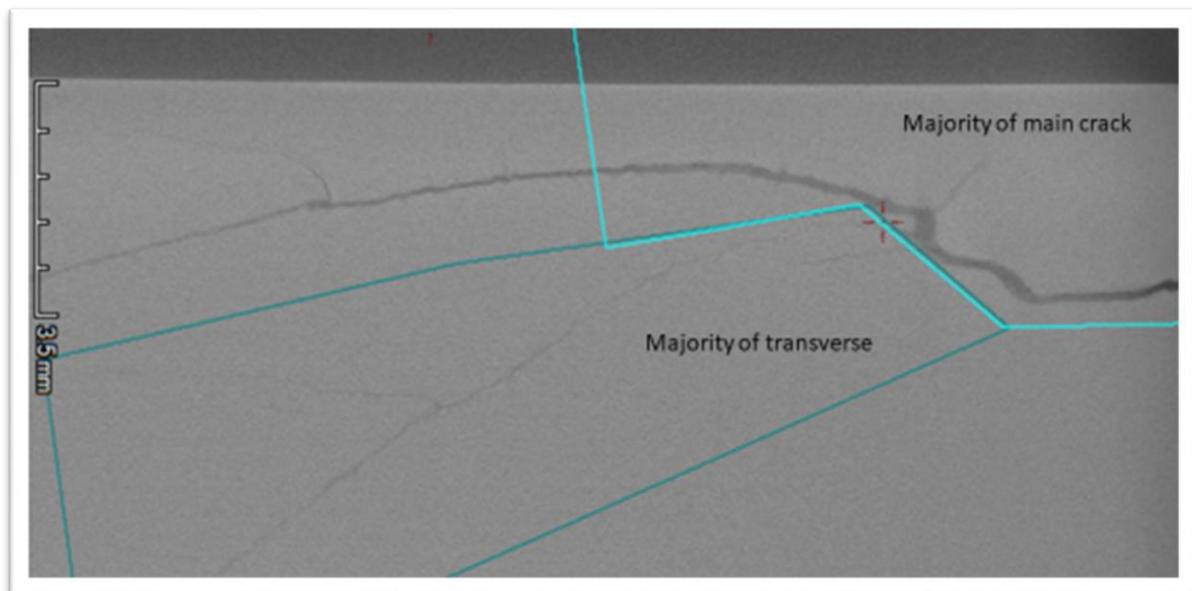

Figure 64: The sectioning of sample 2 into regions after its high-resolution scan
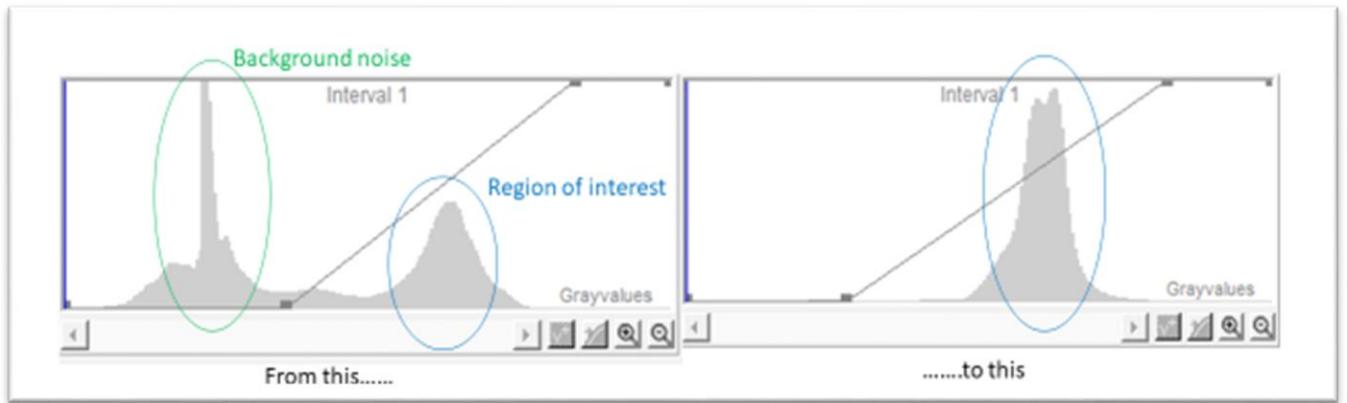
Figure 65: The change in the histogram due to isolation of the regions of interest

Highlighting a crack from the bulk volume by manually drawing around the crack and extracting the region reduced the range of voxels within the region even further (Figure 66). This made it easier for the 'crack segmentation' module to distinguish between the crack and the bulk material. The automatically detected region would then replace the manually highlighted region. There would still be areas where there was a lot of noise, especially near the edges of the steel, resulting in the highlighting of the noise as well as the cracks. These highlighted areas would be manually removed and replaced with regions highlighted using the 'region grower' module (Figure 67). Region grower uses a reference voxel value and identifies other voxels that fall within a specified tolerance (number close enough to reference value) and radius of the reference voxel. Increasing the threshold value will add a larger range of voxels but there tends to be an optimum value, above which scatter becomes highlighted too. This is used to slowly build a region in sections by using tighter tolerances in 'noisy' areas to reduce anomalous voxels being added. Each region would be added to the bulk of the selected crack one small region at a time. This methodology, which was developed by trial and error by the author, resulted in a clearer detection of the cracks with less manual interpretation. This overcame some of the accuracy issues caused by scanning relatively large dense blocks, such as steel. The difference that noise can make to a model's accuracy is highlighted in Figure 68. Noise is reduced with sample size because the signal to noise ratio is improved, due to less material being penetrated, i.e. better transmission due to more photons being detectable.

Once all crack planes have been highlighted and separated from each other they can have colours assigned and be shown with the bulk material around them removed or partially transparent. Other features such as surface damage can be added in the same way until the

model is built. Surface features are easier to add from the aerial view. Surface features suffer from much more scatter effects than most of the volume, so showing the surface in detail is better using an infinite focus microscope.
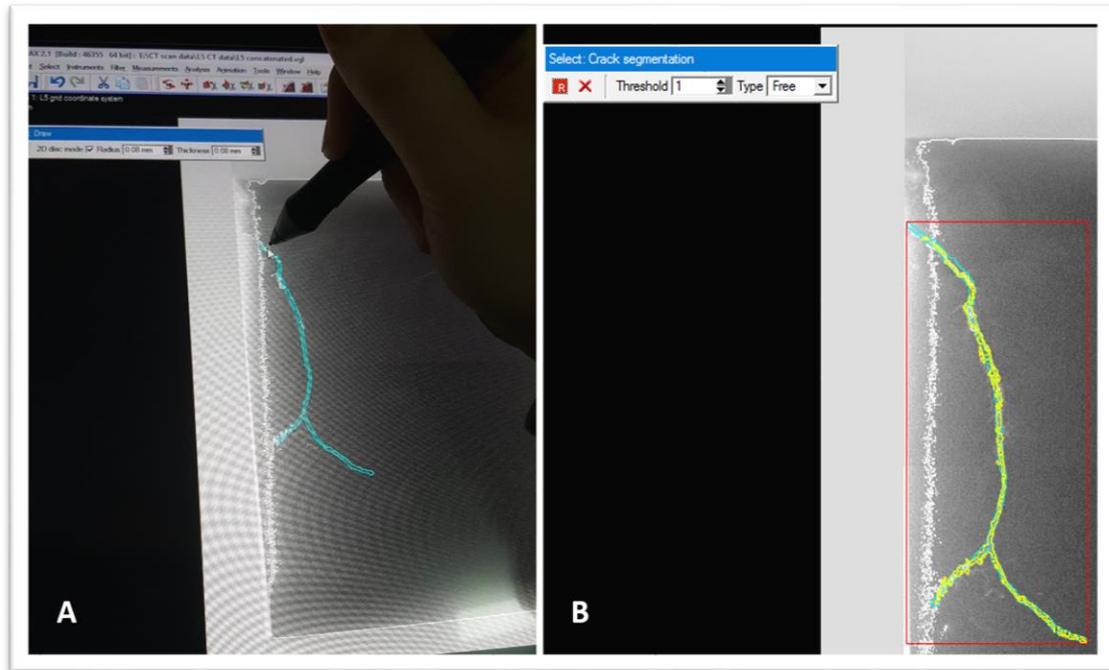
Figure 66: Crack segmentation method a) Manually drawing around the crack b) Using crack segmentation module to select the crack within manually highlighted region. Blue is manually highlighted and yellow is the module selected voxels.
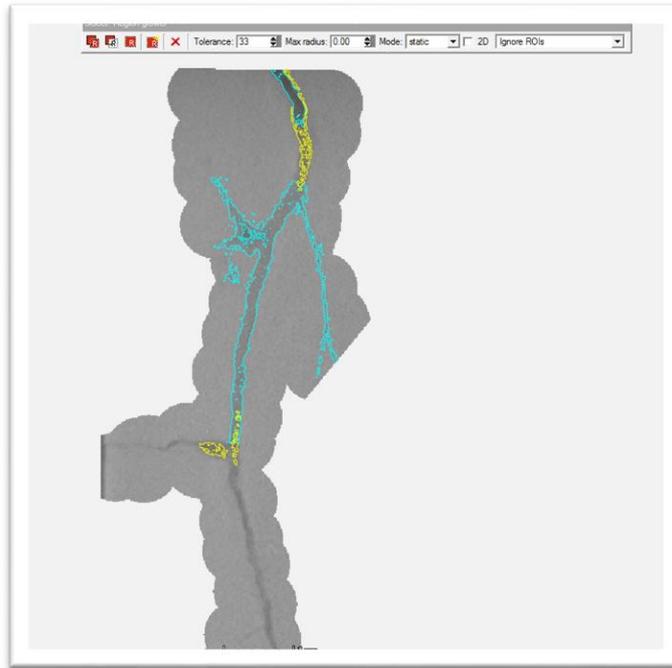
Figure 67: Region grower module: blue is the bulk of the selected crack and yellow is the newly highlighted region, which when added to the bulk will also turn blue, ready for another region to be selected.
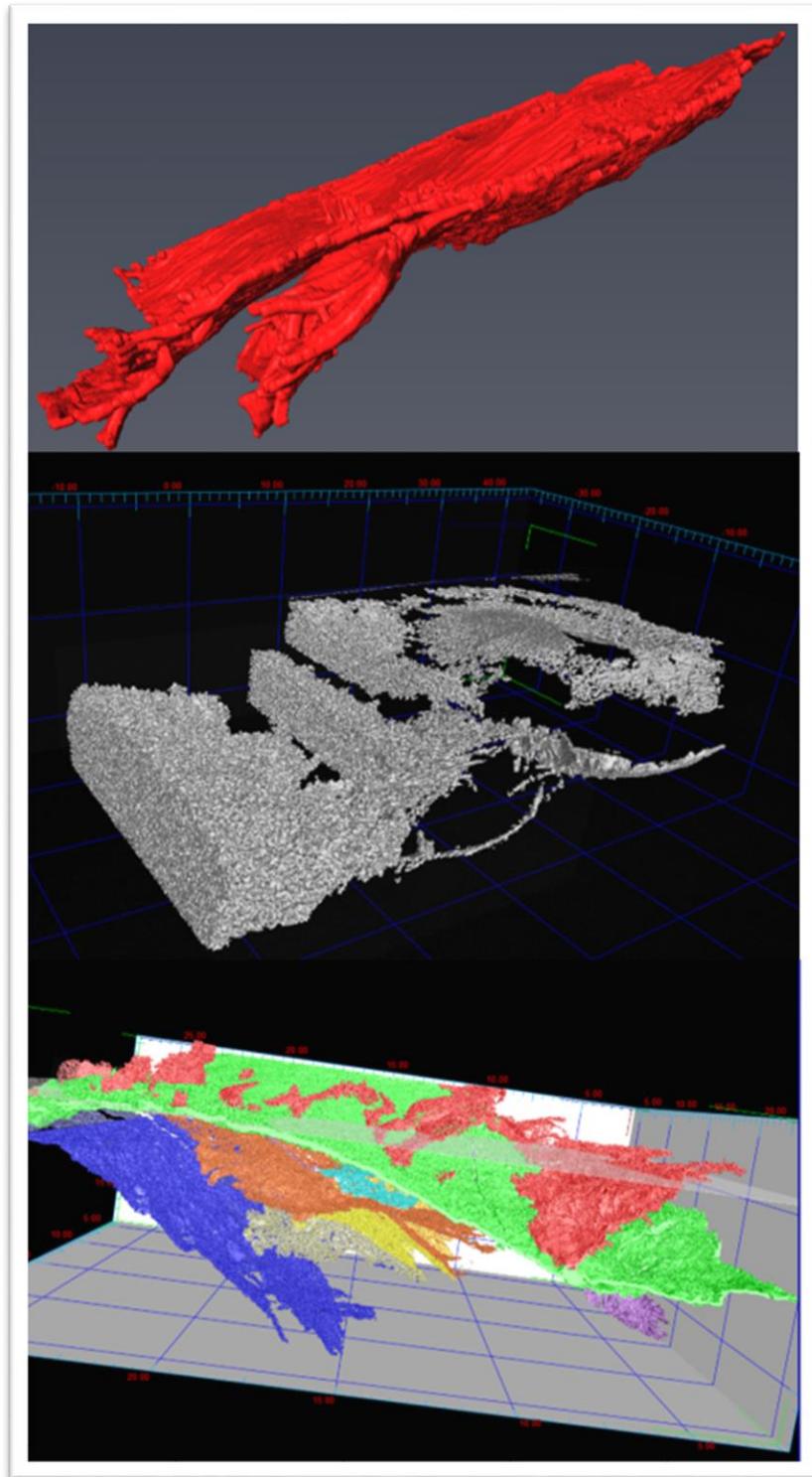
Figure 68: Visual comparison of models created from sample 2; Top) Avizo manually drawn model. Middle) Automated process applied to large sample volume showing the effects of scatter/ noise. Bottom) Automated process of higher resolution scan using crack segmentation and region grower modules.

3.3 Metallography

Traditional metallurgical methods were used to verify the CT scans and to inspect the microstructure of the defects. Sample preparation was typical for rail steel, although the presence of many cracks proved challenging in some samples. First an overview of the typical method used for each sample will be covered before discussing the individual sectioning of each sample. They differed dependant on their identified areas of interest.

Most samples began as a full rail of a varied length, so before micro-preparation, the large rail samples were cut using industrial band saws. Sample 1 was received already cut to approximately 80x100x20mm and sample 3 was only the head of the rail. It is suspected that they were both cut to this size using industrial band saws based on the markings on the cut surfaces.

Micro-preparation involved the processing of the relatively large rail samples from the size of the defect down to etched and polished samples that would show the rail samples pearlitic microstructure under a microscope. The samples were cut using an abrasive disc cutter, mounted in 32mm Bakelite discs, ground and polished down to a 1 micron diamond suspension finish. The samples were then etched in 2% nital (2% nitric acid and 98% ethanol) as is typical for steel to reveal the microstructure. The samples were then visually inspected as detailed in section 3.6. Further details of the samples are discussed in Section 4.2, Table 4.

3.3.1   Sample 1

A lateral cut (top to bottom in Figure 69) was made to separate the two lobes of the defect, making it possible to capture the length of the defect as two halves that would fit into standard 32mm mounts. Longitudinal cuts were then made based on features that had been deemed interesting from the CT scan volumes, usually due to an unusual feature or source of crack branching.

Figure 69: Sectioning diagram for sample 1 with designations used for each piece and the view direction as blue arrows. R7 is the gauge corner, which was removed before the picture was taken.

3.3.2 Sample 2

The key feature of interest in sample 2 was L3 (Figure 70). This was the reduced sample that was scanned again for the high-resolution model. Its location was determined by the need to capture as much of the faint downward growing cracks in a 25mm square as possible. This was measured using the first scan of this sample. After the scan, L3 was sliced longitudinally to verify features within the second scan. L1 and L6 were either side of the cut often seen in literature through the initiation site of the main defect, showing the characteristic inverted-V crack. L5 was a lateral view and L7 was a lateral view of the surface breaking crack. Sample designations containing an 'x' were not mounted or polished.

Figure 70: Sectioning diagram for sample 2 with designations used for each piece and the view direction as blue arrows.

3.3.3 Sample 3

Sample 3 was sectioned to remove the intact lobe from the bulk of the rail before releasing the upper plate of the lobe as shown in Figure 71. The intact lobe was then sliced longitudinally, through what was assumed to be the initiation site (red line), the centre of the lipped region (purple line) and the green line and brown lines to reduce the sample size. No other sectioning was conducted after the initial investigation, as the cause was quickly determined, as discussed in Chapter 5.
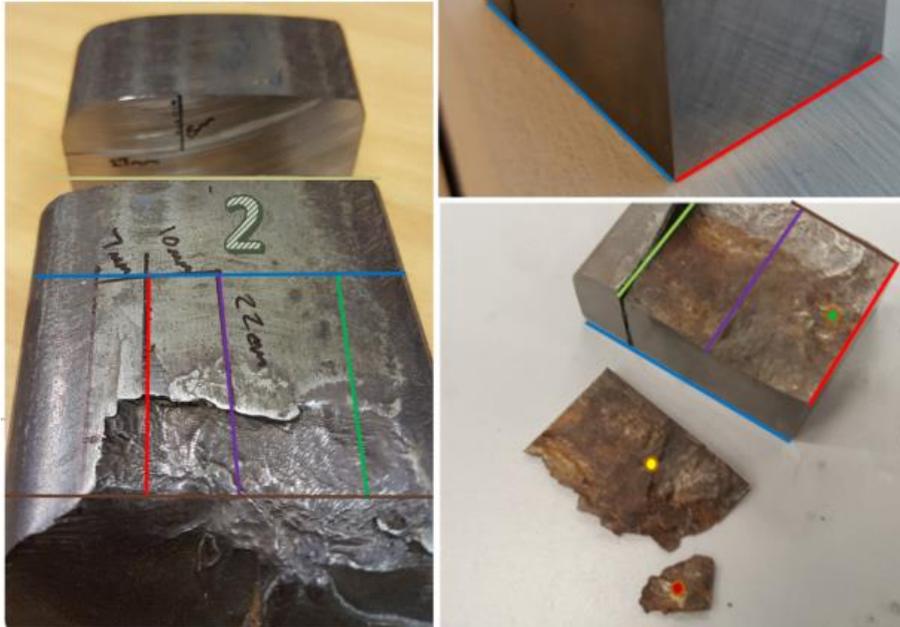
Figure 71: Sectioning of Sample 3 involved 3 cross sectional and 3 longitudinal cuts. The blue and green cuts were needed to release the upper plate of the intact lobe of the Squat. The red line was through the proposed initiation site and purple was through the lipped region. The others were to reduce the size of the specimens.

3.3.4   Sample 4

The first cut through sample 4 was longitudinal through the initiation site (between H3 and H4 in Figure 72). Then a lateral cut was made below H1 and H2. The crack plane was fairly simple so not many sections were needed to verify the scan. HG was a later cut made to investigate how the surface varied along the transition between the rough ground surface and the smoother surface of the weld.
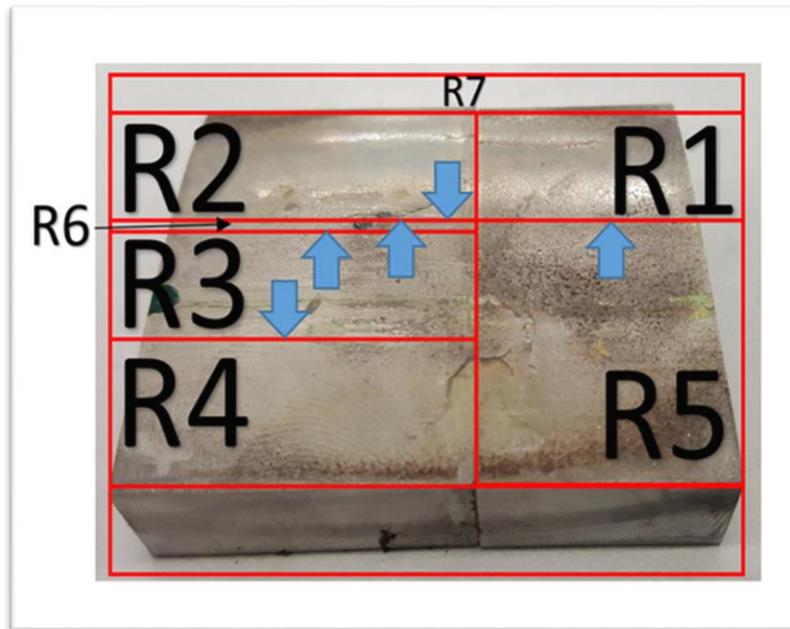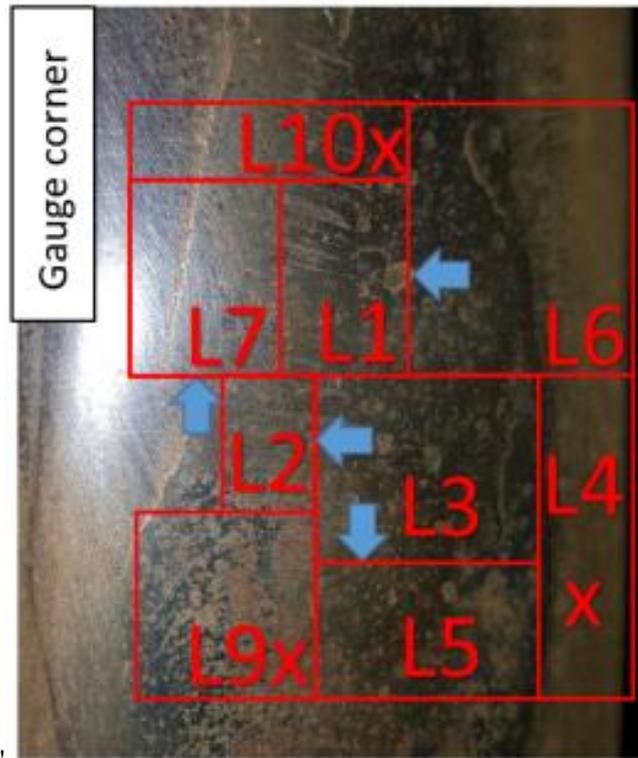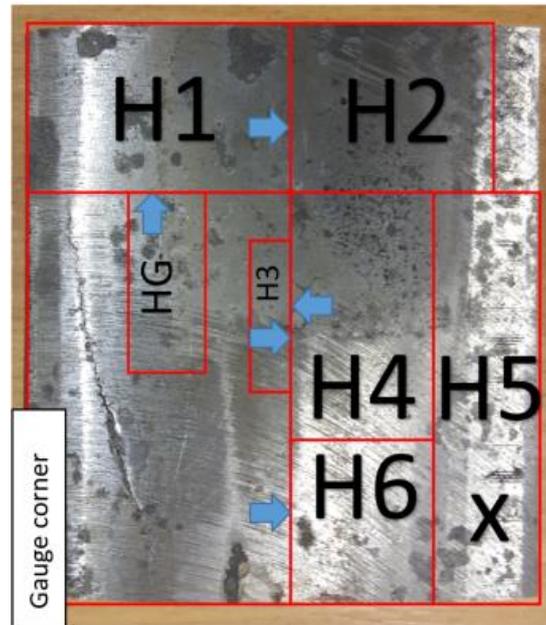
Figure 72: Sectioning diagram for sample 4 with designations used for each piece and the view direction as blue arrows.

3.4 Hardness mapping

3.4.1 Vickers indenting

Hardness variations were measured within the upper layers of sample 1 due to the grain refinement seen and on sample 4 due to the presence of the weld. EmcoTest DuraScan 70 was used with a Vickers indenter, interfaced with ecos Workflow software. Being fully automatic, with a 6-position turret (lenses and indenter) and automatic test cycle system, allowed the mapping of a pre-specified pattern (e.g. a 10x10 point grid 1mm apart) with repeated conditions and minimal influence from the user.

A dwell time of 5 seconds and an optimised approach setting were always used. A grid pattern was initially used but would be adjusted when the points fell on or close to a crack or surface edge. Point spacing was kept to between 0.3 and 1mm apart, dependant on surface obstructions and the size of the area of interest. The indenter operated at a load of 1kgf with an optical zoom of 20x for assessment. Hardness results that were flagged up in red at the end of the indenting would be manually adjusted if need to get an 'OK' status from the software. The results were then exported to Microsoft Excel, then MATLAB for visual presentation.

3.4.2 Nano-indentation

Nano-indentation was used in two areas of interest. On a deformed region around a crack in sample R3 and on a WEL in sample R6. They were performed using a Bruker TI Premier.

For R3, six traverses were performed using a load of 2mN and a spacing of 10μm between indents. For R6, three traverses were performed using a load of 2mN, but the spacing and indents per traverse varied. Traverse 1 contained 9 indents with an indent spacing of 10μm. Traverse 2 contained 12 indents with an indent spacing of 4 μm. Traverse 3 contained 10 indents with an indent spacing of 4 μm.

3.5 Microscopy (Optical and SEM)

For optical inspection, samples were prepared as covered in section 3.3 to reveal the microstructure. Then the microstructure was inspected and documented using a Nikon Eclipse LV150 optical microscope with Buehler Omnimet software and Zeiss Axio Imager.A2m optical microscope with AxioVision4 software.

SEM images were obtained using an Inspect F Scanning Electron Microscope by FEI company. Samples were prepared in the same way as detailed in 3.3 except they were mounted in conductive Bakelite to allow discharge from the sample during SEM use. A beam voltage of 20kV and a spot size of 2.5 were used to capture secondary electron (SE) images.

Energy dispersive X-ray (EDX) spectroscopy was conducted on a FEI Nova NanoSEM 450 to identify the elemental constituents of cracks and pits within the steel samples. SEM images were taken at 15kV with a spot size of 3.5.

3.6 References

[1]   A. H. Compton and D. Ph, "THE SCATTERING OF X-RAYS TIIE present paper will be confined to a discussion of some of the points which present to us a revolutionary change in our ideas regarding the process of scattering of electromagnetic waves . No attempt will be made to describe t," vol. 198, no. 1183, 1923.

[2]   Iowa State University NDT Education Resource Centre, "Sources of Attenuation," 2014.   [Online].   Available: https://www.nde-ed.org/EducationResources/CommunityCollege/Radiography/Physics/attenuation.htm. [Accessed: 07-Jul-2019].

[3]   H. Chen, M. M. Rogalski, and J. N. Anker, "Advances in functional X-ray imaging techniques and contrast agents," *Phys. Chem. Chem. Phys.*, vol. 14, no. 39, pp. 13469–13486, 2012.

[4]   R. Nave, "X-ray production." [Online]. Available: http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/xtube.html. [Accessed: 18-Jan-2019].

[5]   J. A. Seibert, "Flat-panel detectors: How much better are they?," *Pediatr. Radiol.*, vol. 36, no. SUPPL. 14, pp. 173–181, 2006.

[6]   E. N. Landis and D. T. Keane, "X-ray microtomography," *Mater. Charact.*, vol. 61, no. 12, pp. 1305–1316, 2010.

[7]   J. F. Barrett and N. Keat, "Artifacts in CT: Recognition and Avoidance," *RadioGraphics*, vol. 24, no. 6, pp. 1679–1691, 2004.

Chapter 4 : CT scanning to compare Squat type defects

As a strategy to understand the cause of Squats, multiple Squat defects were received courtesy of rail-track operators, to compare to each other. The investigation aimed to identify common features in the hope that this would point to any common causes for these defects. The four samples examined were from three different traffic environments, which was expected to result in differences within the defect.

A volumetric comparison was achieved using μ-CT scanning, allowing the overall structure of the defect to be examined, rather than the usual 2-dimensional section view. However, μ-CT scanning has not been used for such large samples before, so the initial scan results needed to be checked to ensure that the results are representative. Therefore, this Chapter addresses the first two research questions:

1. Can μ-CT scanning accurately detect crack networks on a large scale within a rail?
2. Are all of the features captured during scanning real, or do artefacts make the results deceptive?

The causes of the defects proved to be different despite the developed defect appearing similar. There were also differences in the morphology of the defects, although attributing features to traffic type was not as obvious as expected.

4.1 Introduction

This Chapter documents work that provides a visualisation of the 3D nature of Studs in a way not seen before, as a 3D model of the crack network from an actual in-service defect. Squats and Studs are defects in railheads that share features but have different causes. This Chapter examines four defects from three different traffic environments, to compare features using μ-CT X-ray scans, a non-destructive volumetric inspection tool [1]. One sample was not scanned as it had already fractured and shelled but was included because it was the only Squat out of the four defects.

μ-CT scanning has been used before to successfully investigate the form of Squats by Naeimi et al., [2], who also give good justifications for why CT scanning is a powerful NDT tool, maximising the information that can be retrieved from a sample. This Chapter builds on this previous work by expanding the scan to the entire defect. This requires a 450kV μ-CT scanner due to the difficulty of penetrating substantial thicknesses of steel due to its density.

The scanned volumes create a map of the defect, allowing accurate targeting of regions of interest. These regions are then inspected further using more traditional metallographic methods, which also provide verification of the CT scan accuracy. All defects contain similar superficial features but the depth and severity of the subsurface damage varies. The scan volumes were later used to create 3D models of the crack networks.

The 3D models of the defects visually illustrate variations within the category of squats. Samples 1 and 2 showed the influence of hollow wheels, as the defect seemed to initiate on the field side, grow down and towards the gauge side, before resurfacing as the longitudinal crack noted in all four defect samples. Sample 2 also showed a second initiation site, from which a crack developed and linked up with the first initiated crack. Sample 3 failed in–service and was the only Squat. Unfortunately, unlike most documented Squats, this one failed due the well-known failure cause of containing many inclusions, attributed to the steel being ingot cast. Sample 4 was on a weld and is believed to have developed due to contamination of the weld providing initiation sites. This is different to the typical Squats on welds explanation of hardness differentials [3].

Of the four samples, three were believed to be Studs and one was a Squat (sample 3). Each defect developed for different reasons, although the two metro samples (1 and 2) were similar. One of the Studs shows branching of cracks that, based on its changing angle of growth and the depth it had already reached, could continue to grow into transverse defects, breaking the rail. The three defects that were scanned would all be classed as Studs, but their crack morphology varies, possibly because they are all from different traffic environments. They also show slight differences to other Studs in literature.

This Chapter will introduce the four samples briefly. Then μ-CT scanning will be discussed as a method for inspecting the samples, with some initial results just to show that the technique worked accurately. Finally, features of each sample will be commented on in more depth.

4.2 Introduction to rail defect samples

The samples were sourced from metro, mainline and high speed track in the UK and France. The samples have been characterised using 3D microscope mapping, X-ray μ-CT, optical imaging and scanning electron microscopy (SEM) in order to provide a comparison between them. These characterisation techniques also aid the distinction between Squats and Squat Type Unidentified Defects (Studs). As a simple comparison, Studs do not have the heavily deformed layer on the surface like Squats do. Squats take much longer to form than Studs: ~100MGTs of traffic to ultrasonic detection as opposed to ~30MGTs for a Stud [4]. The four samples that were compared are shown in Table 4. Actual chemical analysis was performed as it was assumed that the steel grades would have been tested and therefore within specification.

Table 4: List of samples examined

Sample number	Location removed from	Traffic type	Grade
1	France	Metro	R260
2	UK	Metro	R220
3	France	Mixed	R260
4	UK	High speed	R260

The samples were subjected to a variety of tests to assess and capture the characteristics of the defect. The tests are shown in Figure 73, although not all samples went through every stage. Surface analysis was carried out to document the surface of the sample and choose points of interest on the sample. It was also used to compare the surface of the samples to each other and is discussed in relation to the scan data in Chapter 6. μ-CT scanning was not conducted on sample 3 as the surface had already shelled. All samples were sectioned, prepared and analysed optically with the most interesting regions being inspected using SEM. Surface analysis is covered more in Chapter 6 and sample preparation is covered Chapter 3.
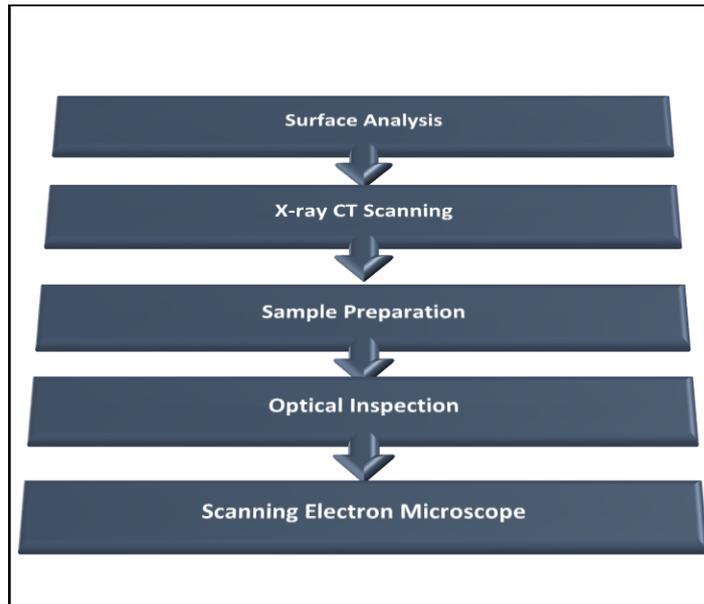
Figure 73: The sequence of tests conducted on the Squat samples

4.3 μ-CT scanning

Details about the technique of μ-CT scanning is covered in Chapter 3, but a brief summary to accompany this Chapter is provided before moving on to the verification of the scans.

Micro-focus cone beam X-ray CT (μ-CT) is a non-destructive volumetric imaging method, which works by acquiring a series (thousands) of 2D projection radiograph images based on X-ray photon absorption and deflection, as the specimen is rotated around a single axis, usually through 360º. The 2D projection images are then reconstructed into a 3D volumetric dataset using mathematical tomographic reconstruction algorithms, commonly based upon Filtered Back Projection (FBP) [1]. The voxel intensity (grey scale value) in the reconstructed volume slice images, reflects a combined function of the variation in X-ray absorption, (which is a function of the specimen's physical and radio-density), and CT artefacts from the acquisition and reconstruction process [1], [5]. Therefore it can be inferred for the reconstructed images in this work that the brighter voxels represent the dense metallic material, and the darker pixels represent less dense materials, i.e. air, cracks or voids in the sample. Care must be taken when interpreting images as artefacts can provide distorted images [5].

4.3.1 Scan verification

The presence of CT artefacts in the volume images, from effects within the scanning and reconstruction process such as photon scatter, partial volume effect, image noise and beam hardening, could lead to spurious features within the reconstructed volume. A visible artefact when viewing some slides is the apparent 'shadows' from cracks that are very close to each other (streaking artefacts), which can make it look like there is a small void in the material. In order to be sure that the scan does not show artefacts, it must be compared to micrographs of the scanned segment. Shadows that occur along a crack were of particular interest as they may hold clues as to why the crack plane develops in certain directions.

Measurements were taken using the VGStudio Max software caliper tool and then the sample was marked up with cutting lines in preparation for physically cutting the sample to expose some of the interesting features found in the scans. The samples were prepared as described in section 3.3 and then compared to the slices to see how accurate the scans were.

4.3.1.1 Results

In Sample 1, an orthoslice of interest was one that showed a crack that branched down from what appeared to be a small void on the surface, which was visible from the surface scans. The crack then displayed interesting behaviour by branching in a circular shape as though propagating around something that it could not penetrate. There were also small shadows on the scans, which were investigated to see how real they were. Figure 74 shows the two comparisons of the scan slices mentioned and their accompanying micrograph. The upper red circle shows a region where the finer crack is not resolved, but the larger crack is seen and the lower circle shows how a branching crack can create a shadow.

Sample 4 also showed a shadow near the surface breaking feature in the centre of the rail and the verification of that scan is shown in Figure 75. Figure 75 shows a shadow found on the crack within the CT scan, the cause of that shadow and verification of the structure shown in the green box, which is responsible for the C shaped crack that breaks the surface of the running band. It also shows a zoomed image of the area that caused the shadow to appear, associated with a small void. These voids are believed to develop from the small 'islands' like the one labelled "shadow from crack branching" in Figure 74b. It is believed this is due to rubbing of

the crack surfaces wearing away and eventually breaking up the 'island' (Figure 76a) into small voids (Figure 76b). The broken debris from the island should cause enough scatter in the X-ray to appear like a void before being washed out during sample preparation.
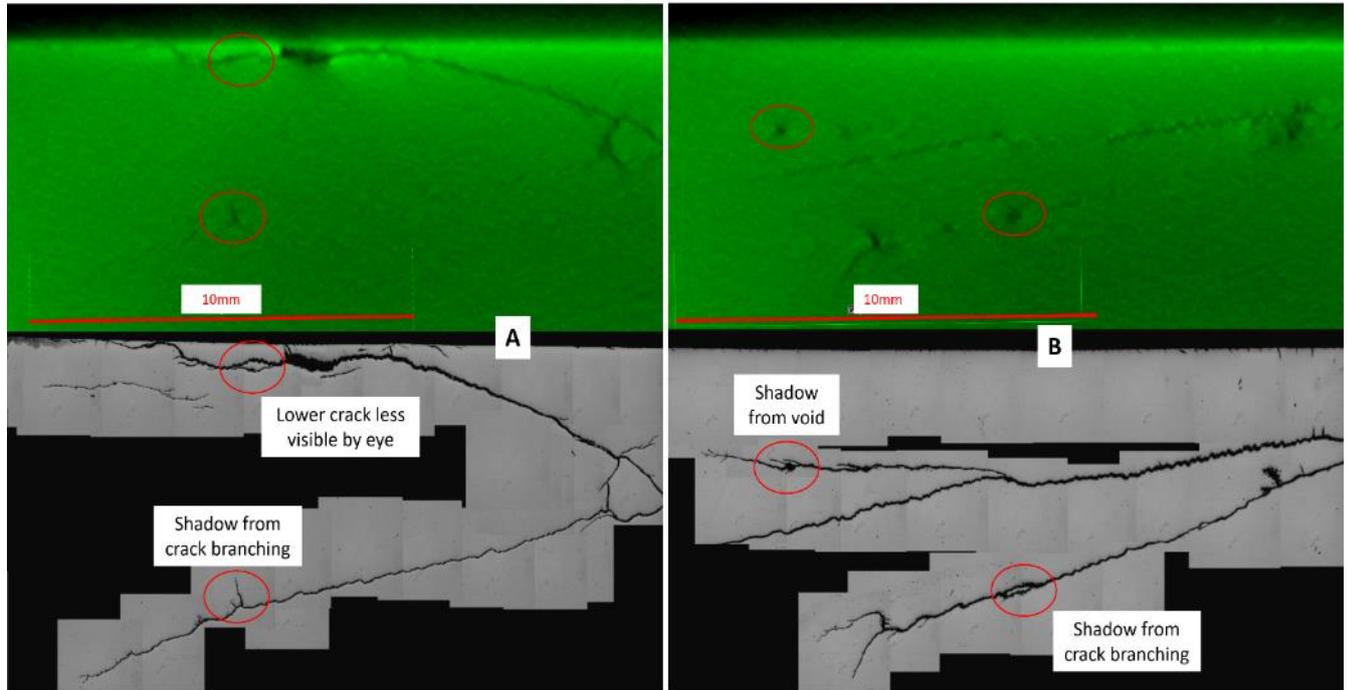
Figure 74: Comparison of longitudinal CT scan orthoslice to longitudinal section in sample 2 a) The surface void visible from the surface scans with a shadow and its cause highlighted b) Branching that shows shadows in the scan and the same cause as (a) in the micrograph.
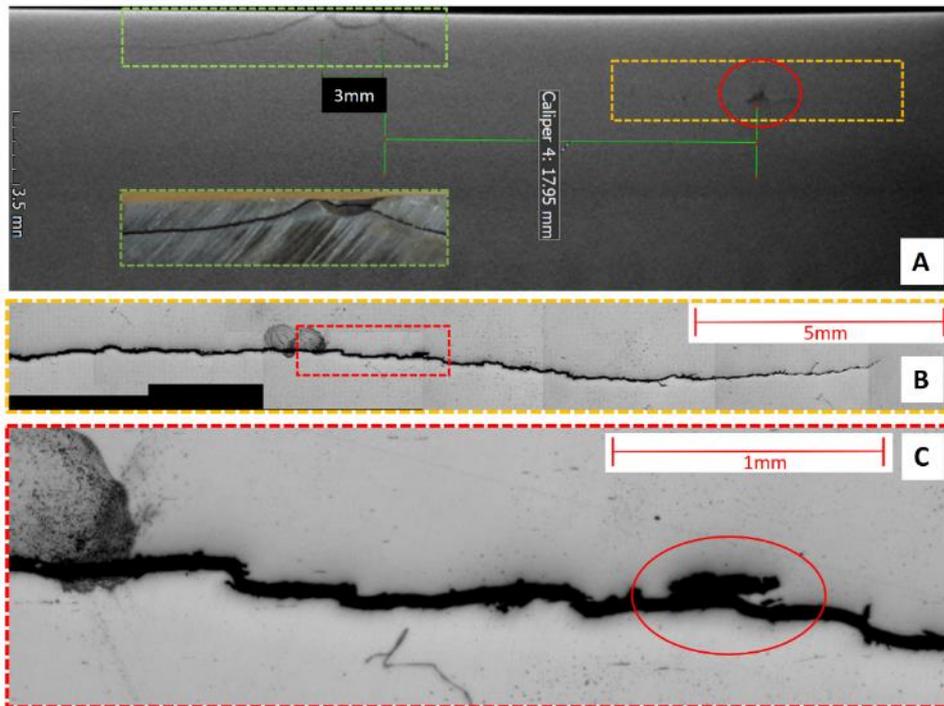
Figure 75: Longitudinal slice through the c-shaped crack in the centre of the surface of sample 4. a) CT scan orthoslice showing the surface breaking crack, highlighted in the green box, with a photo of the sectioned sample inset in the lower green box b) An unetched micrograph of the area highlighted in the orange box c) An expanded view of B, with a red circle showing the cause of the shadow in the CT scan orthoslice (a).
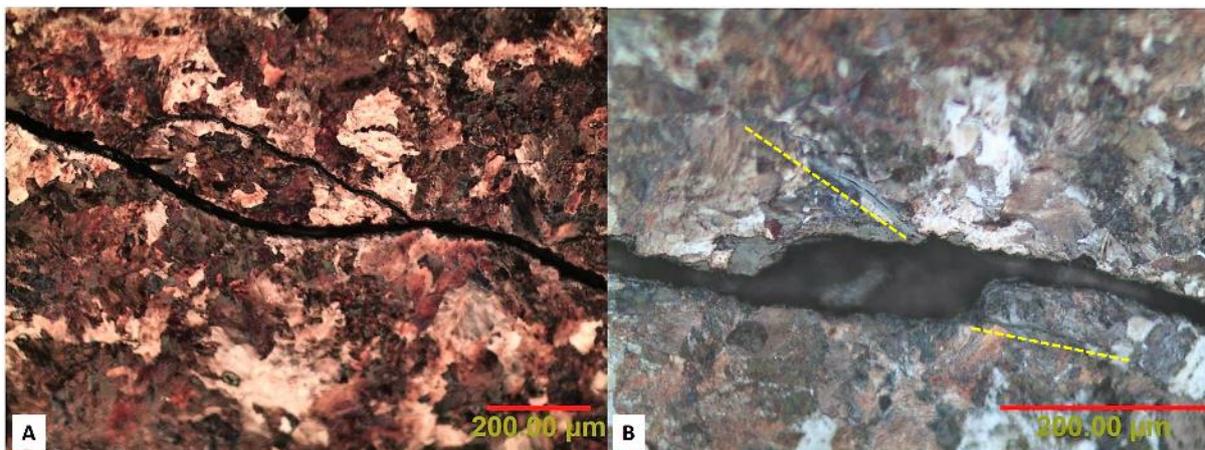
Figure 76: Micrographs showing 'islands' due to branching cracks such as the one seen in Figure 74b. Image (b) is a small void that probably developed to look like (a) before the material detached.

4.4 Rail defect samples

This section introduces each of the samples as they were received in their entirety, with some comments about features noted in the visual inspection, before scanning or destructive testing began.

4.4.1 Sample 1

Not much is known about the track location and traffic experienced by Sample 1 (Figure 77) other than it was from an inclined track near a station on a metro line. It shows signs of sliding surface damage and the dark spots are quite well developed. It contains a lot of surface damage including four obvious surface cracks, streaks of white etching layer (WEL), a pit in one of the cracks, what seems to be 'snakeskin' damage just outside the widened running band. 'Snakeskin' is also known as running surface checking or flaking and is a precursor to spalling of the rail surface [6]. There is also the very common V-shaped crack that Squats are known to exhibit, although the angle within the V in this crack is more obtuse than expected. There is a faint series of small cracks along a ridge that extends from the apex of the V-shaped crack out between the two lobes which is often observed as Squats develop [7].

On initial inspection, the presence of some lipping on the gauge corner of the sample combined with the typical black spots and v-shaped cracks, made this sample appear to be a Squat. The features that makes it more likely to be a Stud are the lack of any significant WEL on the surface of the rail and the minimal plastic deformation. The only plastic deformation found was in the vicinity of cracks and the decarburised lip on the gauge corner. The decarburised lip suggests strong gauge corner contact during sliding conditions and so the inclined track may have also contained a curve.
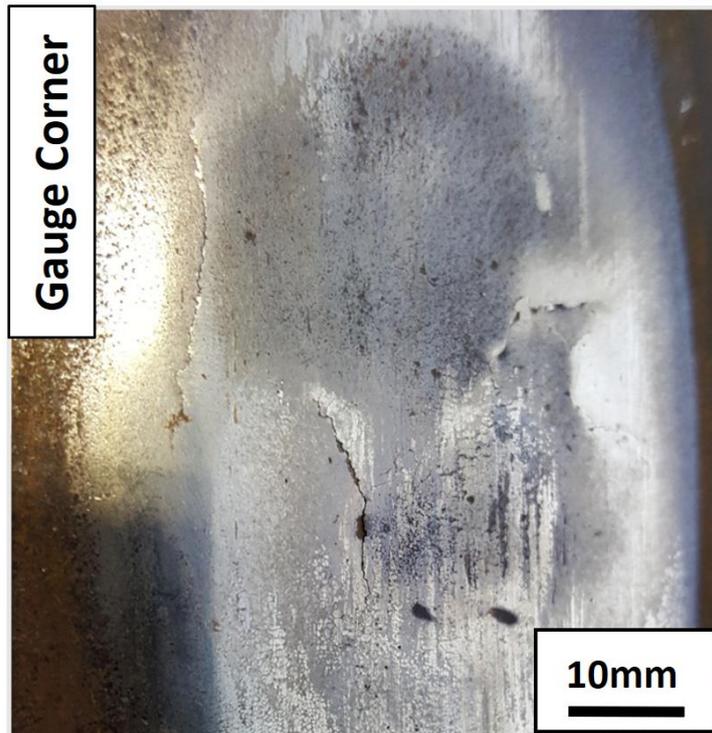
Figure 77: The surface of Sample 1. The dark regions are corroded as the surface has depressed and the bright areas show where wheel contact has occurred recently.

4.4.2   Sample 2

Sample 2 was taken from the high rail of a uni-directional rail in a gentle curve just over 1km along a 3.2km distance between two stations. This track has relatively high line speeds and is very busy regarding passenger trains compared to other lines on the same network. However, this sample came from a fairly rural overground section near the end of the line so the train volume may be less.

Sample 2 is shown in Figure 78, and is post MPI (magnetic particle inspection), a non-destructive test that uses magnetic ink on a white contrasting background to highlight defects in the surface. The expected V-shaped crack typically seen in a classic Squat is present in two locations, signifying the possibility of two defects. The overall defect is well developed as it contains a long crack on the gauge side where the crack plane has surfaced. The crack also has sharp edges unlike the embedded cracks found on the other samples, possibly a sign that the crack surfacing is very recent.

Sample 2 also has significant pitting/ discreet material removal. These look like corrosion pits, but are more likely localised delamination and flaking characteristic of "ratchetting" [8]. Ratchetting is incremental plastic deformation of the upper surface of the rail material due to cyclic strain build-up. These pits are discussed more in Chapter 5.
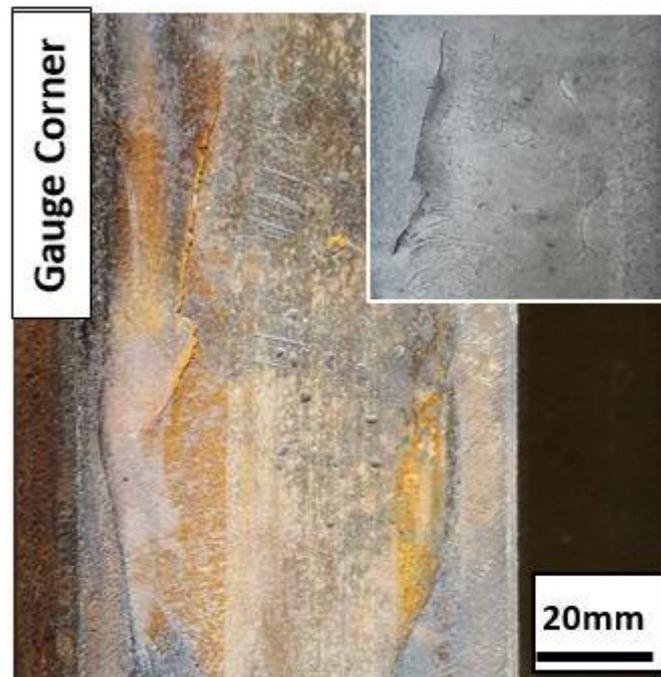
Figure 78: Sample 2. Inset) magnetic particle inspection image (MPI) to show the cracks

Unique features in Sample 2 are possibly because of traction control being used by the rolling stock on this line. Traction control has been linked to Squat development, or more specifically, thermal damage of the rail [9]. With just under 11MGT of traffic, initial thoughts were that it is very likely a Stud rather than a Squat. Studs are detectable from ~10MGT whereas Squats are detectable from ~40MGT. The surface also feels smooth to touch unlike Squats, which feel rough to touch in one direction due to the accumulated strain [4]. There was also no noticeable corrugation on the 1m section of rail, which would add to the dynamic loading and is often found in the vicinity of Squats. However, two V-shaped cracks can be made out and there is a crack along where the ridge should be. The two V-shaped cracks correspond to what appears to be two defects, one deeper than the other, that have then developed into one large defect.

4.4.3 Sample 3

Sample 3 is from a broken rail of unknown MGT, which failed due to a transverse defect, within a Squat defect, shown in full in Figure 79. Most of the Squat suffered extensive damage during the break, removing surface evidence of the expected V-shaped crack. There was not much information to be gained from the shelled lobe either. It would have been ideal to know if the downward branching crack that broke the rail had another crack above it like Sample 1 or was a single crack plane. However, the unbroken lobe of the Squat was still available to view. The second lobe was very close to shelling and so presented the opportunity to remove the almost separated plate of surface material from the top of the defect. This allows the inspection of the crack plane below the Squat. This plane was covered in oxide, requiring dry ice blasting to remove the oxide. The plastic flow shown by the folded material right next to the shelled section makes this most likely to be a Squat, although it is unclear whether this folding occurred before or during the rail failure. The significance of this sample is the presence of the definite transverse crack. However, information gathered from the cracks will be very limited due to the oxidisation, crack face rubbing and violent break removing most information.
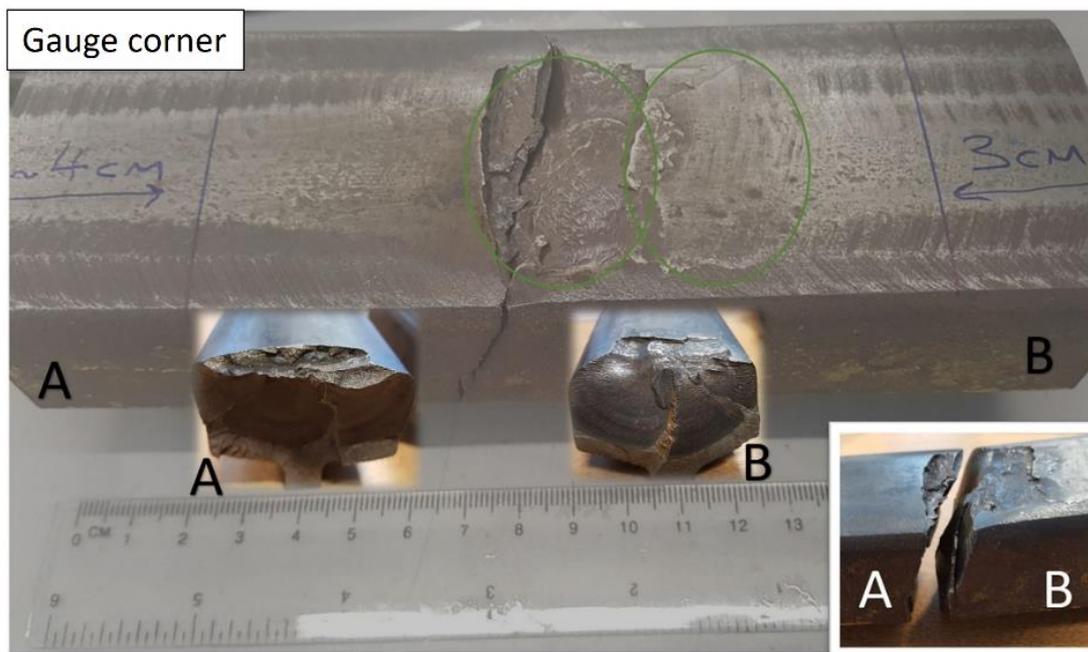
Figure 79: Sample 3 in its entirety split into pieces A and B. The lobes of the Squat are circled. The figure shows how the crack initially grew at a shallow angle (the shelled region) before taking a drastic change of direction and heading down through the rail. Future references to Sample 3 will be to piece B only as it contains the intact lobe of the defect.

4.4.4 Sample 4

Sample 4, shown in Figure 80, was identified by the track operator as a Squat on a weld. Following an initial ultrasound test after its removal, it was believed to be a combination of a Squat interacting with a weld defect. The longitudinal crack that runs along the edge of the sample shows that the Squat is slightly elongated, possibly linked to the higher train speeds [10]. The lack of deformation in the Squat and the traffic total of ~28MGT makes it likely that this is in fact a Stud, although a study has shown that 17% of Squats are found at welds [11].
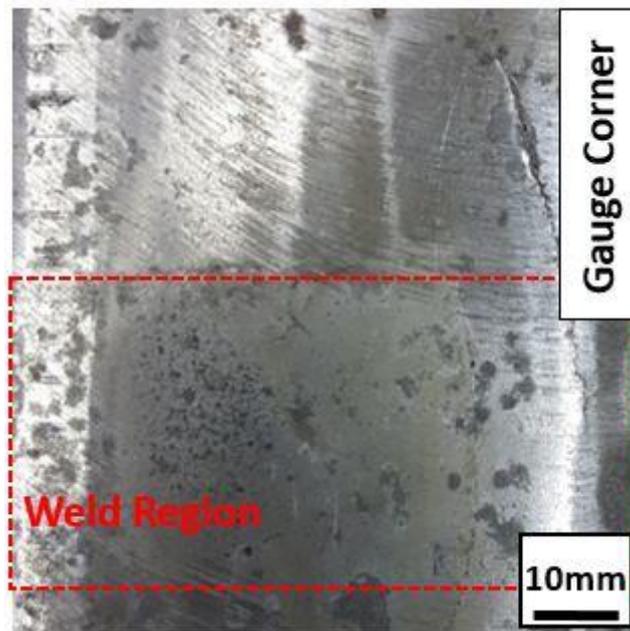
Figure 80: Sample 4 formed on the edge of a alumino-thermic weld.

As with Samples 1 and 2, Sample 4 has a large crack on the gauge corner side, but not the v-shaped crack that is usually expected of a Squat. The sample showed extensive grinding marks across the surface, not including the depressed region of the defect, which was in the weld region. Grinding marks were more extensive than the other samples, but this could simply be because grinding had occurred more recently. Grinding marks and other features that increase the roughness of a surface, such as crack mouths and the pits seen in Sample 2, can increase the subsurface stress by a factor of eight compared to smooth surface contact [12].

4.5 Results

The results of the IF microscope and X-ray µ-CT scanning are discussed in partnership to view the samples in a new way. The IF surface maps complement the poor near surface imaging inherent in µ-CT scans.

4.5.1 Sample 1

On Sample 1 there were quite a few differences in surface finish, such as the shiny running surface next to the dark corroded region of the black spots, as well as the cracked 'snakeskin' region. There are four obvious cracks on the surface, identified in Figure 81, with the scanned images of them are in Figure 82. The longitudinal and obtuse V-shaped crack are found in the other samples, but the L-shaped and Y-shaped cracks are not (shown in Figure 81). The Y-shaped crack has been seen before in a Squat sample reported on previously by Tata Steel [13] along with the 'snakeskin' surface damage (see Figure 83). In the Tata Steel report the snakeskin was also found on an area of the rail that did not contain a Squat. The dark spot within the V-shaped crack is believed to be a spalled piece of snakeskin as the snakeskin can be faintly seen close to the crack (Figure 82c). Snakeskin is shown more clearly in Chapter 6.
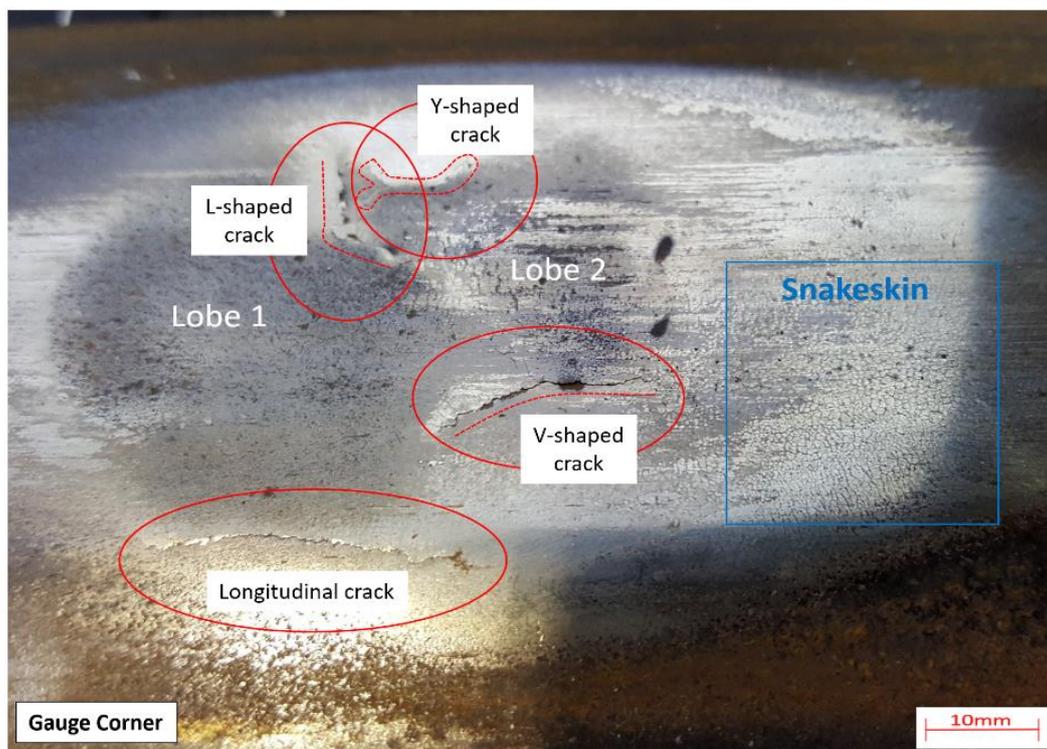
Figure 81: Sample 1 with the cracks circled in red and the snakeskin boxed in blue.
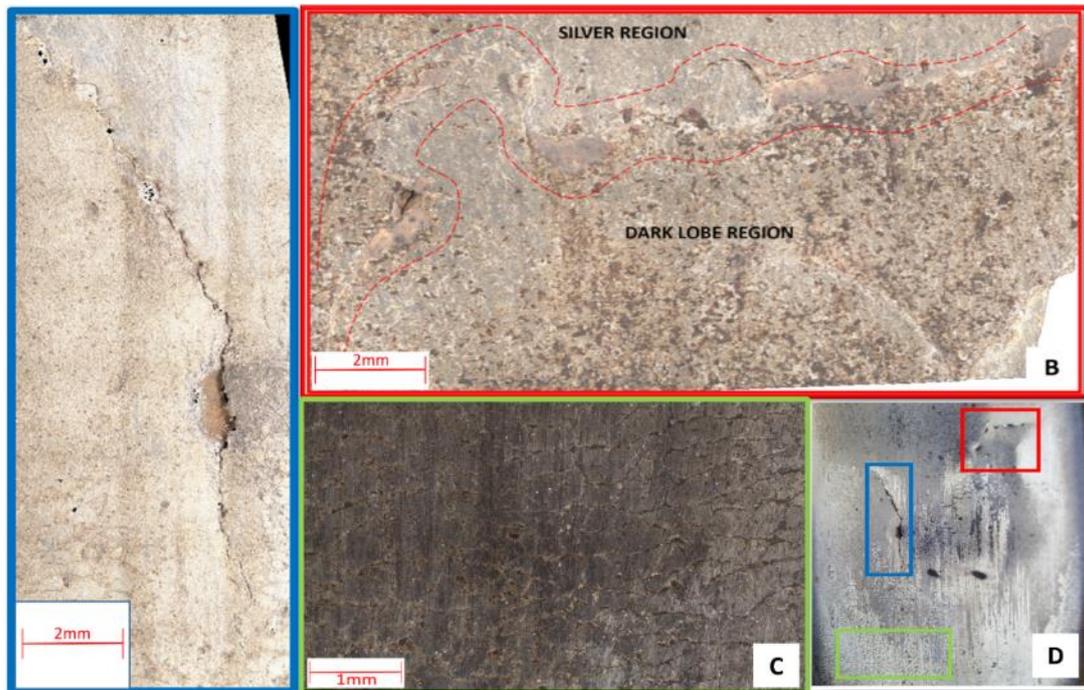
Figure 82: Some of the surface features imaged from Sample 1 a) The L-shaped crack with 'islands' that appear pink in the scan b) The V-shaped crack with a pit that penetrates into the substructure c) 'Snakeskin' microcracks across the surface of the rail just beyond the defect d) A colour coded overview showing where the images are taken from.

Figure 83 shows the Y-shaped cracks in Sample 1 and the Tata Steel report as surface and subsurface defects. As the Y-shaped crack was the only surface crack in the Tata Steel report it can be assumed that this crack forms first. The shallow depth of the defect in the report compared to Sample 1 also suggests that Sample 1 is much more developed. Both samples are believed to be from the same metro network. The presence of the two cracks suggests the possibility of one crack initiating subsurface whilst the other initiated on the surface. Using the X-ray scanned volume that Figure 83d was taken from, it can be seen that moving through the sample towards the field corner shows that the cracks stay separate, although they may have connected within the region that was worn away. Moving through the sample towards the gauge corner shows the two cracks connecting as a very shallow crack parallel to the surface that makes up the bottom portion of the Y shape (when viewing the Y as written here).
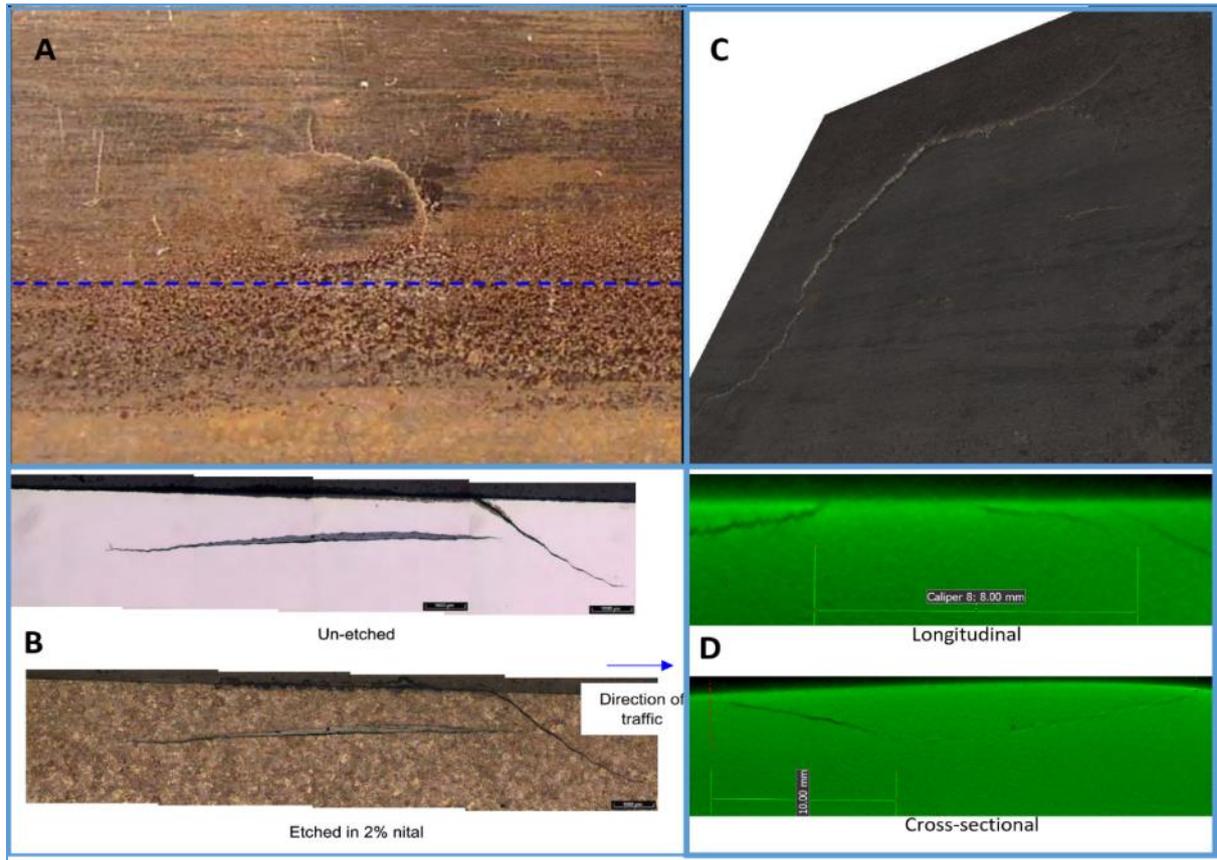
Figure 83: Comparison of features of a Squat from a Tata Steel report [13] *with scan results from sample 1. a) The surface of the Squat from the report, b) Longitudinal cross sections from the report, c) Surface scan from Alicona of sample 1, d) Longitudinal and cross sectional X-ray scans of sample 1 (blue dotted line shows where orthogonal slices cross).*

Comparing images from the report to the scans makes it highly likely that the initiation was in this area, and the branching of the longitudinal crack occurs within 4mm of the longitudinal slice in Figure 83d, moving towards the gauge corner. Given that the Y-shaped crack is closer to the field corner than the gauge corner suggests that a two point contact between the wheel and the rail may have been responsible. High surface temperatures occur with lower creepage and train speed when the contact patch is reduced like in a two point contact [14]. Sample 2 also contains a similar Y-shaped crack in a similar region but the clarity of the two cracks is not as clear, although a higher resolution scan of that area is planned for future work due to the branching of a possible transverse crack being in the same region. The L-shaped crack tracks the boundary between the dark lobe of the Squat and the shiny surface of the running band, it

also appears to contain islands of a different phase of steel, possibly martensite. This crack will be the subject of further investigations.

For each of the 3 samples scanned, a 3D model was initially created both in Avizo and in VGStudio MAX. The Avizo models were less accurate, more difficult to visualise and the software was more difficult to access, so progression was only made using VGStudios MAX and its lighter read-only counterpart, myVGL. The model of Sample 1 (Figure 84) was created in VGStudio MAX.
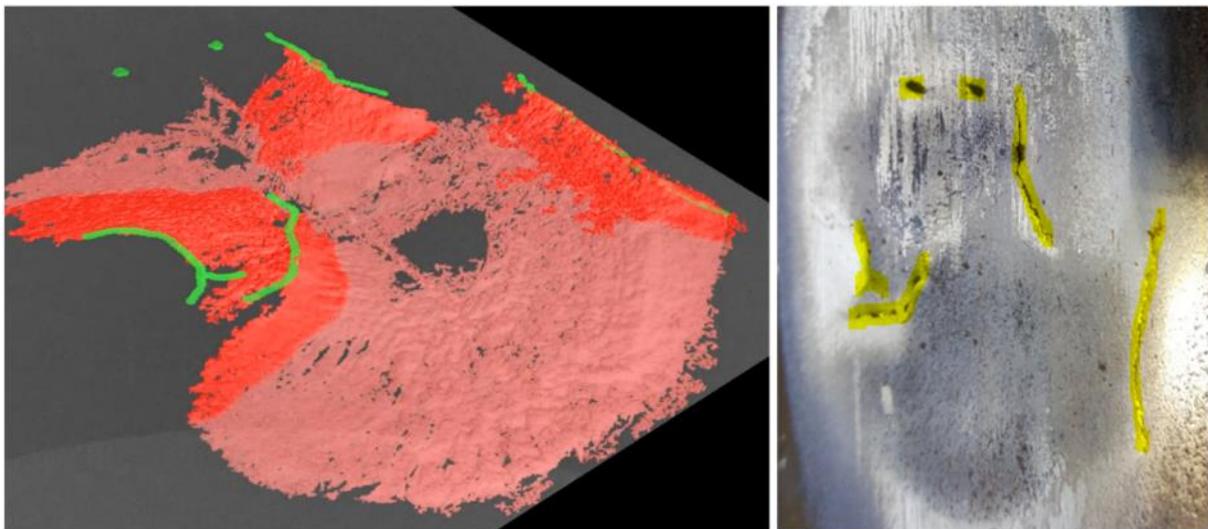
Figure 84: Left) A model of the crack plane in sample 1, created in VGStudio MAX. Right) a photographic reference of the surface cracks that are green in the model and yellow in the photograph.

Figure 84 shows that the basin of the crack plane with the hole in the centre lies directly beneath the larger dark corroded lobe. The other basin is not as developed and this is mirrored in the smaller size of the corresponding dark lobe near the two spots. The crack plane continues through the hole in the centre of the model: this region is undetectable by the automatic construction module, but the crack was discernible by eye. The crack plane also extends slightly further beyond the two green dots on the model but not in enough clarity to be automatically detected. Beneath the smaller basin is another crack plane that only showed up faintly in the scan.

Ultrasound also struggles to detect these lower cracks, which is part of the reason they are so potentially dangerous. In Figure 85 the lower branch was visible to the human eye and there were enough dark voxels to be able to determine its path so blue colouration was added to the model to show where manual additions had been made, as these are subject to user discretion whereas the automated function is repeatable by any user.
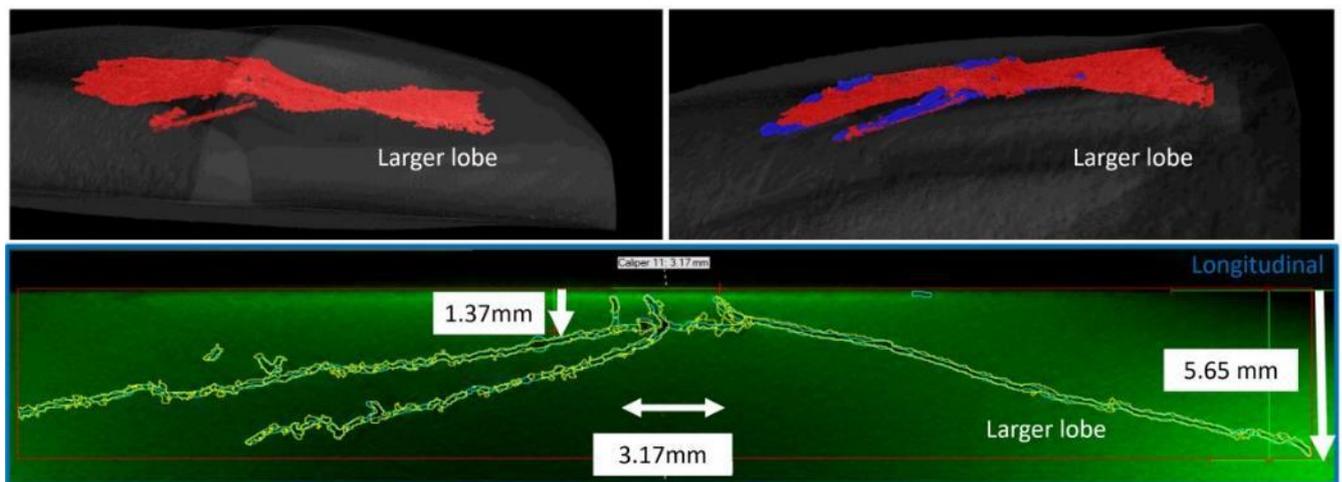
Figure 85: Upper) Models of Sample 1 showing the crack that branches downwards slightly, with the automatically detected sections in red and the manually added sections in blue. Lower) An orthoslice of the defect showing the lower crack after being manually highlighted.

4.5.2 Sample 2

The typical inverted V-shape that is expected in a longitudinal slice from a Squat or Stud is present in Sample 2. Much more damage was found further along the rail from the expected inverted V-shaped crack. This damage was due to multiple branches of cracks that were turning down into the rail and continuing to branch further, as it is common for the leading crack to do. This damage was more significant than that in the other scanned samples, especially when considering the mere 11MGT of traffic experienced by Sample 2. It could be said that the overall defect is made up of two interconnected defects, which could explain its size. This theory is supported by the presence of two V-shape cracks on the surface rather than the typical one. For this to be true, the deeper portion of the defect must have initiated subsurface.

Another detail that showed through the full construction of the models was that the 'O' shaped cracks found in the centre of the surface of Sample 2 was the location where the inverted v-

shaped crack had been truncated by wear (Figure 86). Sample 1 did not have this feature but Sample 3 had a surface breaking crack in the same region but of a different morphology.
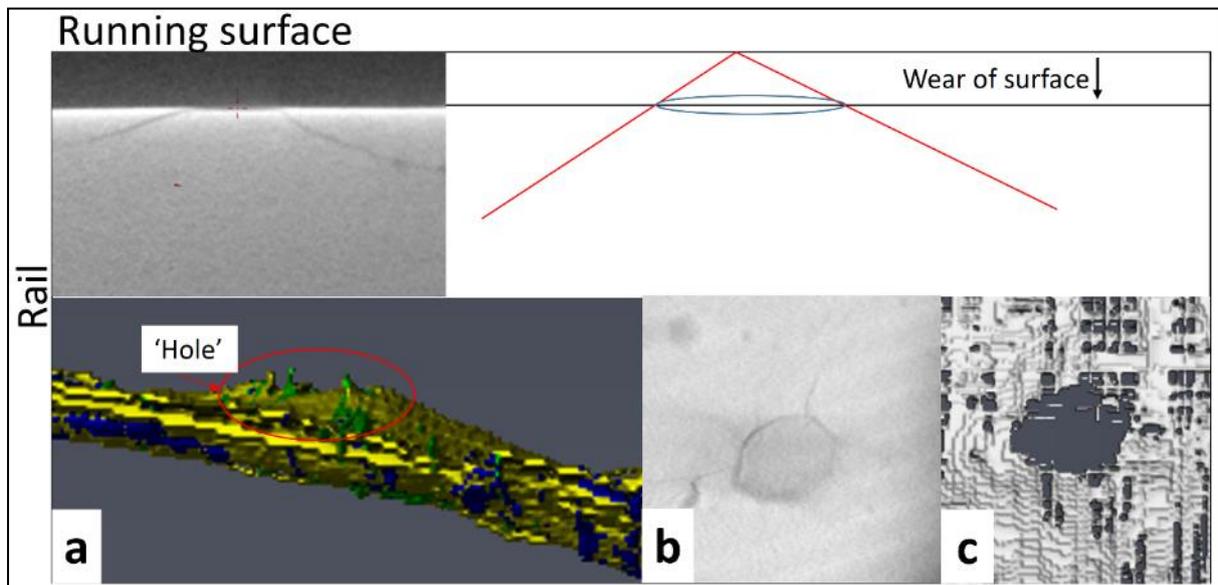
Figure 86: A diagram showing how wear truncates a pair of cracks to leave what is seen in the scan slice. Inset; a) The 'hole' from the longitudinal view as modelled in the Avizo model of sample 2, b) The original scan image of the surface of sample 2, c) The same view as (b) from the Avizo model.
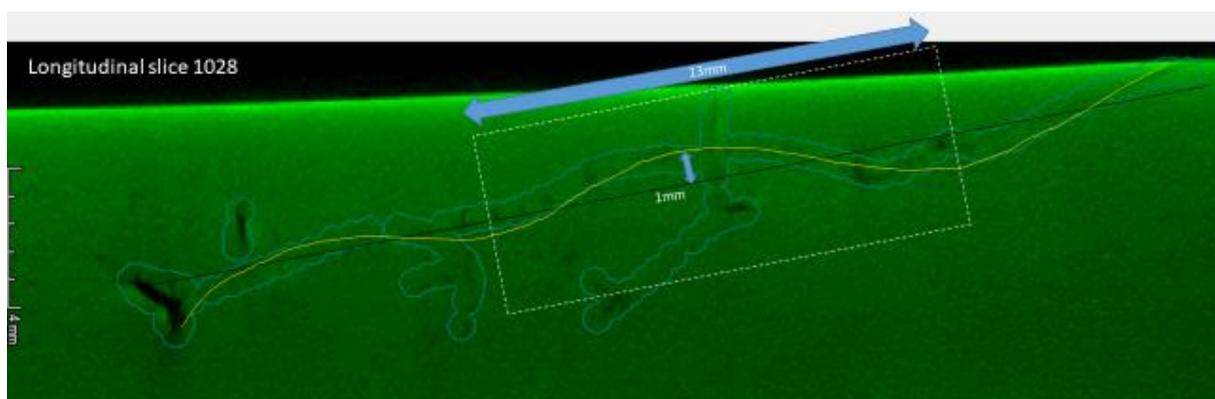
Figure 87: Longitudinal orthoslice of the downward branching section of sample 2. The crack follows a waveform pattern of roughly 1mm amplitude and 13mm wavelength.

Sample 2 had much more damage than expected. Figure 87 shows a longitudinal section from the scan, although the resolution of the scan is low due to the size of the sample, parts of the transverse crack can be discerned up to 8mm deep into the railhead. Cracks can be seen

146

reaching 10mm deep in other regions of lower clarity. The waveform of the defect was noticed after the initial crude model was built, though this observation is not pursued further within the scope of this work. The automated crack detection method developed for this work was used to build the model for Sample 1, detailed in Chapter 3. This method was not successful for the full defect scan of Sample 2 (Figure 88). Therefore, a higher resolution scan was needed in the downward branching region of the defect. The aim was to identify the region that caused the crack to start growing into what could become a transverse defect.
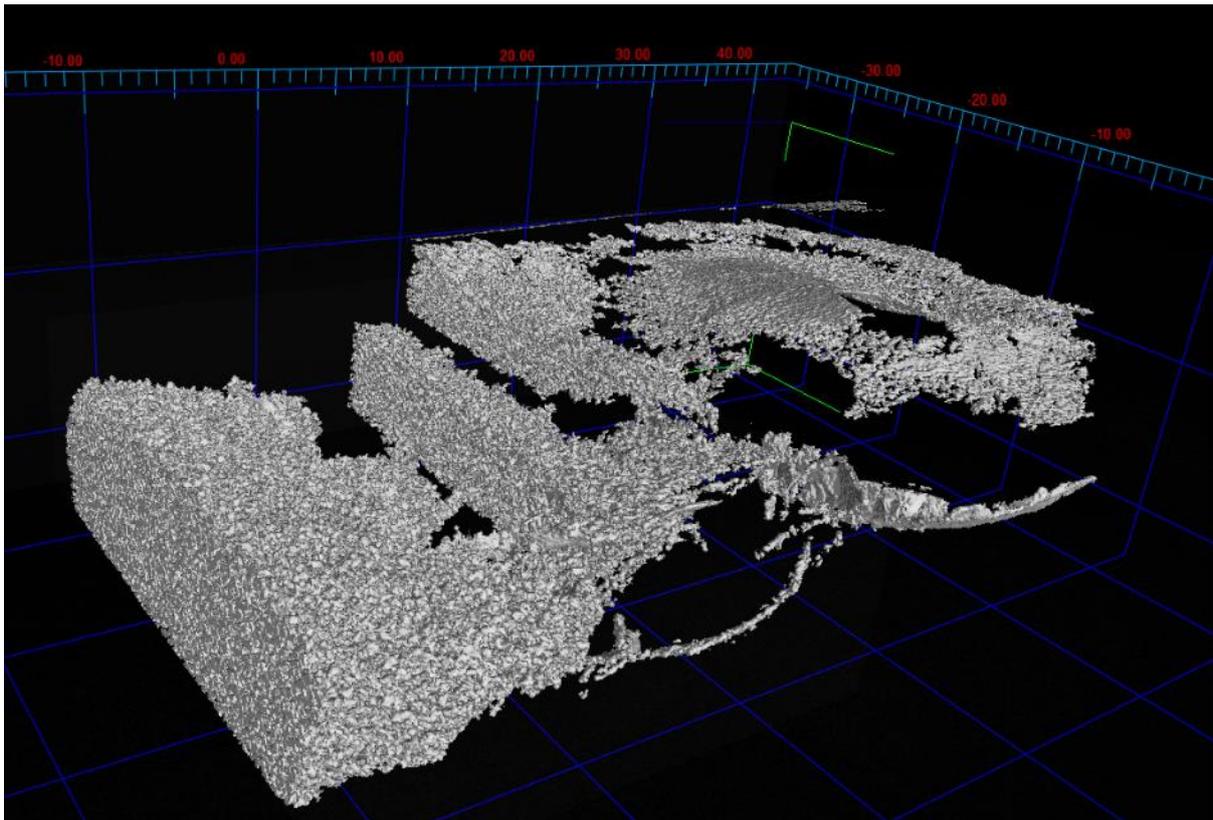
Figure 88: The result of attempting to use the automated crack detection method on Sample 2. Although the top right portion shows detection of some of the main crack plane, distinguishing between the many branching cracks was very difficult. The bottom left portion of the structure was the subject of the higher resolution scan.

4.5.2.1 Higher resolution scan of Sample 2

The higher resolution scan was performed using the same scanner as the full defect, but having the same size detector for a significantly smaller sample provides more voxels per unit area of the sample, providing the higher resolution and clearer crack features. This made the automated

feature much more effective and provided crack plane models that even shared the texture of crack planes. This contrast in results between scan of larger objects and smaller objects is why there is a limit to how large of a defect can be scanned, which is where this work pushes the boundaries of that limit as technology and techniques develop.

Intermittent shadows, which may or may not have made up a continuous crack in the full defect scan, were clarified in the detailed scan as being multiple crack planes branching out like the roots of a plant. Figure 89 shows the model with the main three crack planes showing the upper crack plane in green with two branching downwards in yellow and blue. The two separate branching cracks were hard to discern initially as they link together later in their growth. One side of the crack model was very different in its general shape to the rest and the deciding factor in this dilemma was the discovery of the downward cracks origins. Their origins, or initiation sites, are discussed in more depth along with their causes in Chapter 6. Various crack planes were left out of Figure 89 for clarity including surfacing cracks, but the extent of the branching can be seen in the semi-transparent model in Figure 90.
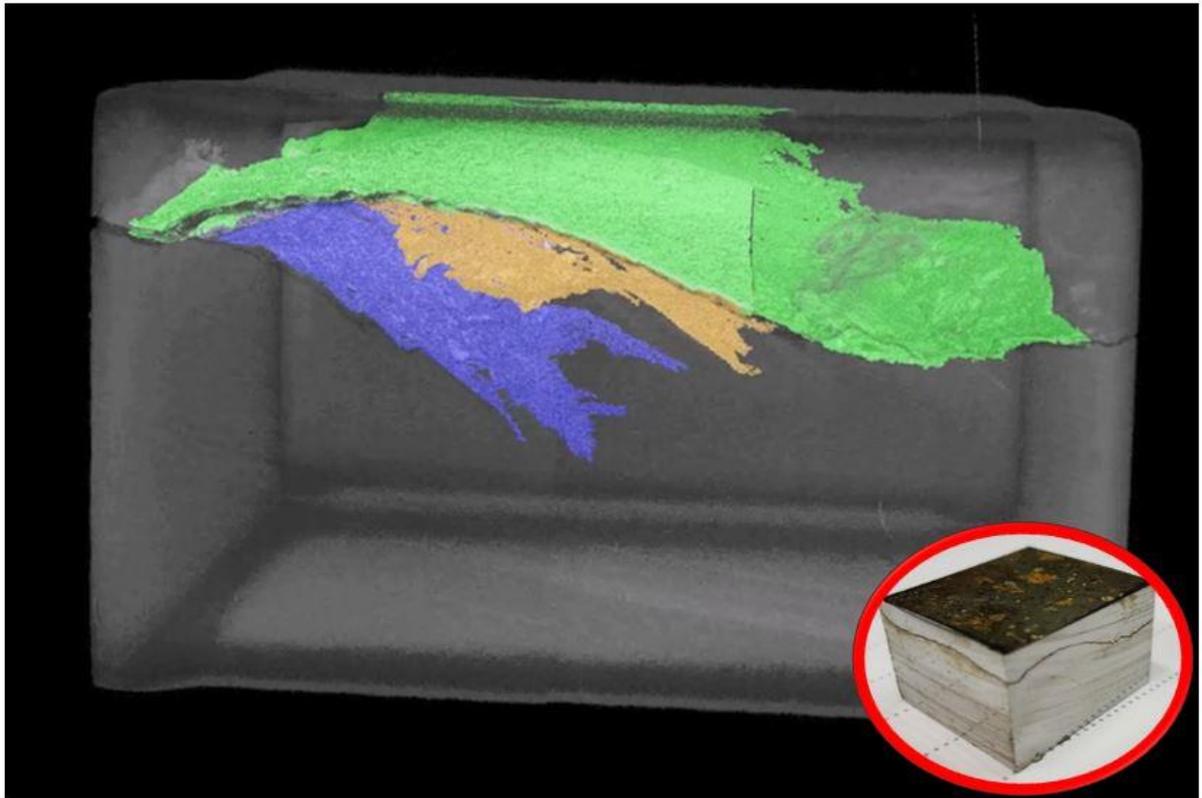
Figure 89: A simplified model created from the high-resolution scan of the region of interest within Sample 2. Inset: the region of interest cut from the defect showing the curved crack that is green in the model.

Figure 90: The same model as Figure 89 but showing all cracks to demonstrate the extent of the crack network that resembles a lightening pattern in the way the crack propagate.

The blue downward growing crack developed at quite a steep angle to the surface, to the extent that it was highly likely to have become a rail break if it had not been removed as diligently as it was. The cause of this potentially disastrous transverse defect became a key focus of this work, as solving the mystery of why some cracks grow downwards whilst others resurface and shell could greatly reduce the potential risk of these defects. Chapter 5 microscopically investigates the paths that cracks take through the microstructure and why they branch. The origin of the downward branching cracks is covered more in Chapter 6.

Isolating the main crack plane (green in Figure 89) and applying a plane parallel to the surface provided a sense of depth, improving the visualisation of the cracks topography. It showed that as well as growing longitudinally deeper into the rail, it grew upwards to the right, and probably left to create the surface cracks seen inset of Figure 91. This is no surprise but it is unclear if the defect began subsurface and grew up towards the edges or if it began on an edge and grew down into the rail.
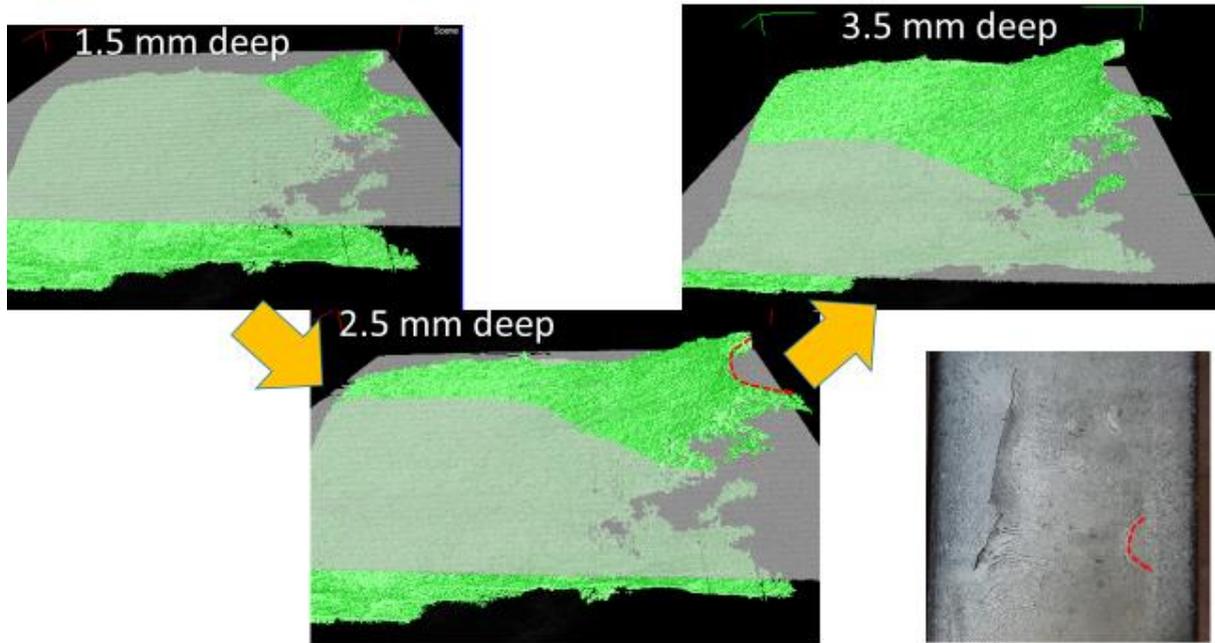
Figure 91: The main crack plane morphology for the high-resolution scan of Sample 2. Measurements are the depth below the highest point of the running surface that the plane is set at. Inset is a photo of the whole defect showing where the defect reaches the surface creating longitudinal cracks.

Taking a longitudinal view of the crack plane, from right to left of the perspective in Figure 91, shows that where the crack reaches the surface on the right (marked by red dashes), the morphology is erratic. If the crack was well developed and was resurfacing as the last stage of the crack growth it would expect to be a fairly smooth planar shape, like the deeper regions of the crack in Figure 91. Instead, it consists of a cliff-like feature (Figure 92) formed from a vertical crack with horizontal cracks branching from its base and peak. The cracks branching from the vertical crack are quite wide suggesting they are from earlier in the defect's development. It is more likely that the vertical crack formed first and then cracks grew horizontally away from it, rather than the crack took two sharp turns. The vertical crack is investigated using metallurgical techniques in Chapter 5.
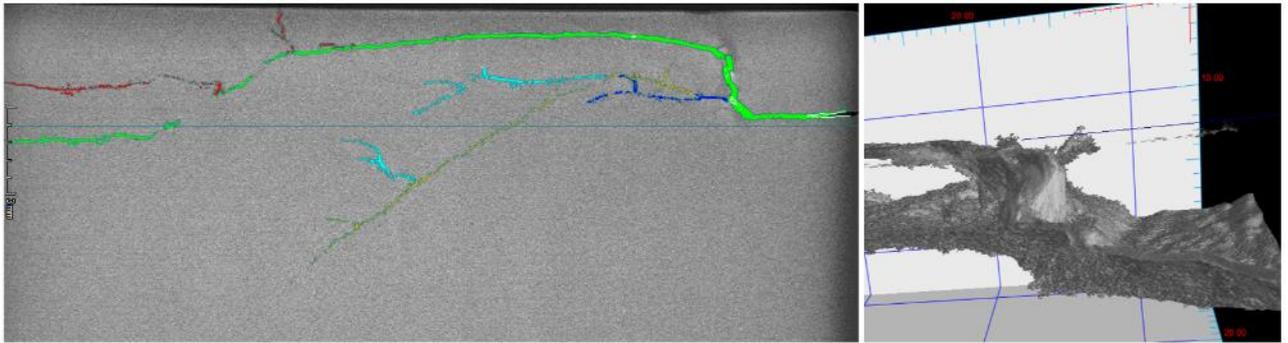
Figure 92: The cliff-like feature made up of a vertical crack with horizontal cracks branching from it. Left) A 2D orthoslice taken from the post-processed volume of the high-resolution scan (scale bar = 3mm). Right) The 'cliff' isolated from the 3D model (grid is in mm).

A second potential initiation site was also identified. This site appeared as a void in both the scan, optical and SEM images (Figure 93). This was identified as a crucial feature as it is where the transverse defect, that could have broken the rail, initiated. The reasons for why the void was believed to be an initiation site, how it developed and its connection with the transverse defect are discussed in depth in Chapter 6.
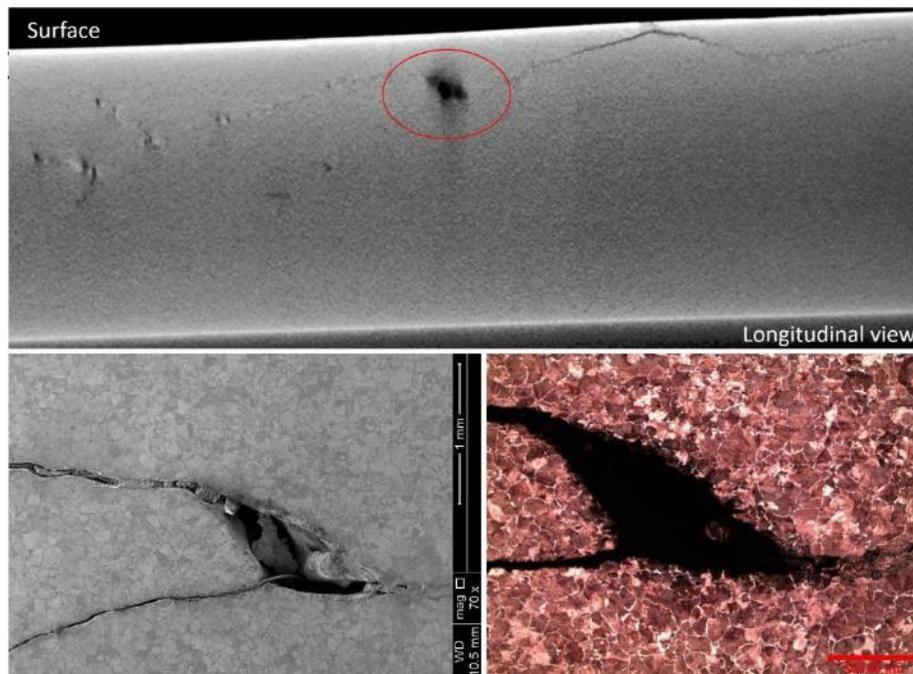
Figure 93: Above) The void (circled) from the CT volume of Sample 2 creating a large shadow. Below) SEM and optical images of the void showing the transverse defect growing from the bottom left corner of the void.

4.5.3   Sample 3

Due to the in-service break, the results obtainable from Sample 3 were limited compared to the other samples. Its inclusion in the comparison was due to the known transverse defect, though a CT scan was not deemed necessary due to the open crack planes. The undamaged lobe of the Squat was still intact and it was possible to remove the upper plate using a cross-sectional cut between two of the longitudinal surface breaking cracks. This revealed a structure similar to the models of the scans along with a discrete vertical crack that broke the surface as a wide crack and penetrated slightly deeper into the rail as seen in Figure 94. The scan in Figure 94c shows the surface area before it was cut, which includes features that appear as two white cracks in the scan due to the information being lost within the crack rather than being captured by the detector. These wide crack mouths are believed to a remnant of the initiation site. Judging by the amount of folded over material just next to the 'white holes' there were high contact forces in this area. The material may have flowed from the ridge that often occurs between the two lobes of a Squat as this ridge experiences very high contact forces. Micrographs of this plastic flow are shown in Chapter 5. Figure 94e shows what appears to be a small void in the crack in area highlighted by the orange box in Figure 94b. This area was scanned (Figure 94f) and the morphology suggests that an isolated region of material ('island') caused by a branching, then converging crack. These 'islands' are found throughout the crack networks of all scanned defects and cause shadows to appear as shown in Figure 74. The microstructure of Sample 3 gave strong indications as to the cause of the defect development and ultimate rail break. This is discussed in Chapter 5.
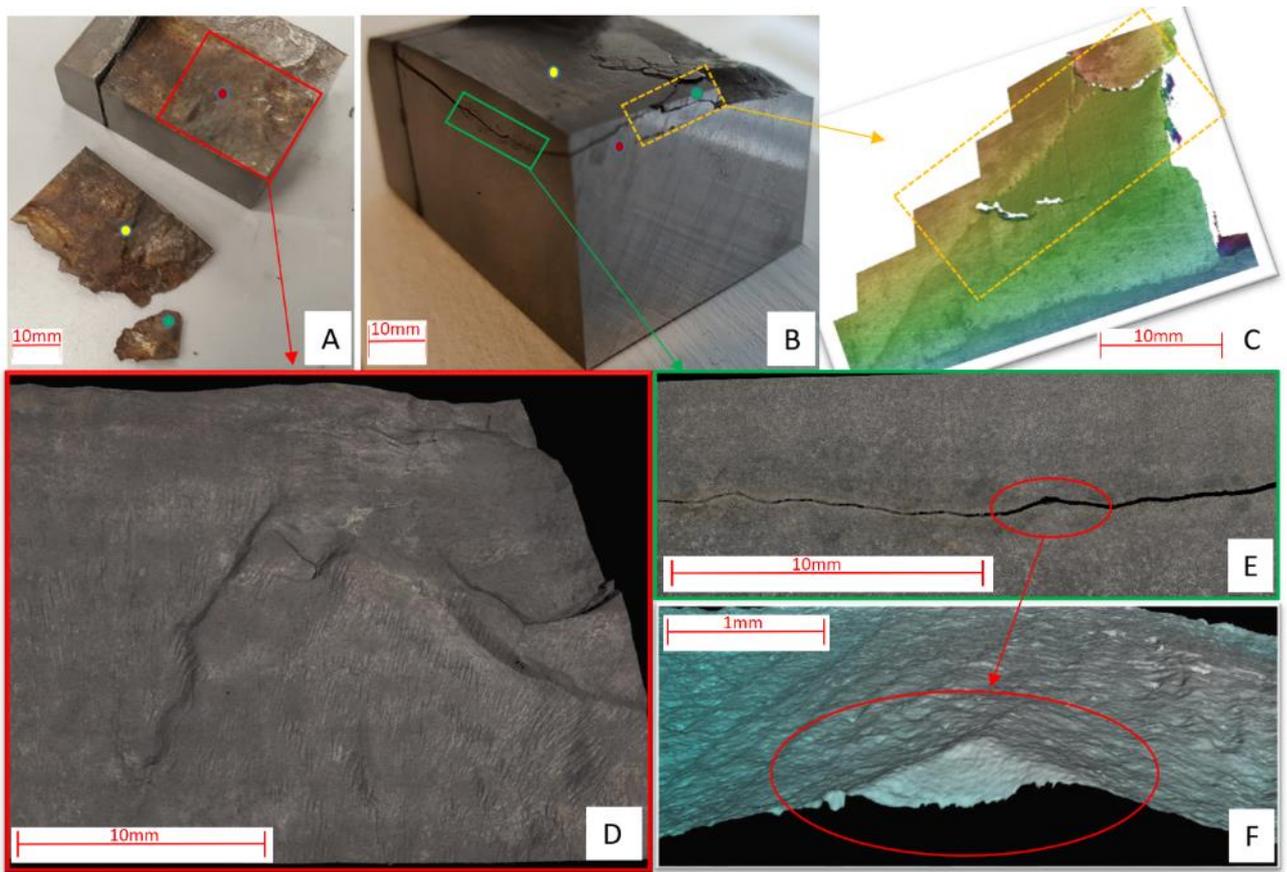
Figure 94: The fracture surface under the remaining second lobe/ spot of the Squat in sample 3. A) The fracture surface immediately after removing the upper surface. B) The sample before the removal of the surface showing the two cracks propagating down into the rail from a possible initiation point. C) The 3D model of the possible initiation area taken before cutting with the orange box showing the common area D) The 3D model of the crack plane shown in A. E) An enlarged view of the green box in B showing where material has spalled away during cutting. F) A 3D model of the 'void' in E.

4.5.4   Sample 4

Sample 4 had much less damage than expected, based on the results of the ultrasound tests. The ultrasound gave reason to expect a weld defect as well as the surface defect. However, there was no sign of a weld defect in the upper 26mm of the head and the welds heat affected zone (HAZ) did not seem to have much influence on the crack planes overall morphology. The crack plane was extensive but travelled at a fairly constant angle with few deviations other than

the C-shaped hole (Figure 86a) where it probably initiated. This initiation is on the boundary between the weld and the rail.
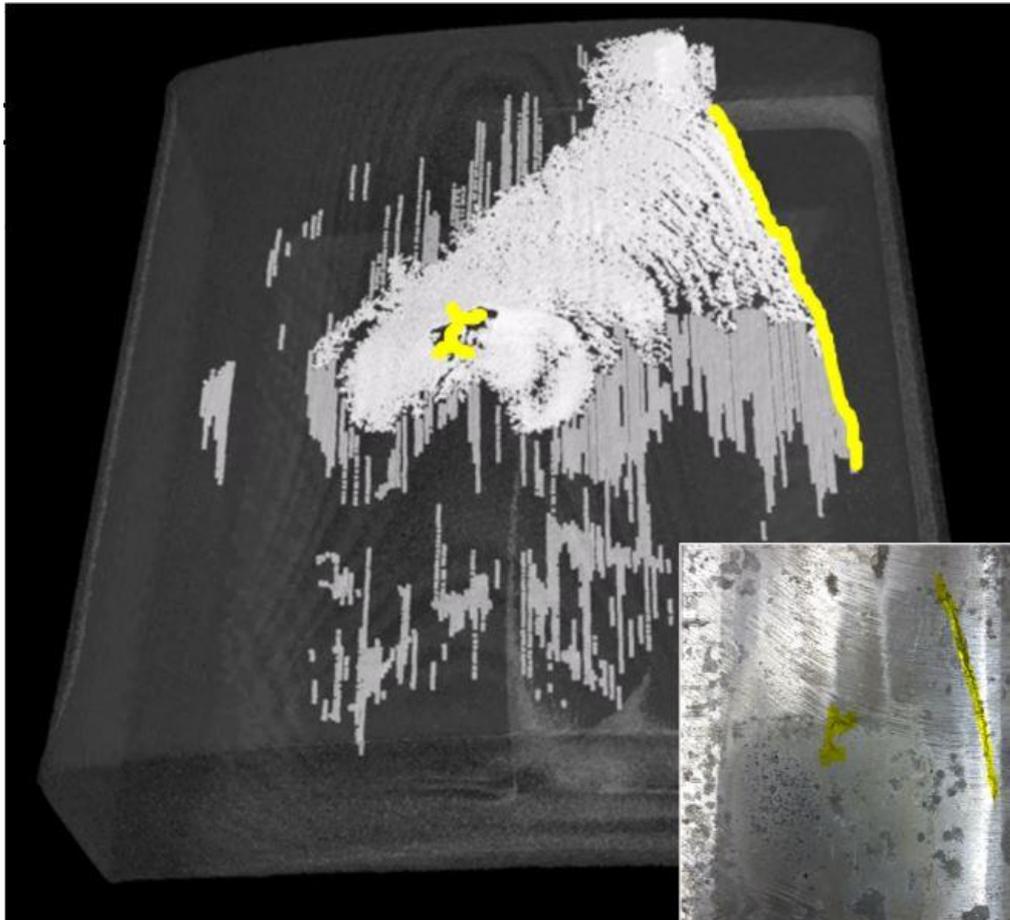
Figure 95: The model of Sample 4. Common features with the inset photo are highlighted.

The clarity of the crack plane in the Sample 4 model was lower than the other two samples, with only the bright white area being automatically detectable. Other areas were noticeable and so were added manually (in dull white), but only the parts that were dark enough to be considered cracks. Reasons for the lower clarity may have been because this sample was 26mm deep compared to the 15mm of Samples 1 and 2 but the weld may have been a factor too. If the areas found were interpolated it would make quite a simple, non-branching crack plane that begins in the centre and breaks the surface at the edge. This simple crack plane morphology shows during traditional sample preparation and is covered in Chapter 5. This makes Sample 4 the most benign of all of the samples as it was the only crack network that did not branch.

4.6 Discussion and comparison

The scans of the samples were successful, with a reasonable level of detail being captured and any artefacts being explainable. The volumetric approach to investigating the defects informed further decisions about where to investigate within the defects to try to understand their causes and differences. The IF surface scans/maps complimented the µ-CT scans well, making up for the less defined regions at the objects surface. The use of more traditional methods of inspection were useful to validate the scans and allowed the viewing of microstructure, which the scans could not do.

Looking at the surface of the defects, although the four defects are from different countries and track environments, they all share a longitudinal crack along the edge of the running band, though Sample 3 is harder to be certain about due to its shelled surface. This shows how heavily developed they are as this is where the crack plane finally surfaces. The kink in the longitudinal crack of Sample 2 is because there are two defects very close together that merge and break the field side of the running band as two separate, but similarly located V-shaped cracks.

Both metro samples (1 and 2) had a Y-shaped crack that was also found in the literature [13]. This Y-crack is believed to be the first to develop of the various cracks found in Samples 1 and 2 and is on the field corner half of the railhead rather than towards the gauge corner. The cooling that occurs after excessive wheel slip is higher outside the contact patch and considering the damage from thermal input is cause by tensile stress during cooling [15], this could explain why early cracks such as the Y-shaped crack occur easier on the edges of the contact patch. The presence of two unconnected cracks under the Y-shaped crack brings in the possibility of two initiations, one subsurface and one close to or on the surface. It is likely that one appeared first then the other as a result, as the leading-trailing crack theory in the literature mentions.

Samples 2 and 4 have a 'hole' in the running band where the conical crack has been truncated as the tip is worn down. This hole had a loose piece of material in the centre that was lost during cutting, so it was beneficial to have preserved this information in the scans. Sample 1 did not have this feature but Sample 3 had a surface breaking crack in the same region, but of a different morphology.

The next few paragraphs will discuss each sample in turn.

Sample 1 was initially believed to be a Stud because of its fairly smooth surface and the likelihood of a bad wheel slip event being from an incline. However, the presence of lipping on the edge of the rail (i.e. plastic flow) brought this into question. Sample 1 is lacking the 'hole' that was noted in Sample 2 and 4 despite having significantly more surface damage than any of the other three samples (or the reported Squat that it was compared to in section 6.1). The lack of plastic flow on the surface, smoothness in both directions and the islands of WEL found make it highly likely that this is a Stud and that the lipping was not due to the defect.

Sample 2 was believed to be a Stud, primarily because of its very low MGT, smooth surface and lack of plastic deformation. Checks for lubricant penetration were not conducted on any of the samples as they were contaminated by cutting fluids whilst trying to expose the cracks. There were pits above the most damaged region with two pits being connected by a surface crack. Investigations into more samples that have this feature would be very beneficial as it may explain why the defect was so well developed considering there was so little traffic on that rail. This is based on the principle that a rough surface causes much higher subsurface stresses. This will be a topic of Chapter 6, which will show through further work that Sample 2 is actually a Grinding Induced Squat (GIS).

The wider crack in the V on Sample 3 is where the cracks meet near the surface making this a likely initiation point due to ductility exhaustion. Sample 3 had long inclusions running parallel to the cracks that caused the failure, so they are believed to have been involved in the failure process. However, the whole rail probably had these inclusions so there must have been an event that initiated it, but not enough is known about the rail or its location. The presence of so many inclusions suggests that the steel was from ingot casting rather than the cleaner method of continuous casting.

Contamination in the very upper part of the weld is believed to be responsible for the initiation of the defect in Sample 4. The contamination consists of just a few spheres/bubbles of unknown composition (probably gas or slag) very close to the surface. Cracks were observed between these spheres/bubbles.

Sample 4 had less crack branching than the metro or mixed traffic samples, possibly due to modern track construction and less traffic variation/better profile matching between the wheel and rail. Table 5 summarises the more comparable points made in the discussion.

Table 5: Summary of comparisons between defects

Feature	Defect 1	Defect 2	Defect 3	Defect 4
Surface cracks	Longitudinal V-shape Y-shape crack	Longitudinal V-shape Y-shape crack 'Hole' from truncated cracks	Longitudinal Probable V-shape	Longitudinal 'Hole' from truncated cracks
Undesirable microstructural content			Large inclusions	Weld contaminants
Other surface features	Snakeskin that left pits upon spalling	Pitting and obvious grinding marks	Folded material. Lost material due to transverse fracture	Weld zone and obvious grinding marks

The scans allow comparison of samples even if the same cuts were not made in sample preparation. Sample 1 was compared to another sample from the literature (Figure 83). Despite the compared sample not being possessed, it was still possible to find the same specific feature, even though it was only ~1mm wide. This was because with a CT volume, it is possible to move through a volume to find specific features/locations. A library of µ-CT scanned defects

158

would allow a more comprehensive comparison of different defects and would aid in categorising them. The details in the scan were verified by comparison to micrographs of the sample after micro-preparation. Figure 74 and Figure 75 show that the shadow highlighted by the red circle is real and the faint traces of the upper crack are real too. The uppermost crack was not obviously a crack unlike the lower two, so the scan did not depict all of the cracks with the same clarity. This is due to the size of the sample scanned, i.e. smaller samples are easier to image clearly.

The Curved Linear Detector Array (CLDA) could be used in place of the flat panel detector (FPD) to reduce scatter effects in future work. However, because the CLDA has an array of 1x2048 pixels rather than 2000x2000 (FPD), it takes a long time to scan large objects. The CLDA is typically more suited to smaller or flat samples.

Focus on the crack path through the microstructure would aid in adjusting the microstructure to resist crack growth. To fully understand why the crack path takes the route it does, the crack needs to be traced in detail from where the initiation begins through to a depth where the crack shows little or no sensitivity to microstructure at all. The need for this work to be carried out requires a good idea of the initiation site and a defect that is not too well developed as to have lost too much information due to the wear and deformation of the surface.

4.7 Conclusions

This Chapter was concerned with determining if X-ray µ-CT scanning can effectively capture the volume of a defect within a rail before destructive inspections begin, thereby increasing the techniques available to an investigator. The fairly accurate scans of defects much larger than any scanned before and of varying dimensions confirms that it is possible. Some detail is lost as the object becomes larger due to the limit of resolution, set by the detector, in this case 20000x2000 pixels. These volumes were then used in conjunction with inspection of the surface to look for clues as to how the defects were caused and what their variations are.

Each defect is different to each other, in cause and a little bit in morphology. This reinforces the idea that Squat type defects may have various initiation causes, as shown in the literature through the increased acceptance of the influence of thermal damage in Studs. Perhaps the less usual causes of inclusions and contaminations should be separated from the more common

mechanical Squat and thermal Stud. Defects from these 'other' causes may cause anomalies in big data attempts to map Squat type defect locations.

It is difficult to be sure how much the traffic type influences the differences seen as track information was scarce with most of the samples. Some common features amongst metro samples were notable though. Studs from the metro networks in France and the UK seem to share a Y-shaped crack feature, containing subsurface cracks that seem to have developed independently rather than branching from a single point. This region is possibly part of the early initiation as the comparison in Figure 83 shows, with the sample from literature only containing the Y-shaped crack and some 'snakeskin'. This crack being located on the very edge of the contact patch suggests possible initiation by a two-point contact/ hollow wheel.

Surface breaking 'holes' in Sample 2 and 4 are truncated inverted V-shaped structures typically found in both Squats and Studs, which indicate some wear has occurred since the crack initiated. Two of the defects had downward branching cracks that could have developed into rail breaking transverse defects. Both branches seem to grow away from the crack plane above them. Why this occurs did not get investigated in this work, but if well-developed cracks tend to grow apart rather than merging, then two cracks running in the direction could be the precursor to a transverse defect. Diverging and merging of early stage branches are seen as 'islands' in Figure 76.

Based on the four samples compared, Studs may be far more common than initially expected, especially on metro and high speed lines. The track operators initially identified all the four samples as Squats. The classification and probable cause of the four samples are shown in Table 6:

Table 6: Defect identity of samples examined

Sample	Identity	Cause
1	Stud	High contact stress
2	GIS (*explained in Chapter 6*)	High contact stress
3	Squat	Inclusions
4	Potentially Squat or Stud	Contaminated weld

The cause listed in Table 6 are in combination with wheel slip for the Stud defects. The wheel slip is the cause of the thermal damage needed to initiate the crack structure. Although believed to be a Stud due to its very low MGTs and some of its features, there are notable deep grinding marks over the defect in Sample 2. Their importance and role is investigated further, leading to the conclusion shown in the table, which will be explained in Chapter 6.

This is just a small number of defects examined compared to how many occur, but there are differences between them despite them appearing to be the same initially. The traffic experienced is highly likely to be a factor considering the load differences between light rail/metro and heavy haul/ freight. The surprise with this work was that one of the lightest axle loads produced a complex crack network, possibly due to a much lower natural wear rate. It did not grow anywhere near as deeply as Sample 3 but the ages of the two samples cannot be compared. Sample 3 is likely much older considering it seems to have been ingot cast.

There would be great benefit in creating more CT volumes of Squat and Stud samples at various stages of development, allowing detailed comparison of defects and aid in searches for specific features. It cannot be determined from this work if a CT scan can discern between a Squat and a Stud as all 3 samples that were scanned are believed to be Studs. The Studs were noted as two being from metro and one from high speed with the Squat being from a mixed traffic environment. It should be noted that the Squat is believed to have failed due to the material as well as the service environment, so it is difficult to conclude accurately where Squats appear compared to Studs just from these four samples.

Microstructural investigations are warranted to inspect some of the features noted in this Chapter. Hardness mapping will also be used to look for variation in hardness from the thermal expansion and contraction experienced by the surface region of the railhead. This, alongside documentation of any white etching layers, will help to support their identity as Studs. This will be presented in Chapter 5 with some focus on the junctions, 'islands' and voids (broken up islands) that caused the shadows in the scans.

Grinding marks have been noted as being present on multiple samples. More work will be conducted into the effects that surface imperfections such as grinding marks and pits have on Squat type defects. Links between surface features and the junctions, which resulted in

potentially downward growing cracks/transverse defects, are investigated further for samples 1 and 2, with the results presented in Chapter 6.

4.8 References

[1]   E. N. Landis and D. T. Keane, "X-ray microtomography," *Mater. Charact.*, vol. 61, no. 12, pp. 1305–1316, 2010.

[2]   M. Naeimi, Z. Li, Z. Qian,Y. Zhou, J. Wu, R. H. Petrov, J. Sietsma and R. Dollevoet, "Reconstruction of the rolling contact fatigue cracks in rails using X-ray computed tomography" *NDT E Int.*, vol. 92, no. April, pp. 199–212, 2017.

[3]   M. Steenbergen and R. Dollevoet, "On the mechanism of Squat formation on train rails – Part II : Growth," vol. 47, pp. 373–381, 2013.

[4]   S. L. Grassie, D. I. Fletcher, E. A. Gallardo Hernandez, and P. Summers, "Studs: a Squat-type defect in rails," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 226, no. 3, pp. 243–256, 2011.

[5]   J. F. Barrett and N. Keat, "Artifacts in CT: Recognition and Avoidance," *RadioGraphics*, vol. 24, no. 6, pp. 1679–1691, 2004.

[6]   NSW Transport RailCorp, "Rail Defects Handbook: TMC 226," pp. 1–83, 2012.

[7]   X. Deng, Z. Qian, Z. Li, and R. Dollevoet, "Investigation of the formation of corrugation-induced rail Squats based on extensive field monitoring," *Int. J. Fatigue*, vol. 112, no. February, pp. 94–105, 2018.

[8]   F. J. Franklin, I. Widiyarta, and A. Kapoor, "Computer simulation of wear and rolling contact fatigue," *Wear*, vol. 250, no. 251, pp. 949–955, 2001.

[9]   D. Scott, D. I. Fletcher, and B. J. Cardwell, "Simulation Study of Thermally Initiated Rail Defects," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 228, no. 2, pp. 113–127, 2012.

[10]   N. Gao, R. S. Dwyer-Joyce, and D. G. G. Grieve, "Disc machine testing to assess the life of surface-damaged railway track," *Proc. Inst. Mech. Eng. Part F-Journal Rail Rapid Transit*, vol. 215, no. 4, pp. 261–275, 2001.

[11]   Z. Li, X. Zhao, C. Esveld, R. Dollevoet, and M. Molodova, "An investigation into the causes of Squats-Correlation analysis and numerical modeling," *Wear*, vol. 265, no. 9–10, pp. 1349–1355, 2008.

[12]   A. Kapoor, F. J. Franklin, S. K. Wong, and M. Ishida, "Surface roughness and plastic flow in rail wheel contact," vol. 253, pp. 257–264, 2002.

[13]   H. M. Smith, "An examination of Squat defects in a 60E1 R260 rail removed from the Bussy Saint Georges test site in the Paris suburbs (RER line)," Rotherham, UK, 2011.

[14]   C. Bernsteiner, G. Muller, A. Meierhofer, K. Six, D. Kunstner, and P. Dietmaier, "Development of white etching layers on rails: simulations and experiments," *Wear*, vol. 366–367, pp. 116–122, 2016.

[15]   J. H. You, "Damage and fatigue crack growth of Eurofer steel first wall mock-up under cyclic heat flux loads. Part 2: Finite element analysis of damage evolution," *Fusion Eng. Des.*, vol. 89, no. 4, pp. 294–301, 2014.

Chapter 5 : Microstructures and crack networks in Squats

Verification of the images gained through μ-CT scanning and comparison of surface features were made in Chapter 4. Some micrographs were used to verify the scans, but the focus was on the scans and their accuracy. This Chapter covers the more traditional material investigation used in support of Chapter 4, being that these methods are well established and were used to verify the scans in Chapter 4. The techniques used are primarily optical and electron microscopy with some hardness indentation results. Analysis of the samples will begin at the surface and then work deeper into the rail.

This Chapter contributes to the next two aims of the project:

3. Are there differences between the crack morphologies of defects?
4. Is there a correlation between the microstructure and the defect?

5.1 Introduction

The scans have given a good insight into the 3D nature of the defects, but some desired details cannot be gathered from scanning. Looking at the microstructure gave an opportunity that the scans could not provide. To look for evidence of thermal input to help to distinguish between Squats and Studs. This is approached by first considering what can be classed as evidence of thermal input into a rail surface. Four factors are considered;

1. White Etching Layer (WEL): a phase change that can be due to thermal input
2. Grain refinement: a possible result of sufficient heat input followed by rapid cooling
3. Vertical surface cracks: seen in extreme heating and cooling of copper
4. Increased hardness: expected to vary if the thermal history of the rail varies along the surface and with depth.

After this a few other factors are investigated. Pitting was noticed on the surface of some of the samples and becomes important in Chapter 6. This is covered in this Chapter due to a connection with vertical surface cracks. Then factors other than those that seem to contribute

to the initiation of squats are discussed. The Chapter wraps up with the causes of some of the shadows seen in the μ-CT scans. They were found to be apparent small voids within the crack network and were found during the verification of the scan quality, so these are discussed in support of Chapter 4. These voids also link into work presented in Chapter 6.

Chapter 4 has already shown that the three defects scanned all showed a different morphology to each other, as visualised in Figure 96. A fifth sample will be introduced in Chapter 6 that also has a different morphology, but this is because it did not develop a trailing crack.
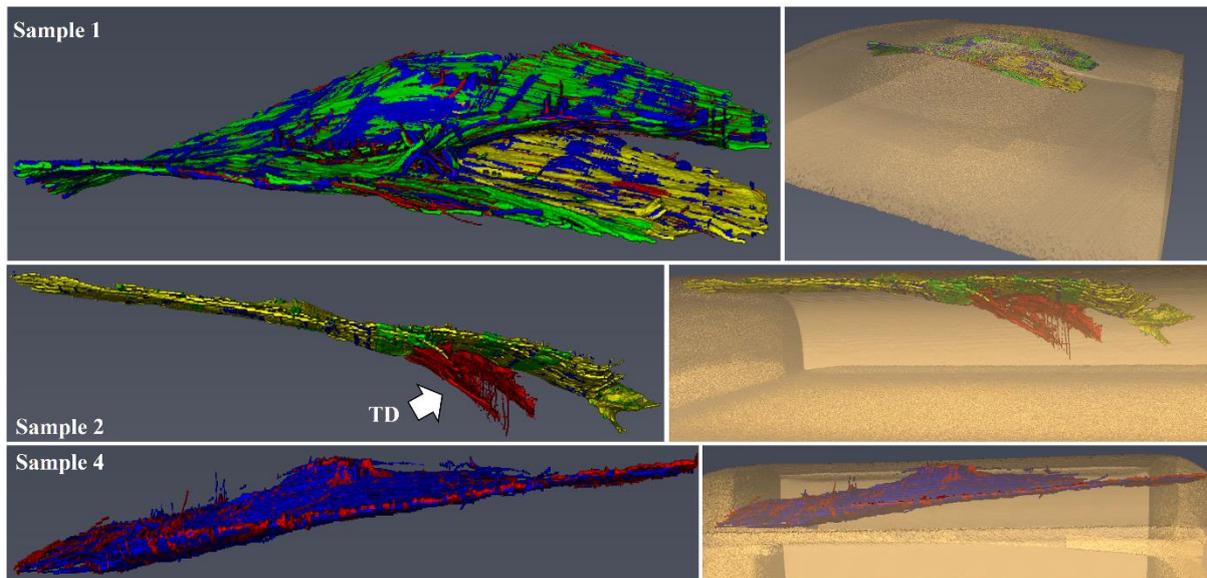
Figure 96: Early models of the crack planes in Samples 1, 2 and 4. Both as free crack networks and inset into the rail: created by manual drawing in Avizo. Colours are not indicative of anything important. The red section of the crack network in Sample 2 is the TD.

The difference in these morphologies could be due to the trains that run on them (weight, speed, suspension type, frequency, technology) and the profiles of the wheels (i.e. contact conditions). Unfortunately, this is hard to explore due to limited information about the trains that ran on the tracks that these samples came from. However, features within the the material and the crack growth through it could also be a strong factor, and this can be examined.

5.2 White Etching Layer (WEL)

WEL has been identified as a strong factor in the development of Squats, as discussed in Chapter 2. It is often present on used rail and due to its brittle nature, is prone to cracking. Although not a focal point of this work, investigating WEL adds to the understanding and cataloging of defects. The presence of WEL does not guarantee a defect, but WEL is often found where defects are.

5.2.1   Sample 1

WEL varied in its appearance; it appeared as both islands and long stretches, with distinct and diffuse boundaries, often within the same sample. An aerial view of the railhead showed long streaks of WEL and some are visible as contrast in the right image of Figure 97. This is probably partly explained by whether the sample is cut in parallel or perpendicular to a streak. The other aspect to consider is that WEL breaks away easily due to its brittle nature. Figure 98 shows one of the more unusual WEL areas found and is taken as a longitudinal slice from between the crack with a pit and the more central of the two black spots (refer to Figure 97). It shows a bright white, fairly continuous band of WEL that seems very brittle as it has spalled in several places. This is on top of another layer of WEL that covers 1.5mm of the 2.5mm long sample. This lower layer has a distinct boundary with the bright upper layer and a more diffuse boundary with the parent steel. The lower layer also seemed tempered by the upper layer, theorised from the darker shade of brown in its upper portion. The dark brown layer was not an etching effect, as it did not exist where the WEL had broken away slightly. It was nano-indented to see if softening had occurred due to the tempering. The results were inconclusive because the sample needed a light etch to reveal where the layers were, but this etch affected the results of the nano-indentation. This multi-layer WEL suggests that the sample suffered multiple heat inducing events of different magnitudes, some possibly incrementally building up the thicker layer.

Figure 97: The surface of Sample 1 with bright streaks of WEL orientated in the longitudinal direction.
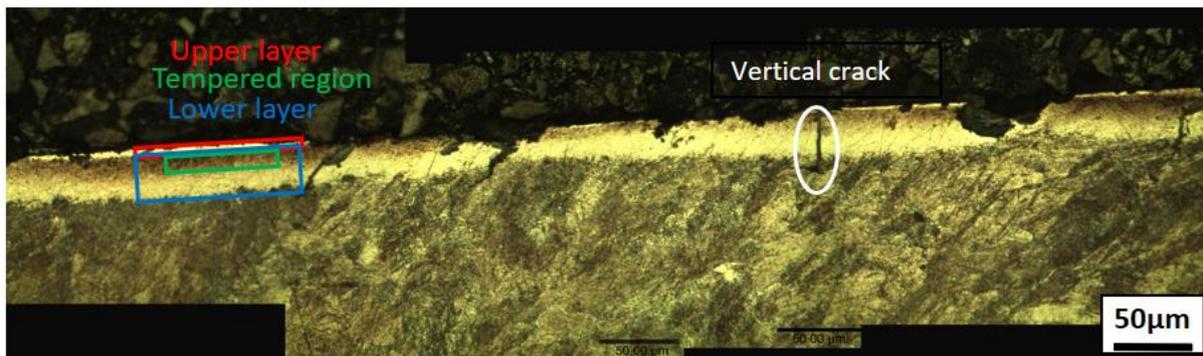
Figure 98: Two layers of WEL. A thin upper brighter WEL and lower duller WEL with a tempered upper surface. The upper layer is ~5 microns thick and the lower layer is 18-35 microns thick. Cracks grow vertically through the WEL and into the parent material.

Figure 99A shows how ferrite, that has been retained within the WEL, provides weaker pathways for cracks to grow along into the parent material. This shows that cracking simply due to the brittle nature of WEL is not the only explanation for the presence of cracks. Retained ferrite is also seen in Sample 2, so will be discussed more when Sample 2 is discussed. WEL is not always smooth either as shown in Figure 99B, which was taken from just next to the crack with the pit. The rough WEL suggests it is a remnant of grinding, reducing the contact area to the peaks of the WEL.
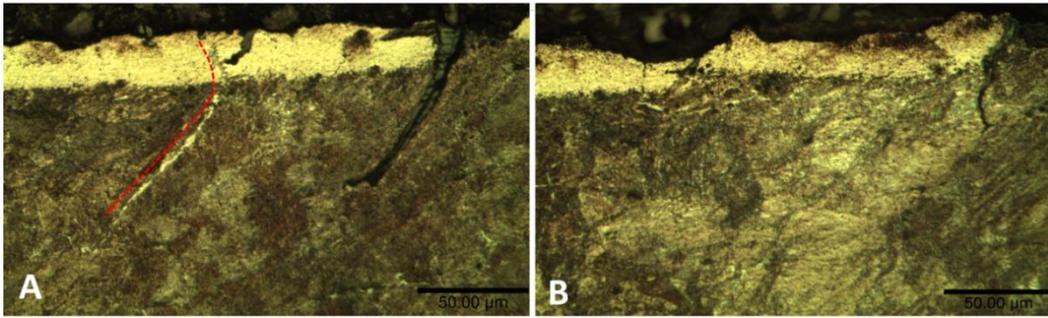
Figure 99: a) Ferrite pathways for cracks to develop along seen next to a crack that has already formed in the same orientation, shown by the dashed line; b) Uneven WEL from within the running band. A crack can be seen where the WEL ends.

5.2.2 Sample 2

WEL was found at a fairly consistent depth of 5-10 µm in some locations inside and outside of the defect. There were some locations where the thickness of the WEL varied even in a very small area, such as in Figure 100. Figure 100 also shows one region that was found very close to the field side of the running band, where the widened band started to narrow again. This region contained WEL with the ferrite from the grain boundary still visible, also seen in the first documentation of a Stud [1].
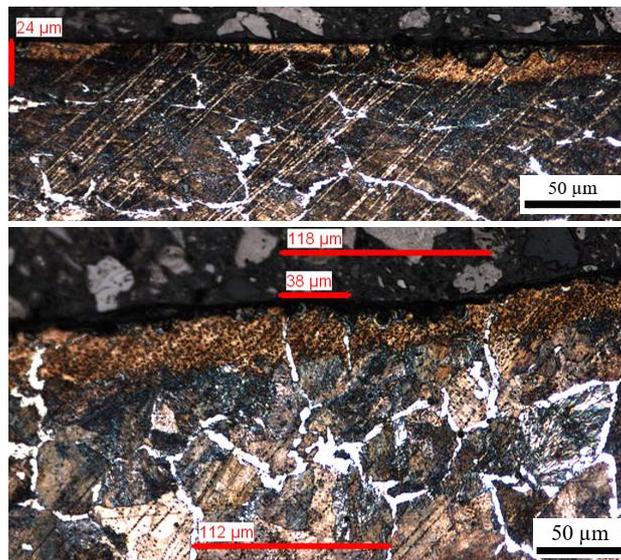
Figure 100: Above) WEL thickness varying from 1 - 24 microns thick in Sample 2. Below) Ferrite grain boundaries still visible within the WEL.

5.2.3   Sample 3

WEL was found in thick islands in cross sectional micrographs from the gauge corner of Sample 3 as shown in Figure 101. This is one of three discrete, but deep, patches of WEL from just before the defect / widening of the band. This is believed to be from rough grinding stones making discrete contact in the area.
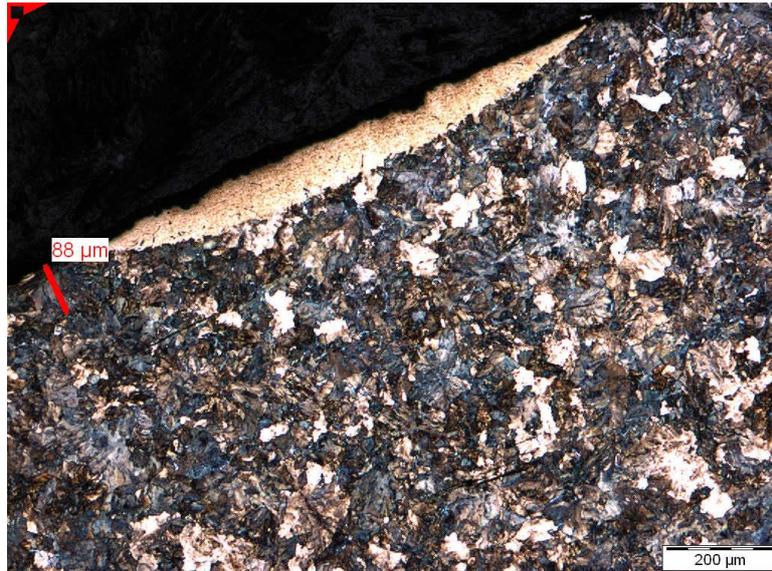
Figure 101: Thick white etching layers on the gauge corner of Sample 3.

5.2.4   Sample 4

WEL was found at a very consistent depth of 5-7 microns inside and outside the defect (Figure 102). The visible grinding marks on the surface make it probable that this WEL is due to the grinding process. The WEL inside the defect had a more distinct boundary with the parent material compared to outside the defect.

The WEL from within the defect was also unusual because there was a layer of WEL within the parent material (Figure 102a). Creating such a discrete area of WEL through heating and cooling seems unlikely, as the entire region from the surface down to the discrete layer of WEL within the parent material should have transformed, like in Figure 102b. This suggests that an event occurred to cause high enough stress to create a transformation at a very specific depth. Lubricated surfaces transfer the peak stress to a deeper region than unlubricated. Strain or deformation induced phase transformations from pearlite to martensite have been recreated in

pearlitic wire [2] and stainless steel. Therefore, coupling the two aforementioned factors suggests that the lower layer of WEL was formed due to high strain causing a crystallographic change from pearlite to martensite. This mechanism has been proposed previously [3]. Replicating this strain-induced martensite in a rail could unlock valuable details about the contact conditions that cause defects. This is not conducted within this work. It should be reiterated at this point that this sample was from a high speed line.
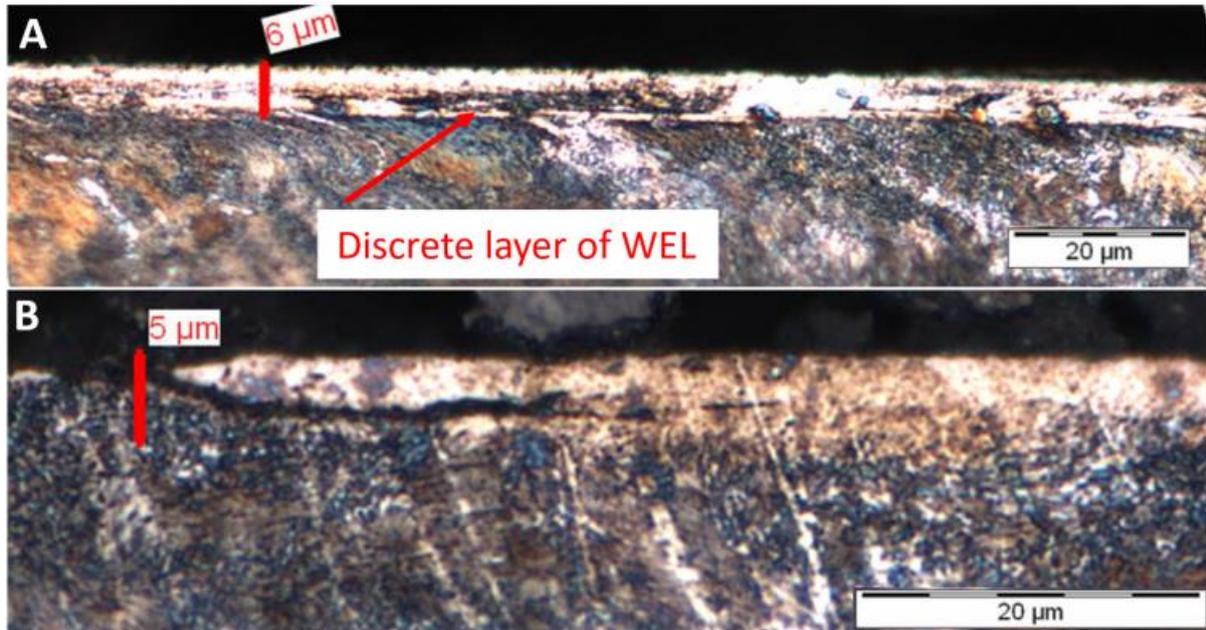
Figure 102: a) WEL from within the defect. b) WEL outside the defect.

5.3 Grain refinement

WEL can be caused by thermal input, i.e. the heating and rapid cooling of the surface of the steel. The effects of this heating is not always so obvious. Heat input can allow recrystallisation of the microstructure, meaning that small grains appear and start to grow, but as the material cools rapidly, these grains do not have time to grow and this creates a finer grain structure/ smaller grains. Although they may be subtle, these grains have different mechanical properties to the parent material. There is also the issue of the residual stress associated with rapid cooling. These residual stresses during cooling can be high enough to crack the steel.
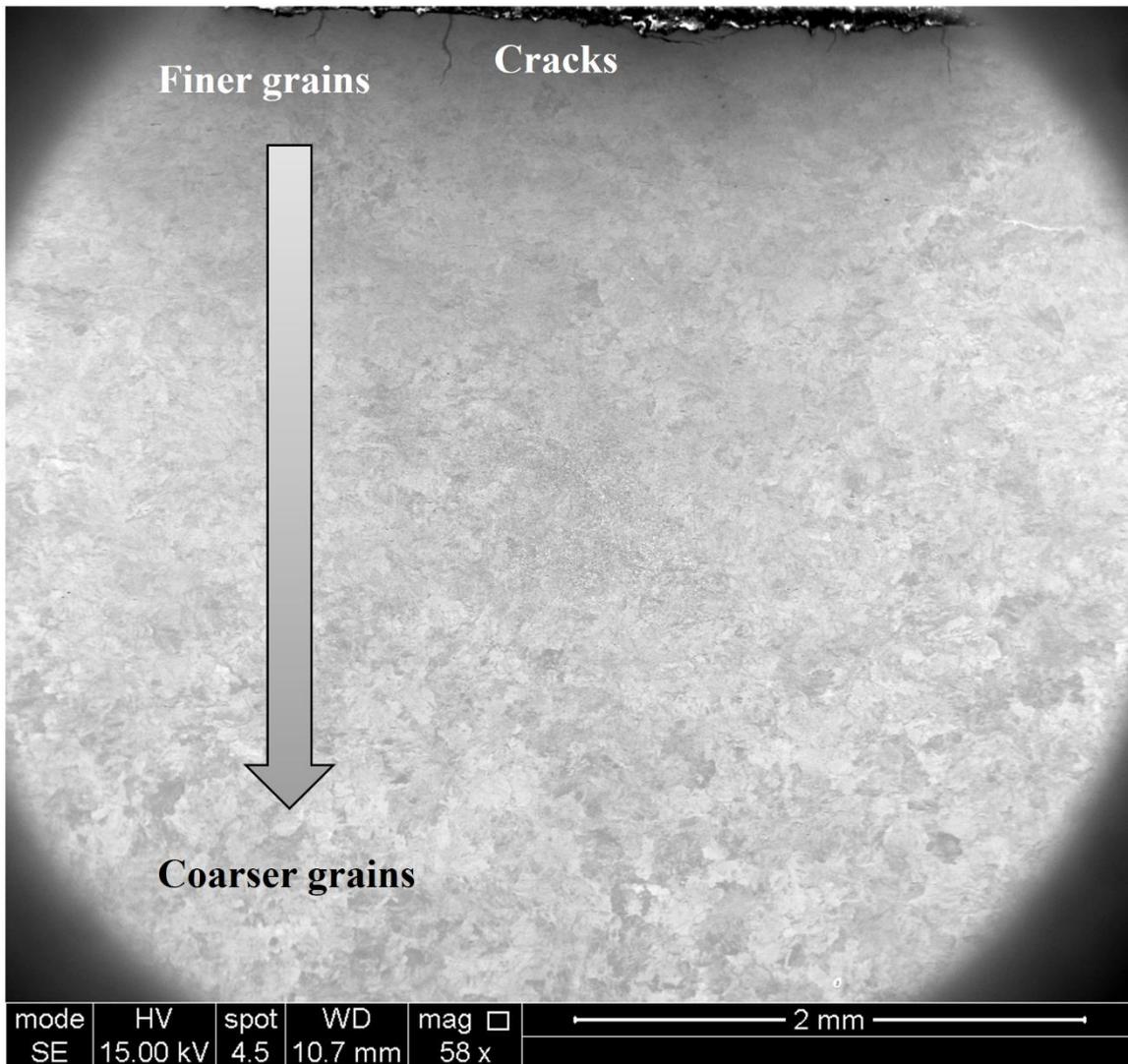
Figure 103: SEM image from the running band of Sample 1. Vertical cracks are at the surface, a finer structure is close to the surface and the original structure is at the bottom of the image.

5.4 Vertical surface cracks

Shallow RCF cracks are understood and are not surprising when found in RCF related defects. However, vertical cracks are less expected. Vertical cracks have been noticed throughout multiple samples. Generally, they are considered a result of thermal input that causes sudden heating followed by rapid cooling. However, vertical cracks also develop from pits and grinding marks (Figure 104). As seen in Figure 104 and Figure 106, some cracks begin as typical RCF cracks that are acutely angled with the running surface before developing into a vertical crack path. It could be that cracks that appear purely vertical may have begun as RCF

cracks, but the acute angled section has worn away. However, it is still also possible that vertical cracks are due to thermal shock. The presence of many vertical cracks grouped together and some grain refinement (Figure 103) is more suggestive of thermal shock whereas isolated vertical cracks could have developed from RCF cracks. Even if the cracks began as RCF, it is still curious that they develop into more of a vertical crack, rather than levelling out and running parallel to the surface as seen in Squats and shells.
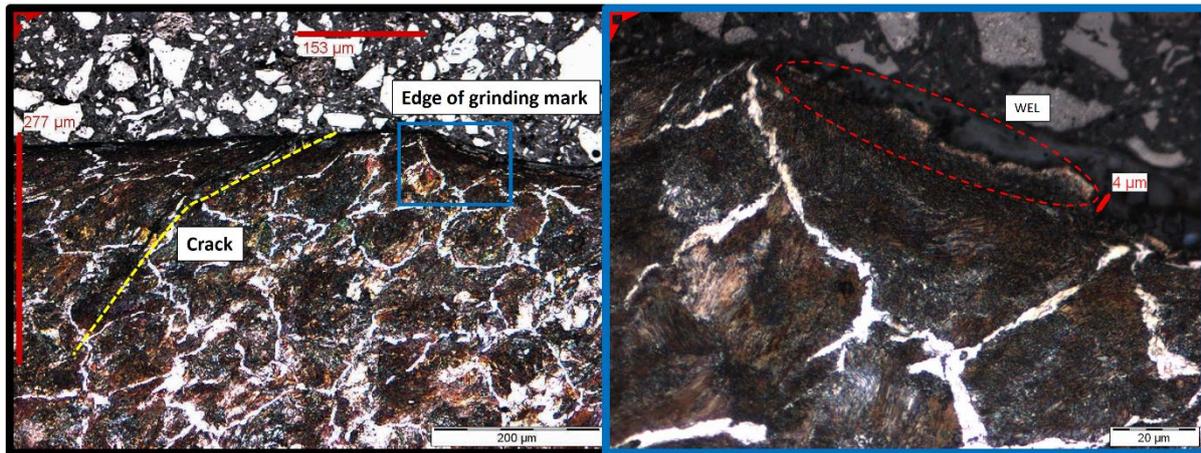
Figure 104: A grinding mark in Sample 2 with a crack developing next to it. The zoomed image in the blue box shows the WEL that made the distinction between a grinding mark and a shallow pit.

Grinding marks create an area of very high contact pressure of the edge of the pit, so cracking is not surprising. However, the combination of partially developed delamination pits and vertical cracks (Figure 105) is less obvious because the contact patch of the wheel has not yet been reduced by material loss. The presence of a vertical cracks on pits and grinding marks raises questions about how many vertical cracks do not initiate independently, but because of surface irregularities such as grinding marks and pits. Figure 105 is an example of pits developing in the same region as vertical cracks. Crack propagation will be determined by a combination of the weakest path, which appears to be the prior austenite grain boundary (left side of Figure 105) and the stress state, which probably accounts for the vertical cracks (right side of Figure 105). Figure 105 shows an example of where the prior-austenite grain boundaries may be. The grains are typically completely round or elliptical.
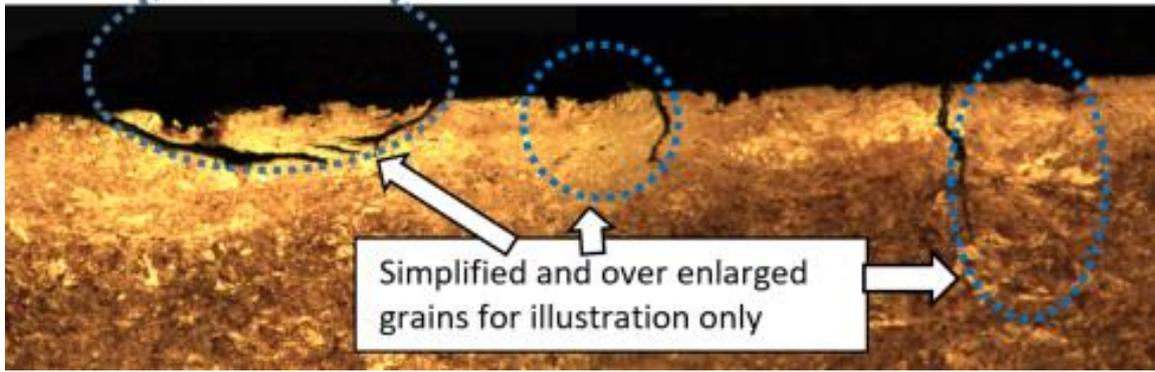
Figure 105: Illustration of cracks following prior-austenite grain boundaries to explain crack variations.

Sample 4 gave an excellent opportunity to investigate the difference in surface cracks in smooth and rough samples because it contained a transition from a patch covered in grinding marks to a smooth region. It was possible that any damage in the smooth area was created before it became smooth, as it will probably have been ground too at some point, unless it was a dipped weld. However, the cracks seen in the smooth area were noticeably different to those seen in the rough area. The rough area had vertical cracks developing, often from grinding marks. The smooth side had no vertical cracks, only typical RCF cracks that ran at a shallow angle to the surface of the rail. The rough cracks grew at an acute angle to the surface before becoming vertical. If this sample is representative then rough surfaces with heavily reduced contact areas, meaning higher contact stress, produce vertical cracks instead of angled cracks typically seen in RCF.
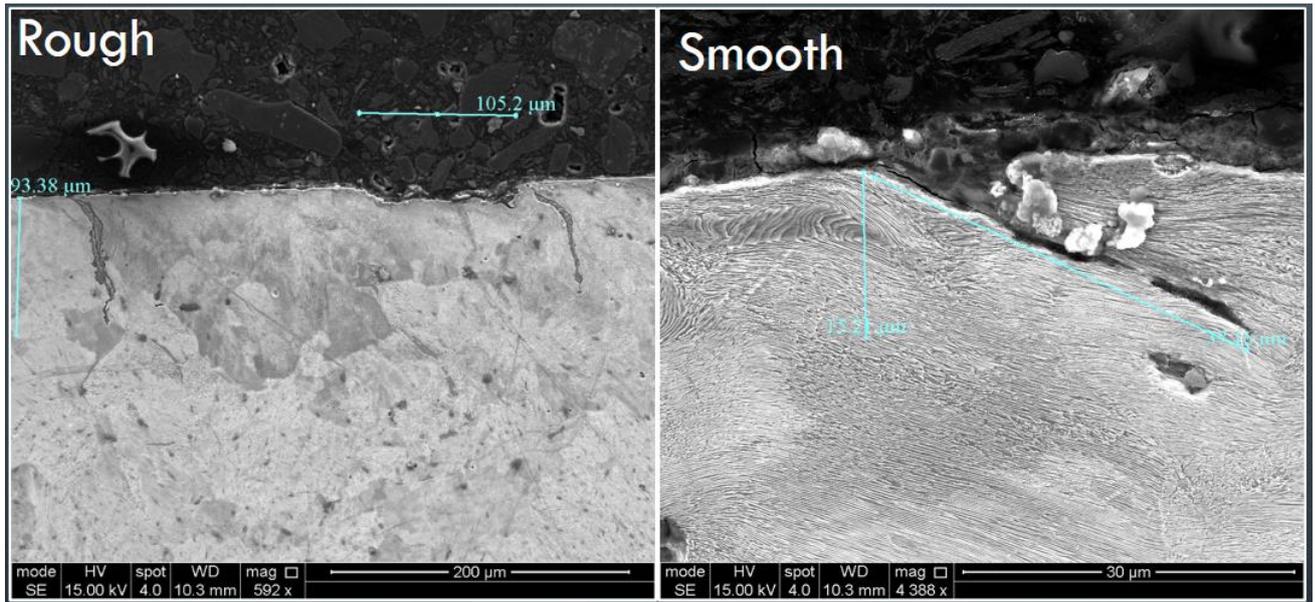
Figure 106: SEM images showing examples of only vertical cracks in the rough region of a sample and only a few RCF style cracks in the smooth region.

5.4.1 Microstructural sensitivity

Figure 107 shows an optical micrograph of Sample 3 with a surface breaking crack. The crack travels fairly vertically through the material rather than at an acute angle to the surface as is usually expected from RCF cracks. This is seen in thermally damaged steel [10]. The interesting behaviour regarding the crack path is that unlike Squats, which typically follow inter-granular ferrite, this crack seems to cut across the grain as documented by Grassie et al. [1]. However, unlike the Studs investigated by Grassie et al. this behaviour also occurs within 450 microns from the surface as well as below.
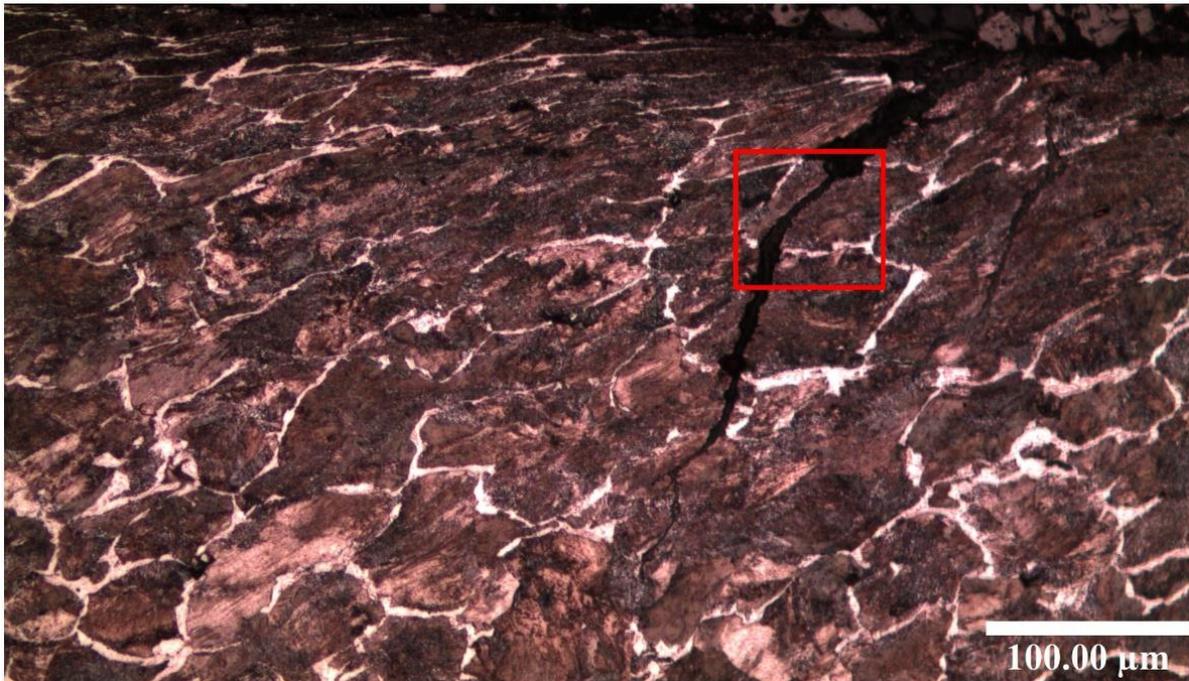
Figure 107: Optical micrograph of sample 2 showing a surface crack with apparent non-sensitivity to microstructure (intragranular fracture). The red box shows where SEM images were taken from (shown in Figure 108).

It should be noted that although the crack seems to ignore the grain in this orthoslice of the rail, it may follow the microstructure more near to its origin, and lose that sensitivity as it propagates out in 3D. The SEM was used to look closer at the crack in Figure 107 to see if it is completely insensitive to the grain structure.

As seen in Figure 108, the cracks travel between lamellae plates a majority of the time, but not always. They seem to be able to change direction when the plates become small and more spheroidal. There are places where this is not true such as the bottom of the right SEM image where the cracks cut straight through plates. This could be because the crack was already orientated parallel to the lamellae of a neighbouring grain and so continued in that orientation as it propagates. Looking at the CT scan data in 3D the crack plane grows down into the rail and diagonally across the rail as it propagates longitudinally and laterally at the same time. This means that the initial crack path through the microstructure cannot be viewed in its entirety within one micrograph.
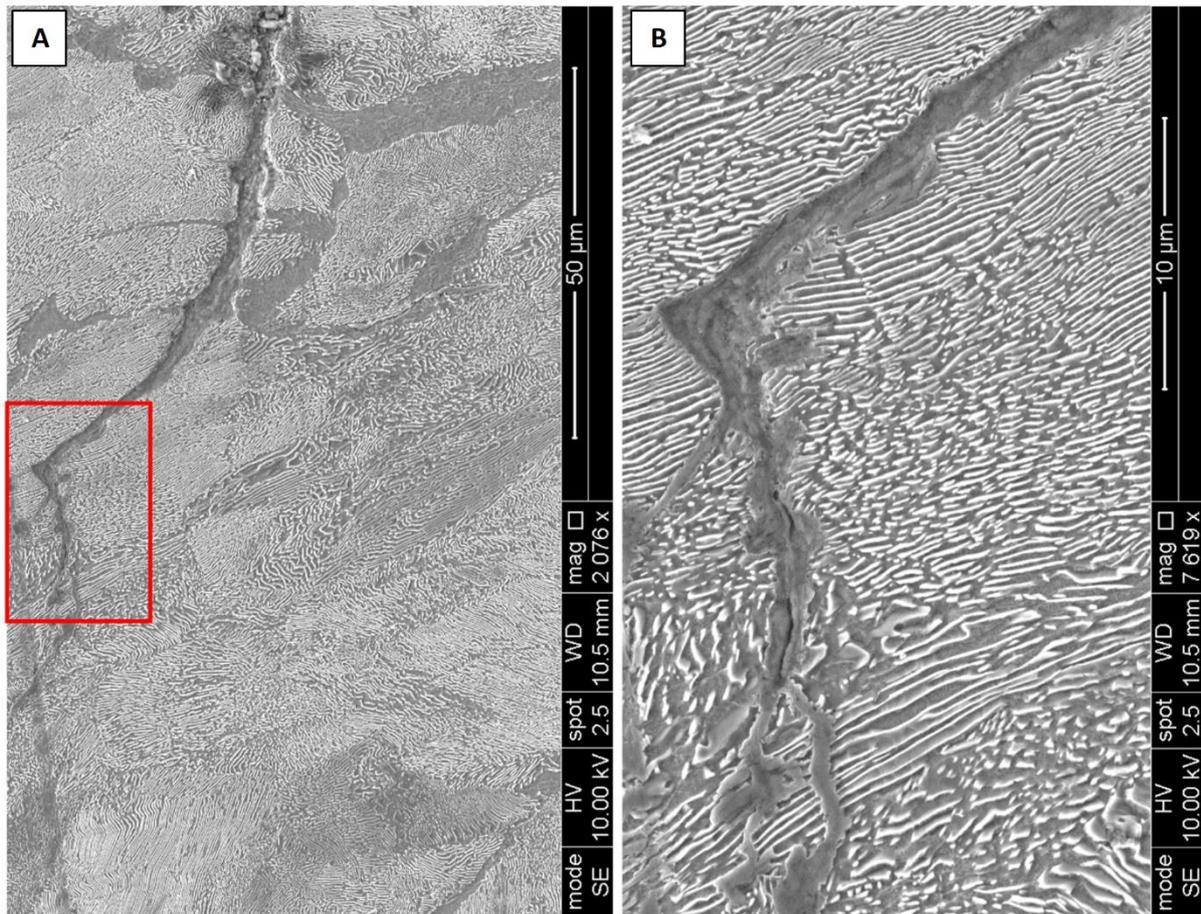
Figure 108: a) SEM image of the box from Figure 107 showing that the crack still prefers to travel between the lamellae plates of the pearlite when cutting across grains; b) magnified view of the box in the left image.

5.5 Hardness mapping

The presence of surface features such as WEL and vertical cracks led to an initial investigation to see if there has been any significant changes/ hardening of the upper surface of the rail. Sample 1 was of keen interest due to the grain refinement that could be seen within sections from the running band (Figure 103). Sample 4 needed to be investigated because of the presence of the weld.

Hardness maps from various areas of each sample provided grid patterns that could give indications of hardness change with depth or surface location. These grids were produced using a Durascan 70 automatic indenter as detailed in Chapter 3. Many maps were produced but only

the maps that have meaningful interpretations are shown here, i.e. maps that show a consistent change in hardness in a direction.

The running band of Sample 1 has two regions: a brighter region where the wheel clearly made contact (R1) and the darker half of the defect where the lobe is (R2) (See Figure 69 and Figure 77 for clarification). Figure 109 gives a clear indication that there is a hardness differential present in the rail. The lobe region shows some increase in hardness longitudinally and with depth. The variation is not significant though. Work hardening is expected to harden the upper surface consistently. Thermal input can soften the material, so if work hardening has occurred, it may have later been softened in places by thermal input.
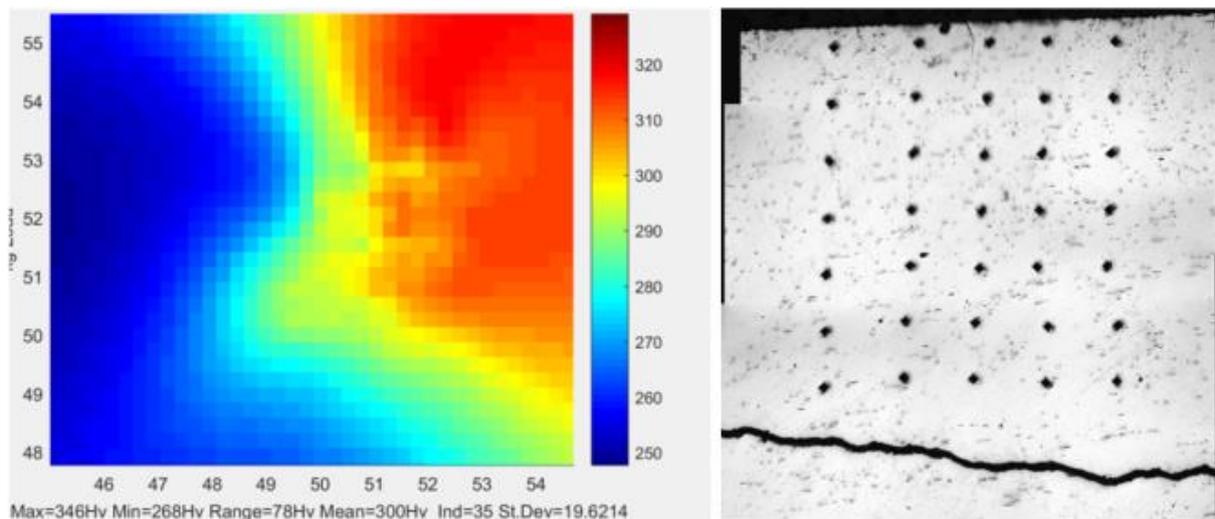
Figure 109: Hardness map for R1 in Sample 1, the lobe region. Points are 0.5mm apart.

R2 has a clearer reduction in hardness with depth, which was what was expected (Figure 110). The high 383HV reading affects the pattern slightly. This is probably due to a harder grain, which could be due to carbon content or crystal orientation. This gradient of hardness is consistent with the gradient of grain size seen in Figure 103.

An area of interest from within the running band was also mapped, to see if there were hardness variations that could provide clues towards the unusual cylindrical crack feature. This sample, known as R3, is believed to have initiated due to a subsurface crack that makes up the lower branches of the defect, before developing a surface breaking crack between the cylindrical

feature and the high surface damage region where the pit was. The lower branch is clearer in Chapter 6, where the defect is discussed more.
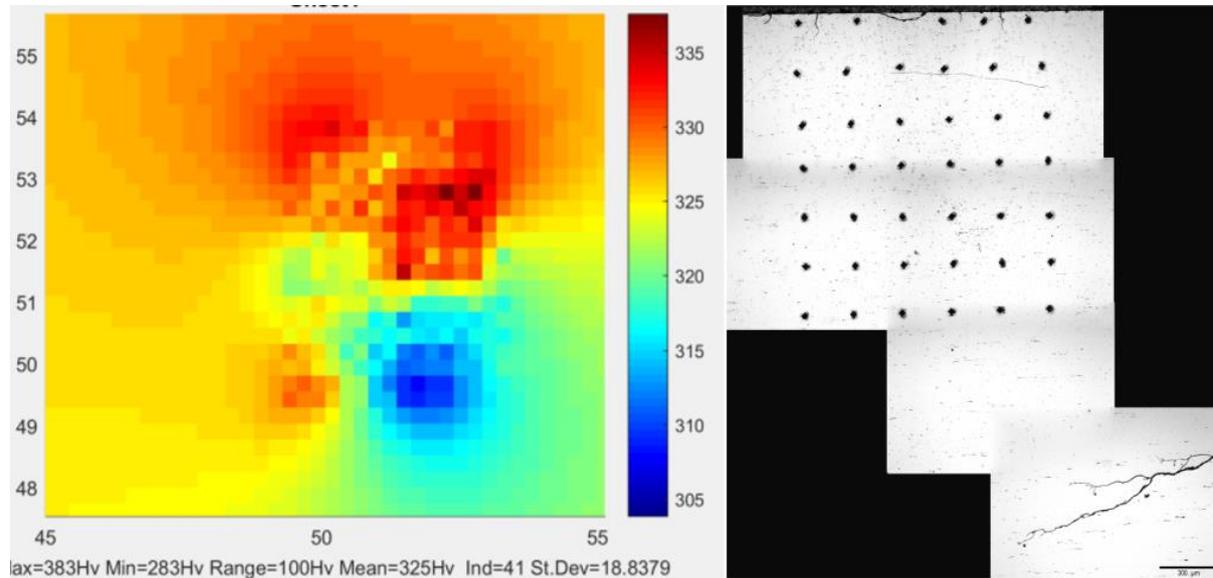
Figure 110: Hardness map for R2 of Sample 1, the brighter region. The same region seen in Figure 105. Points are 0.5mm apart. Scale bar reads 500microns.

The hardness map for R3 delves 3mm deep into the rail and shows a good decrease in hardness with depth like R2 (Figure 111). The highest values are seen in the thin region between the surface and the crack. This is expected as isolated thin strips of material are often seen to deform dramatically in rails. The contour map within Figure 111 shows the softer blue region is found below the lower crack of the defect, with the crack being somewhere within the green band. This is consistent with the idea that the lower crack developed first due to sudden heating and cooling of the material, with the crack appearing on the interface between the heated surface and the cooler subsurface. The red region is found where the surface cracking and void is, which again matches the expectation of strain accumulation in the upper surface. The red region in the lower right of the contour map may need to be ignored. It is interpolated from the points around it because that area was not indented.
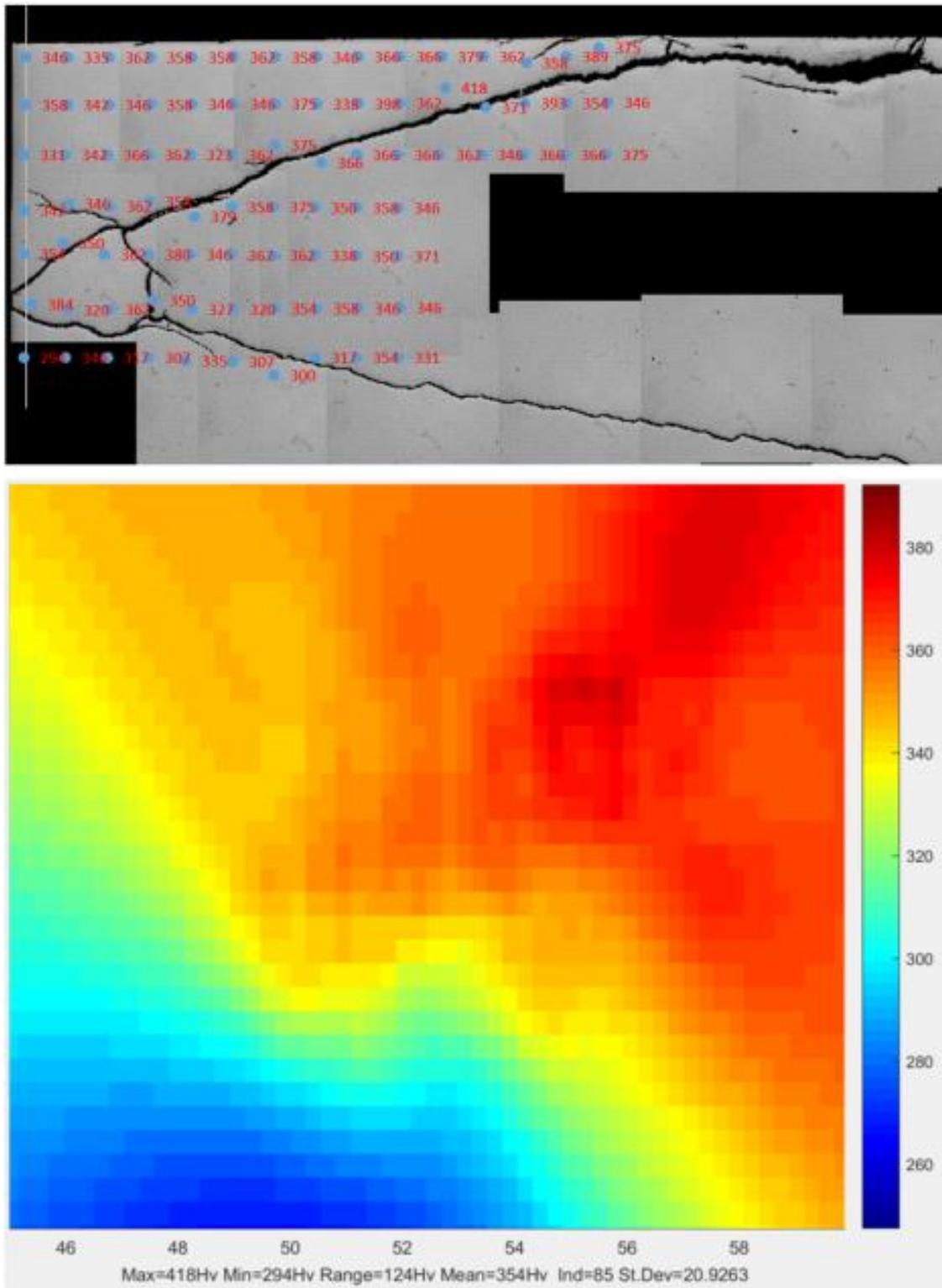
Figure 111: Hardness values and their approximate locations from R3, a sample from the centre of the rail, next to R2. Contains the cylinder feature discussed in Chapter 6. Points are 0.5mm apart but some adjustments were made to avoid cracks.

This gradient from the severely transformed WEL down to the closer to grade hardness of the rail supports the idea that thermal input followed by rapid cooling occurred. However, the lobe region has been out of contact with the passing wheels since dropping below the contact area due to the subsurface crack development. This could explain why the hardness gradient is not found in the lobe region: perhaps the thermal input due to wheel spins occurred after the contact patch was reduced, due to the lobe development. Alternatively, perhaps gentle heating and cooling from nearby regions has allowed the area under the lobe to normalise from a thermally damaged state since its initial development.

5.5.1 Nano-indentation of cylinder of damage
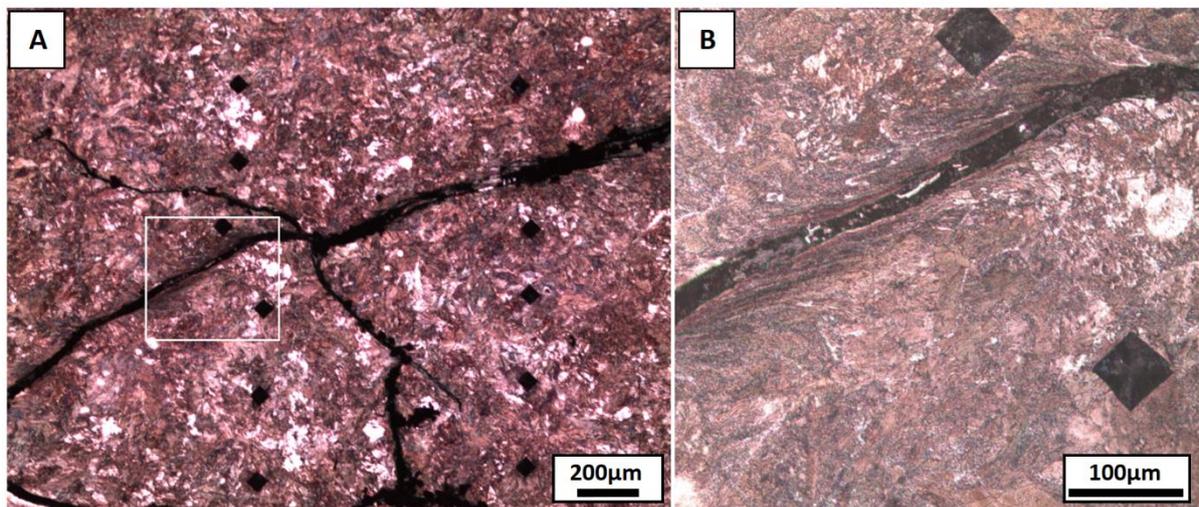
Figure 112: Sample 1: a) The upper part of the cylinder; b) Magnified image of the white box, showing the deformation of the crack faces due to rubbing.

To evaluate the deformed region within the cylinder seen in Sample 1 (Figure 111 and Figure 112) further, nano-indentations were performed. The cylinder is discussed in more detail as part of Chapter 6. Hardness measurements were taken from the region (visible in Figure 112 as black diamonds) were unable to detect any significant variation in hardness. With nano-indentation, 20 measurements were taken per line from above and below the crack magnified in Figure 112. Reference measurements were taken from the centre of the circle and 2mm above the very bottom of the entire sample. The results of these rows are displayed in Table 7 as a summary and in more detail in Figure 113.

Table 7: Minimum, maximum and average hardness values for the nano-indentation of the rubbed cracks faces in Sample 1. HV values from nano-indentation are ~50HV higher than equivalent Vickers indentation. The thick black line represents the location of the crack.

Row name	Av H(HV)	Max HV	Min HV
Above Crack 1	497	563	436
Above Crack 2	536	601	482
Above Crack 3	588	625	552
Below Crack	528	594	444
Centre Circle	476	520	444
Reference	393	427	348
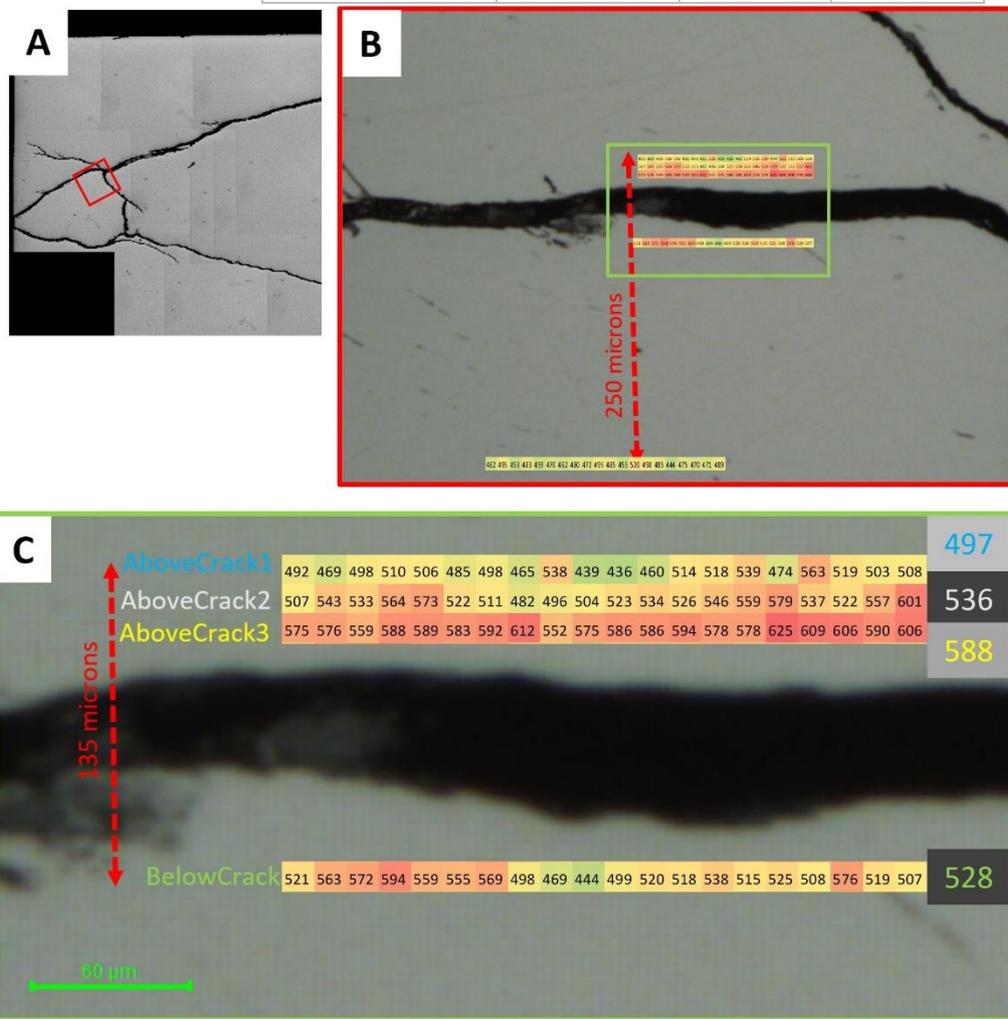
Figure 113: a) An un-etched micrograph of the area; b) Overlaid results for the centre of the circle, above and below the crack; c) A magnified image of the crack to show numbers and the mean hardness for each row.

The results are not surprising with regard to the hardness increasing with proximity to the crack face. The higher hardness for the 'CentreCircle' reference (compared to the main reference) shows that the hardness increase was not localised to the deformed crack edge. The hardness was quite high close to the crack edge, consistent with the expected 800HV+ expected from WEL. WEL was in very thin layers on the crack edge where the measurements were taken from. Other areas along the crack showed no WEL or deformation, but had a rough edge. This suggests that the deformed part of the material had broken away due to the rubbing. Only the indented region showed deformation around the crack edge.

5.6 Pitting

This section is the investigation into the surface pitting that will be discussed more contextually in Chapter 6. Pitting is found both inside and outside of the running band of Sample 2 as well as Sample 5 (introduced properly in Chapter 6), a sample from the same railway track section as Sample 2. Figure 114 shows the surface of a 25x25mm piece of Sample 5 covered with pits of varying morphologies. Pits are distinguishable from grinding marks because they are usually quite smooth and have a layer of oxide. Grinding marks often have WEL in them and a rough morphology to a pit. The large quantity of pits led to an investigation of how they could have formed other than by typical corrosion pitting.

Figure 114: The running band of Sample 5 as a 25x25mm piece with various pits, grinding marks and cracks.

"Delamination pitting" was a possibility and was observed in the partially transformed upper surface of the sample micrographs (Figure 115 and Figure 116). The cause of the pits forming in such shapes is believed to be due to the cracks following the prior austenite grain (PAG) boundaries of the microstructure. The PAG is more prone to crack growth as it is softer [4]. Initial etching in picric acid showed that some cracks followed prior austenite grain boundaries. This could explain the varying morphologies seen in a localised region in Figure 115 as well as the branching of the cracks into vertical offshoots. Figure 115 is from sample 1, which has few pits, but does have snakeskin on the surface. Snakeskin that spalls would produce pitting. Figure 116 is from pitted Sample 2, though most observed pits had already shelled and were oxidised, finding that micrograph of one almost formed was by chance.
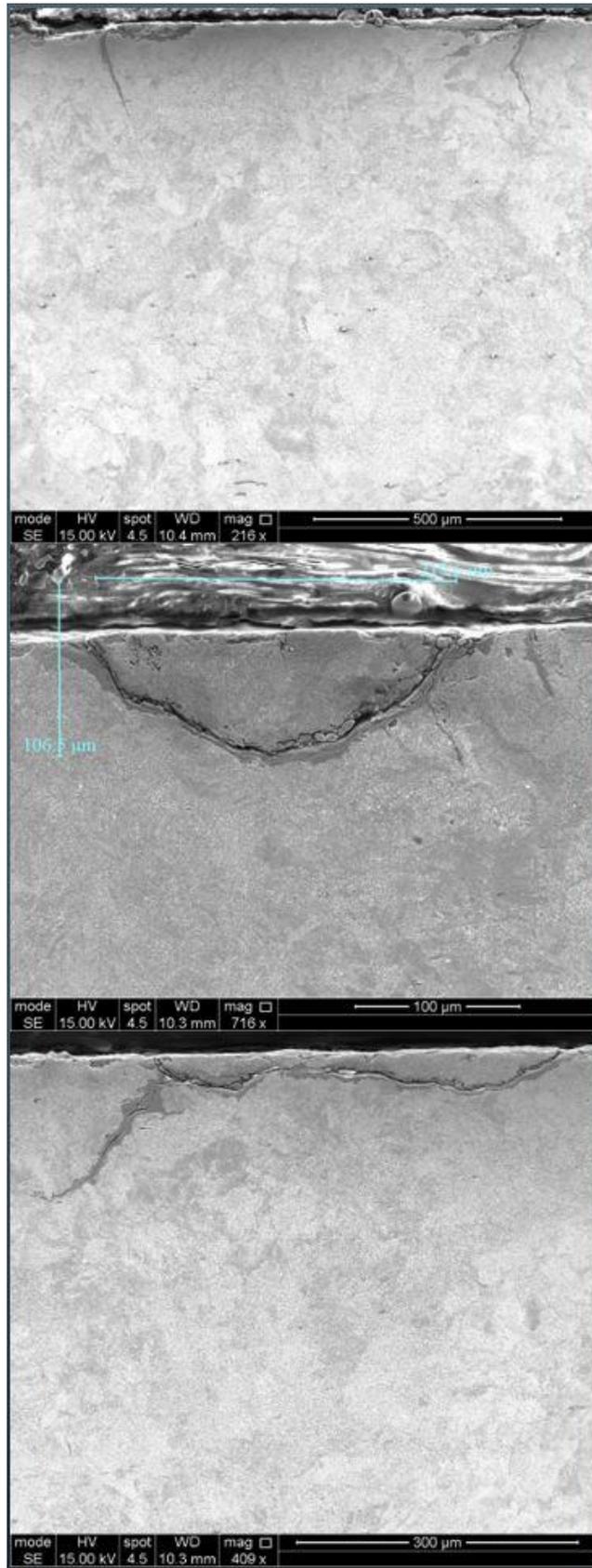
Figure 115: SEM images of pits developing in different locations within Sample 1 due to delamination. Vertical cracks grow from the pit interface.
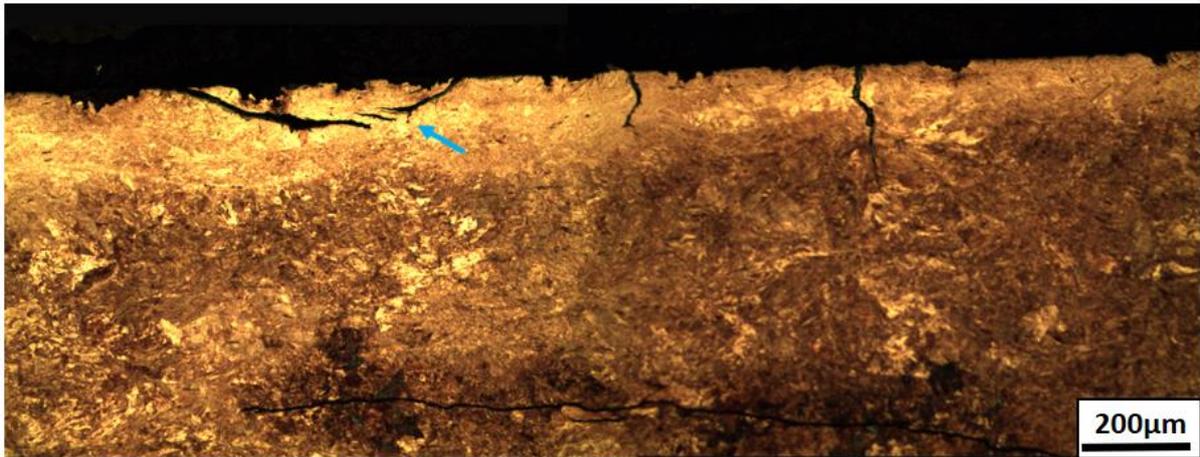
Figure 116: A delamination pit developing alongside vertical cracks in Sample 2. A potential vertical crack has just initiated on the border of the pit, indicated by the blue arrow.
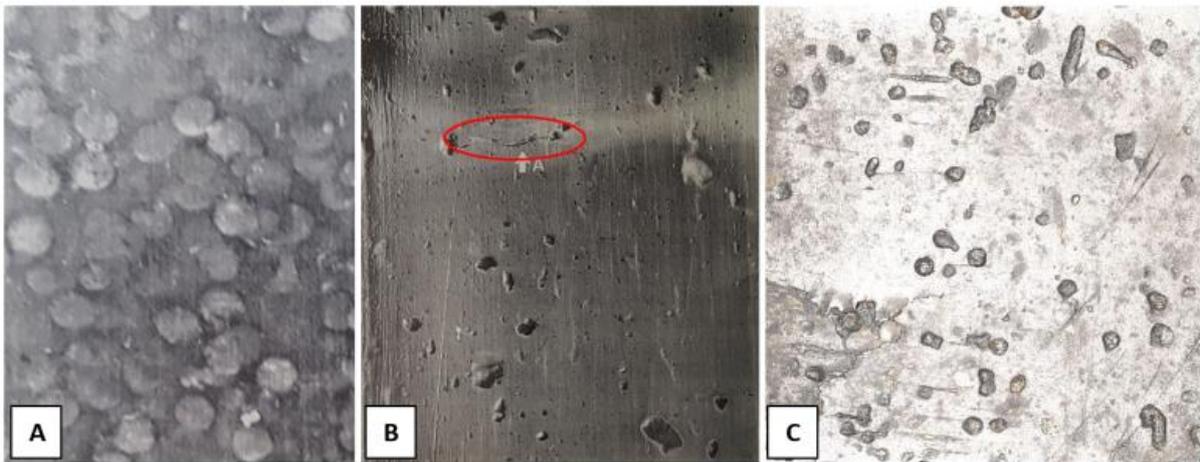
Figure 117: Examples of possible causes of pitting (other than typical corrosion or localised delamination) from literature on bearing steel failures [5]: a) Electric erosion showing very round and overlapping pits; b) Embedded ductile particles producing varied pits and cracks (circled); c) Sample 5 showing various pit morphologies.

Trying to determine what caused the variety of pits found on the rail surface led to a few possibilities (Figure 117). For electric erosion to occur, a current must pass through the contact, arcing from one interface to the other. Samples 2 and 5 were taken from a system that used third and fourth rail, so current from the train should not have been passing through the track that the wheels sat on. However, flashover in the DC traction motors [6], that are associated with the rolling stock on the track where sample 2 and 5 came from [7], provide a plausible

reason. Both the circular pits and the elongated pits can be accounted for when the rolling contact of the WRI is considered [8]. Conversations with two people that have worked on the source railway tracks did confirm they had seen sparking in the WRI on passing trains. So electric erosion is a possibility, but the pits sectioned did not show a heat affected zone (HAZ, which is expected if an electric arc had occurred).

The embedded ductile particle pit morphology is found in bearing steels when contamination gets into the lubricant. It is unlikely that many of the pits were caused by trapped ductile particles as the pits are very circular in nature. Third body particles are more likely to be forced out of the contact that trapped within, although entrapment in a lubricant may be a factor. Some pits may be from embedded particles as the rail is known to have been a soft grade (R220), probably softer than the wheel: the softer of the two contacts tend to become embedded [9].

A thick oxide layer (approximately twice as thick as two of the other pits in the same sample) was found in one of the elongated pits. There was also some signs of mechanical microstructural damage at the oxide-steel interface, which is consistent with gouging of the surface by a third body.

The theorised metal column formations during the DC electric damaging of contacts could also explain indentations. This would be due to the remnants of the columns being left on the contacting surfaces as protrusions, then impacting the rail periodically [8].

In summary, the pits have multiple potential causes of unknown proportions: corrosion, delamination and a few from third body contaminants trapped in the contact (embedding, gouging and indenting). Regardless of their cause, the pits within the contact band were contributors to a reduced contact area in the WRI leading to higher stress in the rail material. This reduced contact area probably exacerbated the crack growth that allowed vertical cracks to appear. Nevertheless, the influence of corrosion cannot be ignored. Figure 118 shows an oxidised ferrite boundary between pits and resembling some of the crack structure shown, full of oxide. Iron oxide is brittle and so can crack easily. Keeping a rail free from oxygen and water does not seem like a reasonable solution to pursue, but minimising the amount of ferrite on a boundary could help to reduce this scenario. Pitting will be revisited as part of Chapter 6.
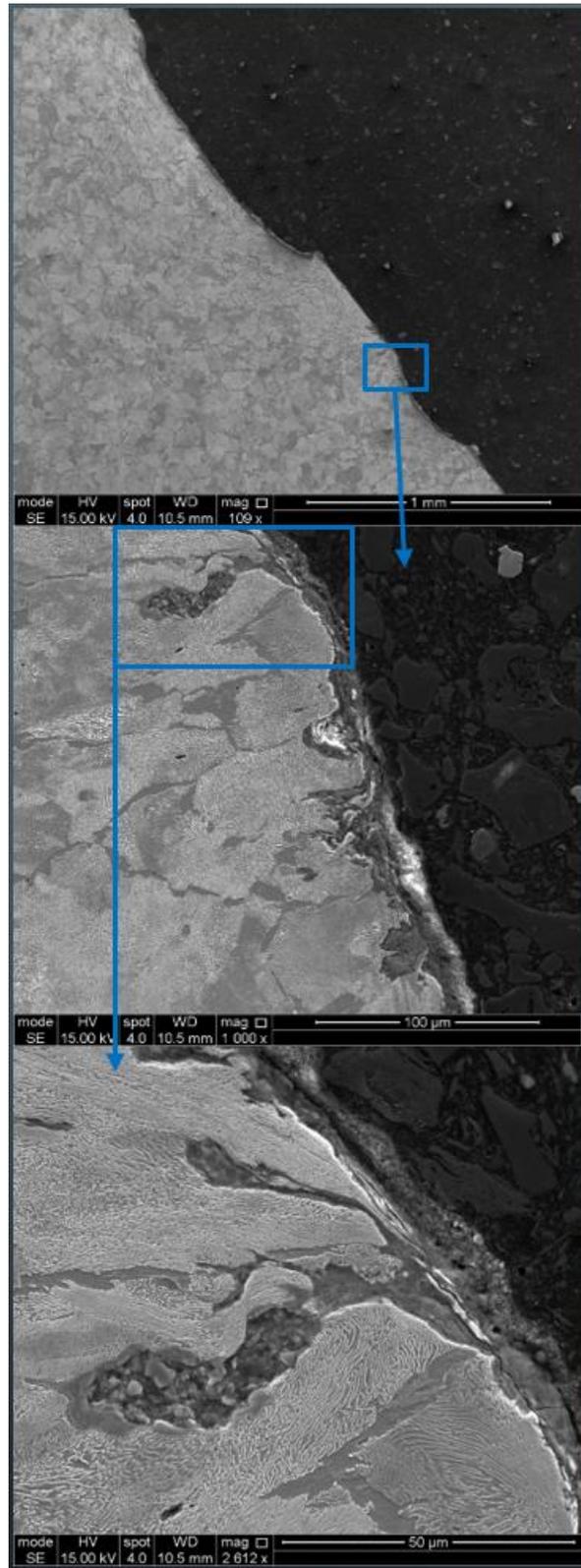
Figure 118: Oxide corroding the ferrite boundary between pits in Sample 5. This oxide would contribute to fretting like that seen creating the voided region in Sample 2. The oxide path back to the surface may be in a different plane to this image.

5.7 Squat initiations

Squats are often confirmed in micrographs by their signature inverted V-shaped crack that causes the two dark lobes on the surface. These cracks grow unevenly, with the leading crack growing first and the trailing crack growing later at a faster pace. Therefore an early stage Squat could be found with only one of the branches developed, which is the case for Sample 5, discussed in Chapter 6. Two small crack structures were found in two separate samples, each matching this early stage Squat like structure. One was only ~200 microns across and was found close to the surface breaking region of the Sample 1 defect (Figure 119).
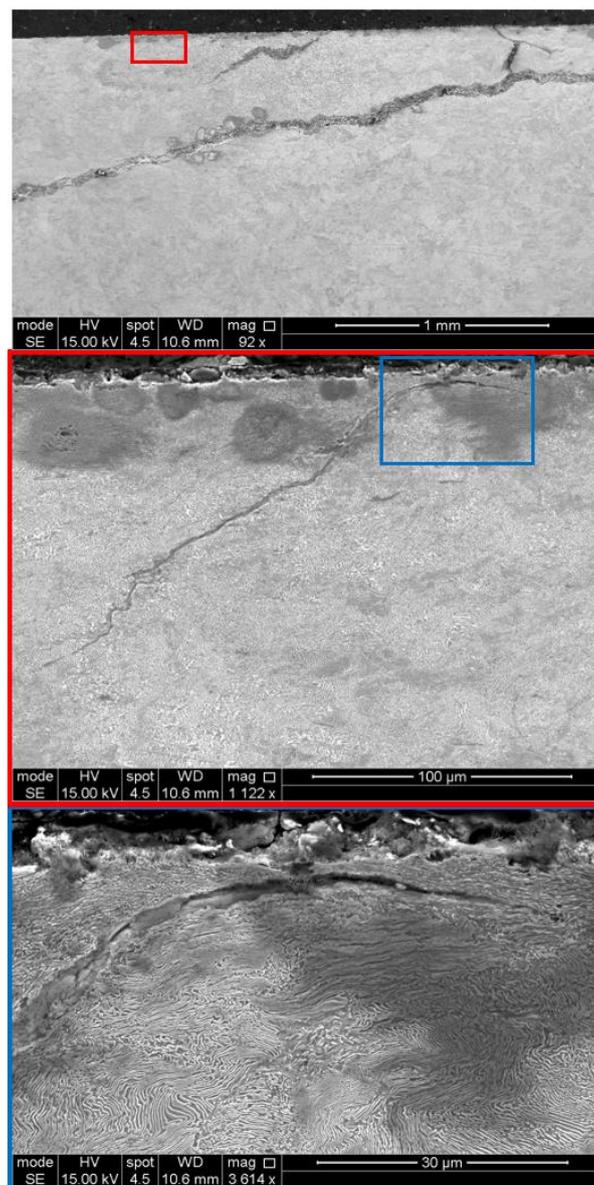
Figure 119: A 200-micron wide crack structure that resembles a Squat type defect.

The magnified images of the crack show that it is just about below the surface, even though some of the thin surface material has broken away. The parallel orientation of the peak of the crack suggests a subsurface initiation. This initial stage to the crack is orientated parallel to the lamellar structure, meaning it appeared between the cementite lamellar (clarified in a higher resolution image in Figure 120). This could mean that subsurface initiations preferentially occur in grains that have lamellar parallel to the running surface. If the grain was perpendicular and a crack developed between the lamellar, this would create a vertical crack, which would surface easily, but may not develop into the structure seen in Figure 119.
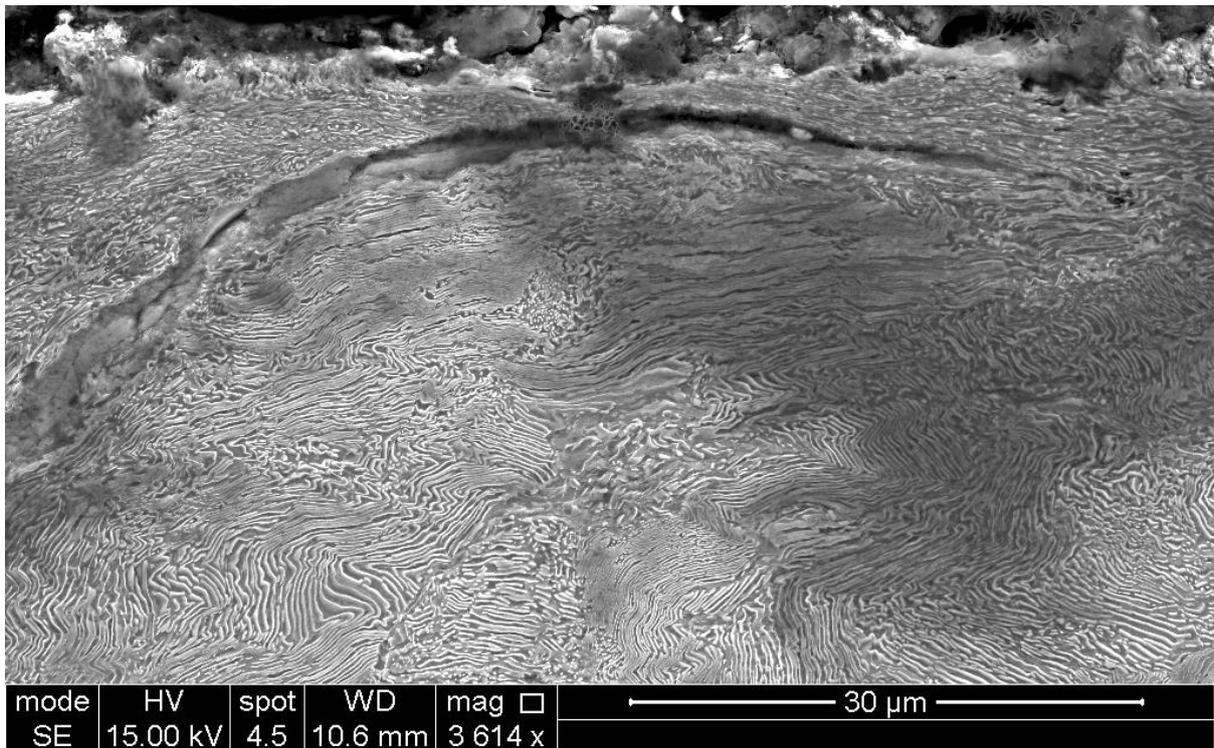
Figure 120: Higher resolution image of the peak of the crack structure in Figure 119.

The presence of such a small crack within a well-developed defect suggests that many more Squats initiate than develop. Perhaps many are worn away before the crack growth rate becomes too high. This does beg the question of why do some grow whilst others do not. It could simply be that the small crack only recently developed. If this crack had grown it would probably have linked into the much larger crack below it.

5.7.1 Contaminants

Squat type defects are generally considered to be caused by the contact conditions that the rail experiences during service. The high pressures exerted by the wheel cause the material to yield and slowly degrade over time, leading to crack structures like those discussed above. However, for two of the samples examined, the primary reason for the failure was determined to be microstructurally caused. The secondary reason was the high contact forces that exacerbated the weakness caused by the microstructural fault.

Sample 3 was of interest due to failing in track via a TD that broke the rail. Microstructural analysis of the steel quickly verified the presence of many long manganese sulphide (MnS) inclusions in the steel. MnS inclusions are known to weaken steel and were found orientated parallel to many of the cracks in the defect. Inclusions are distributed throughout the steel orientated in the rolling direction, the same direction as the unusual ferrite 'stripes' within the microstructure. Most cracks that travelled through the subsurface of the rail also match the orientation of the inclusions and ferrite 'stripes' and so there is a very good chance that this unusual microstructure was the cause of the defect development, shelling and the failure of this rail via a TD.
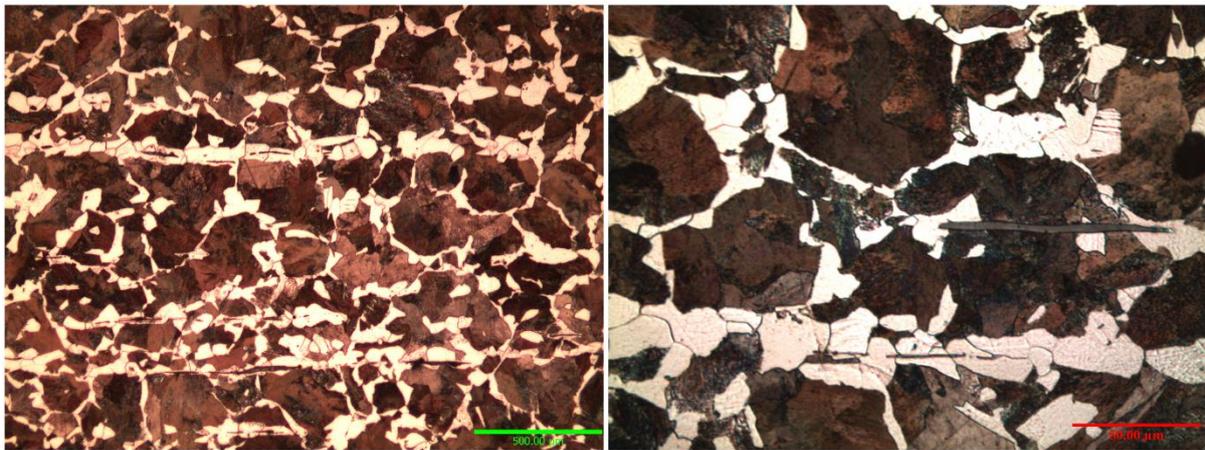
Figure 121: Micrographs of Sample 3 at two different magnifications, showing the unusual microstructure and the inclusions found throughout the steel. Rolling direction was left to right.

The long inclusions running parallel to the running surface also explains a possible reason for the lipped material found at the edge of the break in the rail (Figure 122). The micrographs showed that the lipped material was less deformed than the material below (Figure 123). It was expected that the deposited material would have been deformed more. A reference sample that

did not contain the deposited steel was used to confirm that the identification was correct. This oddity can be explained by considering the lipped material delaminating or shelling from the bulk material, due to a long subsurface inclusion. The train wheel then deposited the detached material onto a region of the rail that had already suffered plastic deformation over time. The deformed region would have been the ridge between the lobes, which is known to take a high contact stress and can crack as a result. It is unclear if this transfer occurred during the rail break or before.

Figure 122: The lipped region next to the rail break, highlighted by the blue circle.
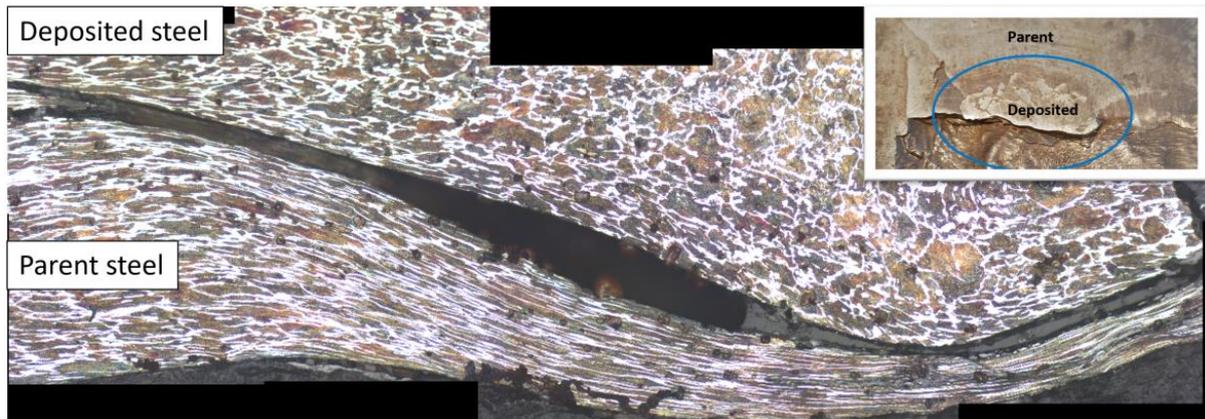
Figure 123: A micrograph of the deposited steel on the parent rail showing the deformation in the parent rail, but not the deposited steel.

Sample 4 is a defect on a weld so a microstructural factor is a strong possibility. The c-shaped crack in the centre of the running band is believed to be the initiation site, with the only other cracks being where the defect surfaced at the edges of the rail. The model produced by the CT scan supports this theory. One of the cuts through the defect was made through the C-shaped crack. Even in its unetched state many little dark spots could be seen just under the initiation site (Figure 124).
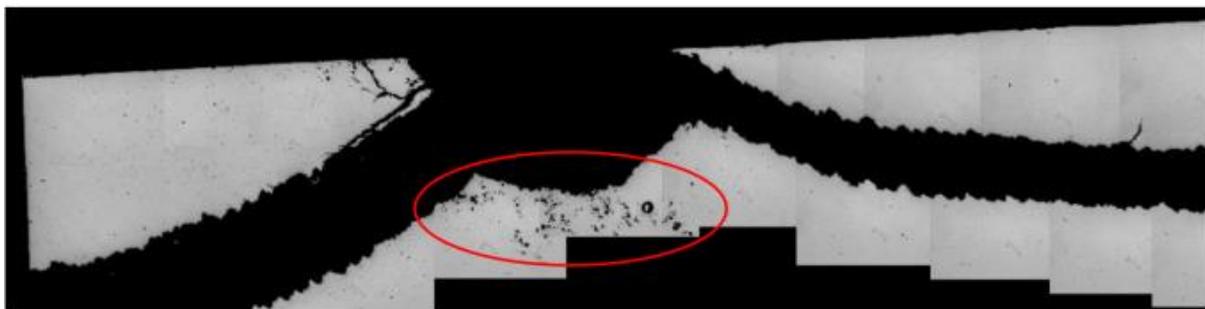
Figure 124: The unetched micrograph of the initiation site showing spots below the initiation. The centre of the C-shaped crack detached and was lost during cutting, hence the large pit.

Figure 124 shows a magnified etched image of the larger of the dots, showing spheres with some material within. These spheres are believed to have been either trapped gas or slag from the alumina-thermic reaction used to create the weld. If the spheres were gas voids, they were filled by Bakelite during polishing, possibly explaining the retained metallic fragments. Small

193

cracks travel between these spheres in the region directly below the cracks. A subsurface initiation is most probable, as these micro-cracks exist in the shallow subsurface.
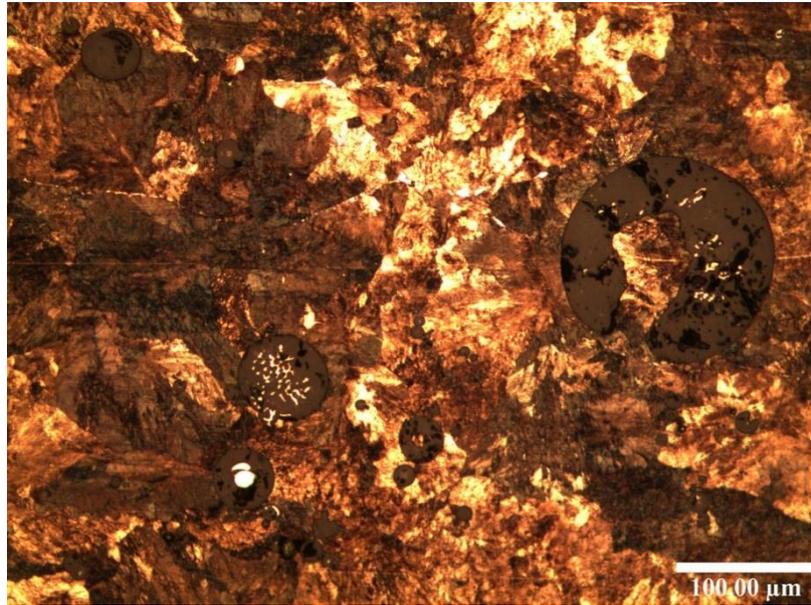
Figure 125: Spheres or bubbles found in the microstructure of the weld edge of Sample 4 within 2mm of the running surface. Probably either gas trapped in the melt. or slag from the welding process.

5.8 Creation of voids through cracking

Chapter 4 showed scans of the crack network with shadows along the crack path, with a brief description of the cause, which will now be extended here. Upon investigation these shadows were seen to be caused by cracks being very close together, such as when one crack branches into two. Sometimes, a recently branched crack would remerge, creating an 'island' that could later break down into a void. Figure 126 shows one of these 'islands' as well as a void that may have formed when one of these islands deteriorated. Figure 126b has two smaller cracks growing from the void, highlighted by the dashed line. This type of cracking used to be seen as star cracking, from bolt holes drilled into rails for fishplates. A problem that was fixed by introducing compressive stresses into the edges of the bolthole when drilling. Consequently, cracks growing from a 'hole' in the material is a known problem, which becomes worse as the material becomes increasingly compromised by a developing crack network.
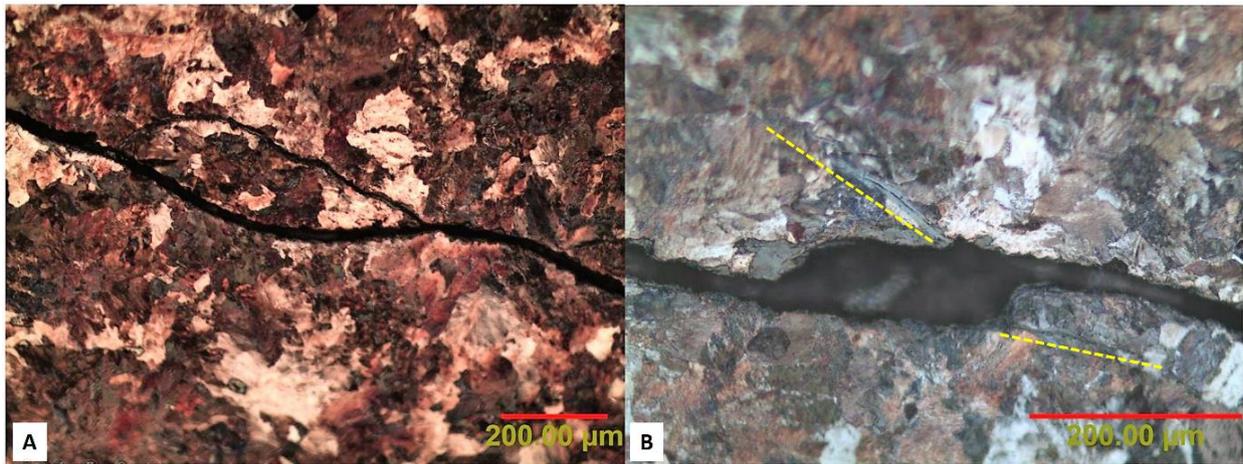
Figure 126: Micrographs showing 'islands' due to branching cracks such as the one seen in Chapter 4 Figure2b. Image (b) is a small void that probably developed to look like (a) before the material detached.

If cracks grow from voids that have formed due to branching cracks, perhaps branching cracks could also lead to small voids developing as shown in the SEM image in Figure 127. The void is from the circular crack in Figure 111, where cracks are branching out in various directions. Deeper into the crack two very thin cracks can be seen entering a large cavern or void. This suggests that sample preparation has made the crack appear wider than it should be near the surface. Moreover, the void could be a result of the extensive cracking or the cracks could have developed from the voids. It becomes akin to the "chicken or the egg" conundrum.

| mode | HV | spot | WD | mag □ | 50 µm |
| SE | 10.00 kV | 2.5 | 10.5 mm | 2 400 x | |
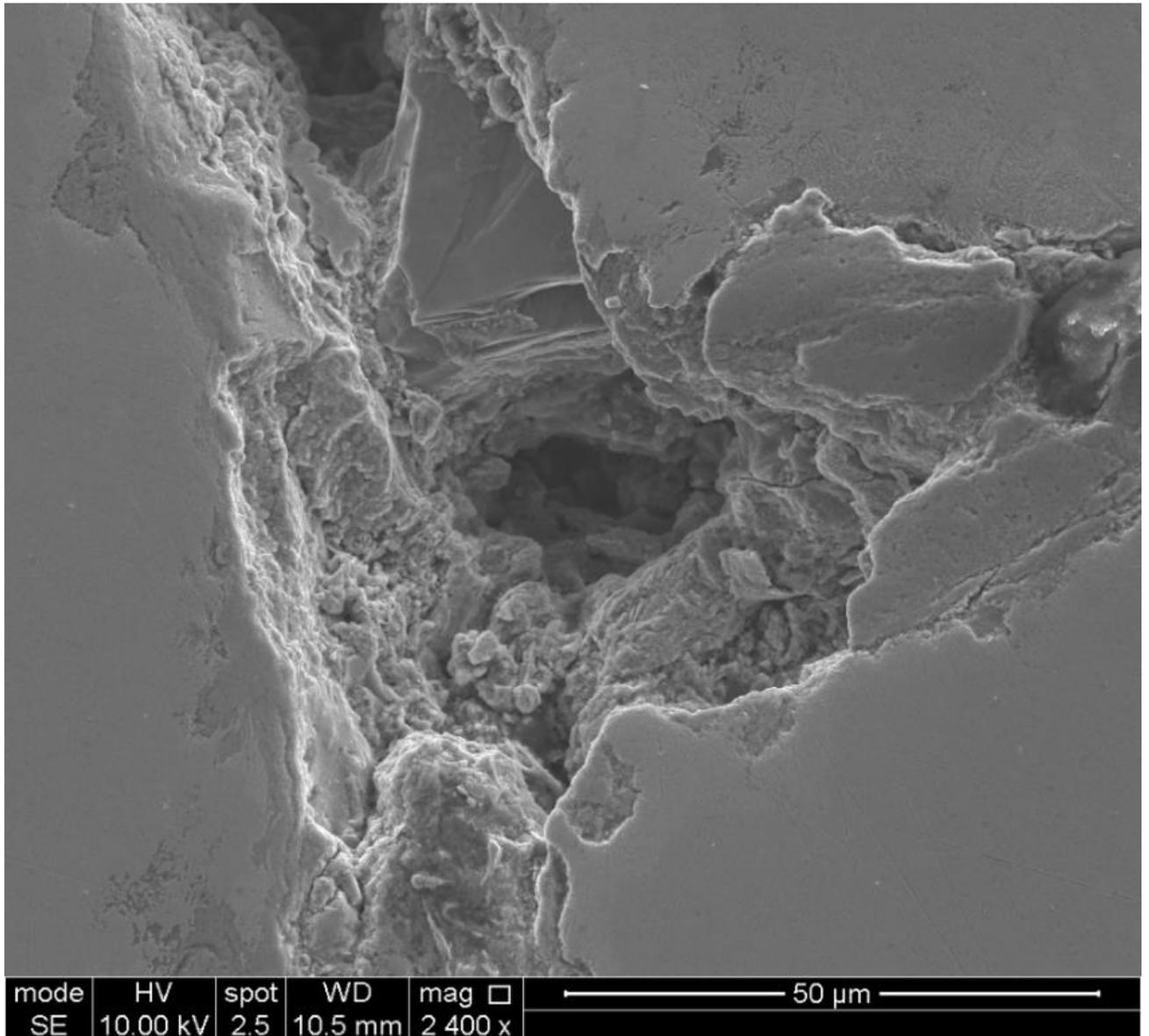
Figure 127: SEM image of a junction where a crack branched off deeper into the rail.
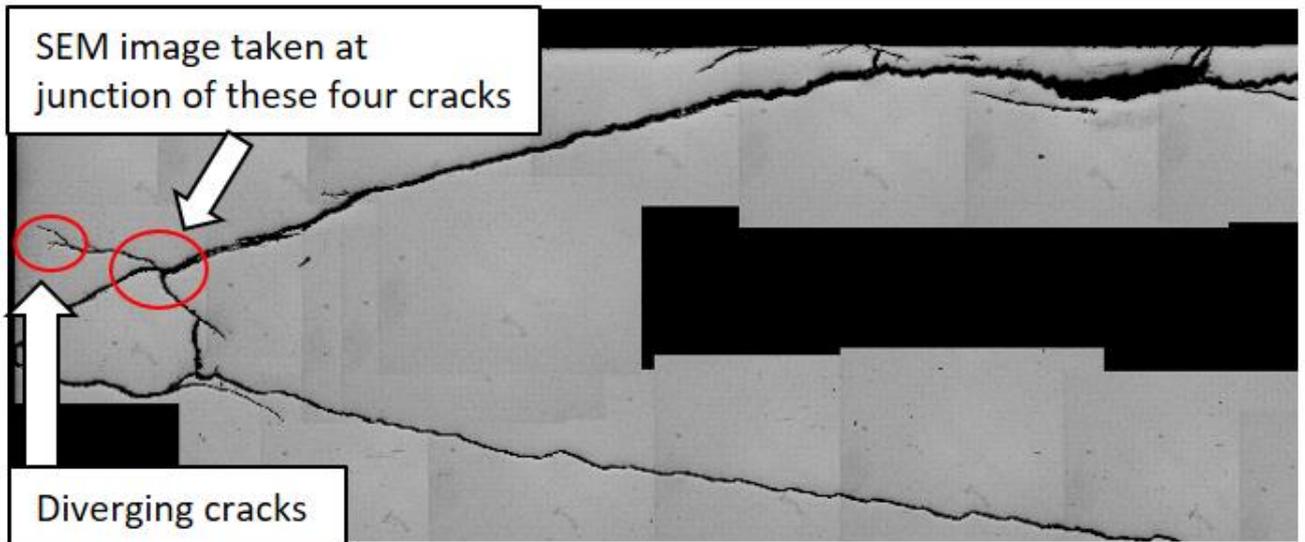
Figure 128: Location of SEM image in Figure 127 and an example of a crack diverging.

Samples 1 to 4 all had the typical inverted-V crack structure expected in a Squat type defect. Nevertheless, the overall topography of the crack planes varied with each sample. The variation is believed to be partly due to differences in service conditions, but this could not be investigated because little is known about where the samples came from. To look for microstructural reasons why the cracks chose to branch, the locations where they diverge (branching points) were investigated.

The difficulty with investigating the point where a crack branched is working in 2D. A micrograph may show where a crack branched in that plane, but it may not be where the crack initially branched in the volume. Therefore, the microstructural clue may not be where the branch lies. The other issue is that the branches from the early development have oxidised over time and widened due to crack rubbing and sample preparation. A way around this issue is to look for recently branched cracks deeper into the defect, but these deeper areas have a different driving force for the propagation of the crack (rail bending rather than contact pressure). No clear microstructural path was found for the defect deeper into the material. This probably reflects the crack propagation becoming continuum mode.

The first branches to be looked at were the small islands that were formed by diverging and converging cracks that created the shadows, covered in Chapter 4. These small islands were found at various locations along the crack path with no obvious preference for where they formed. The only detail noticed with the islands was that there was usually some ferrite near

the branch source, which is the easiest component for a crack to grow through. Figure 128 shows an example of what could become an island in the form of two diverging cracks.

The main benefit of looking at crack branches was the identification of an unusual crack structure in Sample 1 (identified in Chapter 4 and shown in Figure 128). This cylindrical crack seems to have formed from a crack diverging from the original lower crack plane to the surface and then smaller cracks linking these larger cracks together. The reason why the lower crack plane existed first is discussed more in Chapter 6. The erratic path of the lower crack as it passes the bottom of the cylinder suggests that the cylinder is the original cracked region that then propagated out as two subsurface branches. SEM images of subsurface initiated cracks with no observed surface connection are shown in Figure 129. These subsurface cracks are from the centre (longitudinal slice) of Sample 4 and are outside the weld zone. The weld may have caused vertical excitation that created these small crack clusters. Track information is not available to confirm this theory.
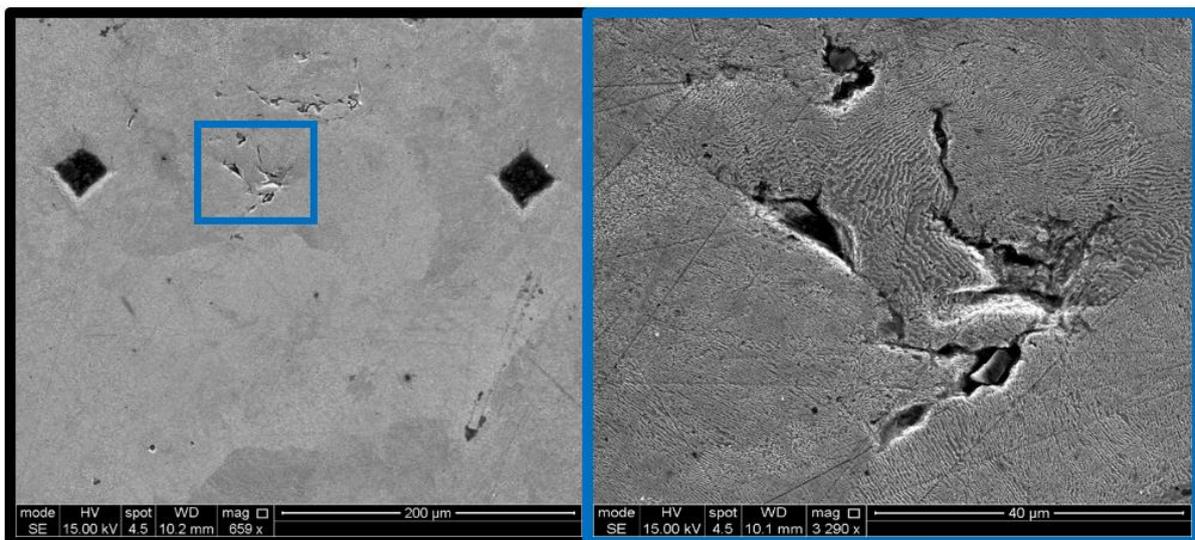
Figure 129: Examples of subsurface initiations growing into clusters of cracks growing in no specific direction. No surface cracks were observed in this region.

The cylindrical crack in Sample 1 shares some similarity to the void in Sample 2. This is discussed in Chapter 6 along with the introduction of Sample 5, another sample with this type of star branching crack. The reason for the interest in these 'cylinders' and voids is because the only observed transverse defect in the samples collected for this work, originated at a void.

5.9 Discussion

This Chapter contains investigations into features that are associated with thermal damage, to try to distinguish between Squats and Studs. First the thermal damage observed will be discussed before crack initiations and subsurface features of Samples 1 and 2.

Squat type defects are not just caused by the elastic exhaustion and plastic deformation of the upper surface of the rail. Sometimes other factors such as a weld contaminant or inclusions are involved. In the case of the inclusions, the steel is compromised, but other factors still have to occur to generate the defect. Many features that are believed to be strong factors in the defect's growth are found subsurface. Subsurface initiations could be due to thermal shock causing initiations between the heat-affected surface and the cold parent rail, or because of rough surfaces driving the peak stress position deeper into the rail.

As already established in the literature, WEL renders the surface of a rail susceptible to crack growth and possible defect development. However, WEL is often found on used rails but not all rails have Squats, so the cause of Squats may include WEL, but it also includes other factors. An analogy would be the fire triangle: oxygen, heat and fuel are all needed for the fire to exist, if one is not there then the fire dies. WEL could be akin to the oxygen in the analogy: it is found in many places but there is not always a fire there.

Some of the WEL was attributed to the remnant effects of grinding, leaving an undesirable surface condition of a rough brittle material. The WEL in Sample 1 seemed to have layers to it, possibly deposited by different WEL inducing events, leading to a tempering of the initial layers of the overall WEL.

Pits also seemed to have an influence on surface crack initiations, but for different reasons than the WEL. A reduced contact area results in higher pressure being distributed across the remaining contact surface. This appears to have led to cracking on the edges of pits as well as some grinding marks. If this crack grows deep enough, the next maintenance grinding pass would remove the evidence of the pit but the crack tip may be left behind, as is the case with WEL. This confirms that the surface quality is very important when looking at where cracks are appearing.

The Stud seen in this work shares most of the features originally documented by Grassie et al. [1], but have a few differences such as cracks crossing grains close to the surface as well as deeper into the rail. There also seems to be less WEL, especially regarding depth. This could be that these Studs experienced a different history of temperature change, perhaps due to a different contact patch or other tribological factors. Both previously documented Studs and the Stud seen here both show signs of thermal influence. This work determined that this was due to grain refinement in some areas, an overall lack of plastic flow and the presence of fairly vertical cracks that break the surface. Vertical cracks are often seen through thick WEL, but none was seen in the presence of these cracks, though that does not guarantee that it was never there.

Both an isolated surface crack and a Squat like cracks structure were investigated by SEM. Most importantly in some areas the believed initiation site for the Squat like crack, cracks were seen growing between lamellar plates, showing that grain orientation could be a factor in the early stages of a defect's growth. Therefore, the microstructure does play a role in the cracking of the rail, though it is harder to see the influence as the cracks grow longer, leading to strong driving forces overwhelming the microstructures resistance.

Sample 1 exhibited localised subsurface deformation of the crack faces around the crack cylinder. This suggests that cyclic shear stresses were transmitted through the area, allowing rubbing of crack surfaces in the area due to the region being structurally compromised by the presence of so many cracks. This is why the 'cylinder' is believed to have been free/ detached but constrained within the railhead. The rubbing of crack surfaces, and the presence of free but constrained material trapped within a crack network, has been seen in the literature [11], [12]. These examples in literature are linked to grinding damage, where multiple defects can lead to long sections of spalling. It is believed that although Sample 1 and 2 have slightly different causes, the cylinder cracking in Sample 1 could later develop into the void seen in Sample 2 (Chapter 4, Figure 22). This would occur by the further detachment of material, already seen around the cylinder due to deformation from the high forces being transmitted through the area. This detached material would oxidise, producing hard oxide particles, creating an even more abrasive contact between the crack surfaces. This would then be rail fretting, eroding at the material in the centre of the cylinder/ void as well as the outer crack faces. The cylinder and the void have similar dimensions, discussed more in Chapter 6. The void in Sample 2 was the source of the transverse crack that could have broken the rail.

5.10   Conclusions

To review the aims that this Chapter was addressing:

    3.   Are there differences between the crack morphologies of defects?

It was shown visually in Figure 96 that the overall morphology of the defects share the familiar inverted-V shape seen in the literature on Squats, but vary as they develop.

Surface quality is important as various surface irregularities are linked to damage features within the rail. These irregularities can lead to high subsurface stresses that can cause initiations beneath the surface. Features with high damage, such as the subsurface crack cylinder in Sample 1 and the void in Sample 2, lead to further cracks developing. The more cracks that develop, the higher the chances of one of them growing transversely. Furthermore, subsurface cracks have longer to develop before the wear of the rail can erase them. Subsurface branching cracks were connected to surface features that are discussed in depth in Chapter 6.

    4.   Is there a correlation between the microstructure and the defect?

Grain refinement can be suggestive of a Stud rather than a Squat. WEL is an indicator of thermal input too, but it is more common and can be caused mechanically. Vertical cracks also seem to be indicative of a severe enough thermal event, as they are seen in steels used in high temperature applications.

Squat type defects can occur because of impurities in the steel, either from manufacturing or from installation (welding). This information is not new, but perhaps impurity related defects should be grouped separately.

The discovery of a small Squat shaped crack gives a glimpse at the very early stages of a Squats initiation. It appears that the initial crack formed subsurface, parallel to the surface and in between the cementite plates of the pearlite. Therefore, in this case, the orientation of the grain was a factor.

References

[1]   S. L. Grassie, D. I. Fletcher, E. A. Gallardo Hernandez, and P. Summers, "Studs: a Squat-type defect in rails," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 226, no. 3, pp. 243–256, 2011.

[2]   S. Djaziri, Y. Li, G. A. Nematollahi, B. Grabowski, S. Goto, C. Kirchlechner, A. Kostka, S. Doyle, J. Neugebauer, D. Raabe, G. Dehm, "Deformation-Induced Martensite: A New Paradigm for Exceptional Steels," *Adv. Mater.*, vol. 28, no. 35, pp. 7753–7757, 2016.

[3]   S. Simon, "Thèse De la dynamique ferroviaire à l' accommodation microstructurale du rail," 2014.

[4]   T. KUNIO, M. SHIMIZU, K. YAMADA, M. ENOMOTO, and A. YOSHITAKE, "the Role of Prior Austenite Grains in Fatigue Crack Initiation and Propagation in Low Carbon Martensite," *Fatigue Fract. Eng. Mater. Struct.*, vol. 2, no. 3, pp. 237–249, 1979.

[5]   T. E. Tallian, *Failure Atlas for Hertz Contact Machine Elements*. New York: ASME Press, 1992.

[6]   Y. Nakanishi, K. Sugimoto, and N. Morita, "Investigation for the Flashover Phenomenon in DC Machines," *IEEJ Trans. Ind. Appl.*, vol. 129, no. 11, pp. 1054–1059, 2009.

[7]   RAIB, "Uncontrolled evacuation of a London Underground train at Holland Park station 25 August 2013," Derby, UK.

[8]   Y. C. Chiou, R. T. Lee, and S. M. Lin, "Formation mechanism of electrical damage on sliding lubricated contacts for steel pair under DC electric field," *Wear*, vol. 266, no. 1–2, pp. 110–118, 2009.

[9]   N. Gao, R. S. Dwyer-Joyce, and D. G. Grieve, "Disk Machine Experiments to Assess the Life of Damaged Railway Track," 2001.

[10]   J. H. You, "Damage and fatigue crack growth of Eurofer steel first wall mock-up under

cyclic heat flux loads. Part 2: Finite element analysis of damage evolution," *Fusion Eng. Des.*, vol. 89, no. 4, pp. 294–301, 2014.

[11]   M. Steenbergen, "Rolling contact fatigue: Spalling versus transverse fracture of rails," *Wear*, vol. 380–381, pp. 96–105, 2017.

[12]   M. Steenbergen, "Squat formation and rolling contact fatigue in curved rail track," *Eng. Fract. Mech.*, vol. 143, pp. 80–96, 2015.

Chapter 6 : Surface analysis of Squat defects

Previous Chapters addressed the first four aims, reiterated below, establishing the tomography of entire crack networks for multiple defects (Samples 1-5) within a railhead. This provides a new perspective on how these defects initiate, propagate and vary. The coupling of the surface maps and the CT volumes provide an opportunity to compare the surface features with the subsurface features, thereby addressing the fifth and final aim of the work.

1. Can µ-CT scanning accurately detect crack networks on a large scale within a rail?
2. Are all of the features captured during scanning real, or do artefacts make the results deceptive?
3. Are there differences between the crack morphologies of defects?
4. Is there a correlation between the microstructure and the defect?
5. What influence does the surface quality have on the defect growth behaviour?

Chapter 4 focussed on the initial CT work and Chapter 5 supported the scan verification and defect identification through more traditional methods. Chapter 6 focuses on the surface to sub-surface connection between features, covering the surface roughness of the four samples from Chapter 4 with the addition of an extra sample (Sample 5), to establish common and anomalous features. The 5 samples are summarised in Table 8.

The subsurface features of three similar defects are then discussed individually, before linking subsurface features of interest together with the surface in the discussion. The three similar defects were from metro railways, which have relatively light rolling stock, so extensive damage was not expected. Focus will be on Sample 2, with comparisons to sample 1 and sample 5. Sample 5 is a less developed defect with only a leading crack, but no trailing crack: a single lobe Squat type defect. The similar surface damage of these three samples will indicate wether the subsurface features are related to the surface or coincidental. The selection of the three samples are explained in Table 9.

Table 8: A summary of the samples. *A volume is the initial 3D raw data from a CT scan and the model was post processed to highlight features of interest and create a clearer 3D image.*

Sample number	Location removed from	Traffic type	Grade	Volume created (Y/N)	Model created (Y/N)
1	France	Metro	R260	Y	Y
2	UK	Metro	R220	Y	Y
3	France	Mixed	R260	N	N
4	UK	High speed	R260	Y	Y
5	UK	Metro	R220	Y	N

Table 9: The three similar samples investigated in this Chapter. All from metro networks.

Sample number	Relevance
1	Shares subsurface feature with Sample 2
2	Contains dangerous Transverse Defect
5	From same railway line as Sample 2

6.1 Introduction

Micro-computed tomography (µ-CT) scanning has been used to investigate the initiation of Squats [1]], the development of head checks [2] (a form of RCF similar to GCC) and to compare the morphology of Squat type defects [3]. Models built from µ-CT scans allow the tracing of cracks back to their origin, to identify the cause of features such as TDs. This origin is then linked to related relevant features on the surface of the rail, through detailed mapping on an infinite focus microscope. This approach allows a track inspector to know what features to look for on the surface of a rail making potentially dangerous defects more noticeable. The identification of the origin of damage and its related surface features are the focus of this work.

Two of the reconstructed micro-CT scan volumes provided 3D models showing complex crack networks, known as Model 1 and Model 2. These models provided a guide for precise cuts through the rail to investigate any regions of interest. Focus was on where cracks branched into new cracks. This is important to understand because if a crack grows down into the railhead rather than upwards back to the surface, it can break the rail and possibly derail a train. The presence of these downward growing cracks, known as transverse defects (TD), have been noted in Squats before. A growing crack either becomes a TD or resurfaces and shells between 3-5mm in depth [4]. The TD crack in Model 2 grew to ~9mm deep, making it very likely to be a transverse defect.

Squats and Studs appear to be very similar both on the surface and in the subsurface crack morphology. Figure 130 shows the surface features associated with Squats, which develop due to the crack growing beneath the defect, meaning that a Stud defect appears the same on the surface. The differences are found in the micrographs and through checking the traffic history.

Studs are subtly different in detail to Squats for various reasons [5]. Most notable are their faster development (~4x), the strong links to thermal damage in the initiation of Studs and the lack of surface plastic deformation. Studs were not a concern at first because unlike Squats, Studs were not considered a sample that developed into a transverse defect (TD) (Figure 131). A TD is a safety risk as they eventually break a rail, risking a derailment. However further investigations showed that when another type of rolling contact fatigue (RCF) defect called GCC was present as well as a Stud defect, TDs were possible [6]. Grinding induced Squats (GIS) are a new classification of Squats, with the cause being linked to grinding quality rather

than the exhaustion of the rail material [7], [8]. A rough finish after grinding causes higher subsurface stress due to the reduced contact.
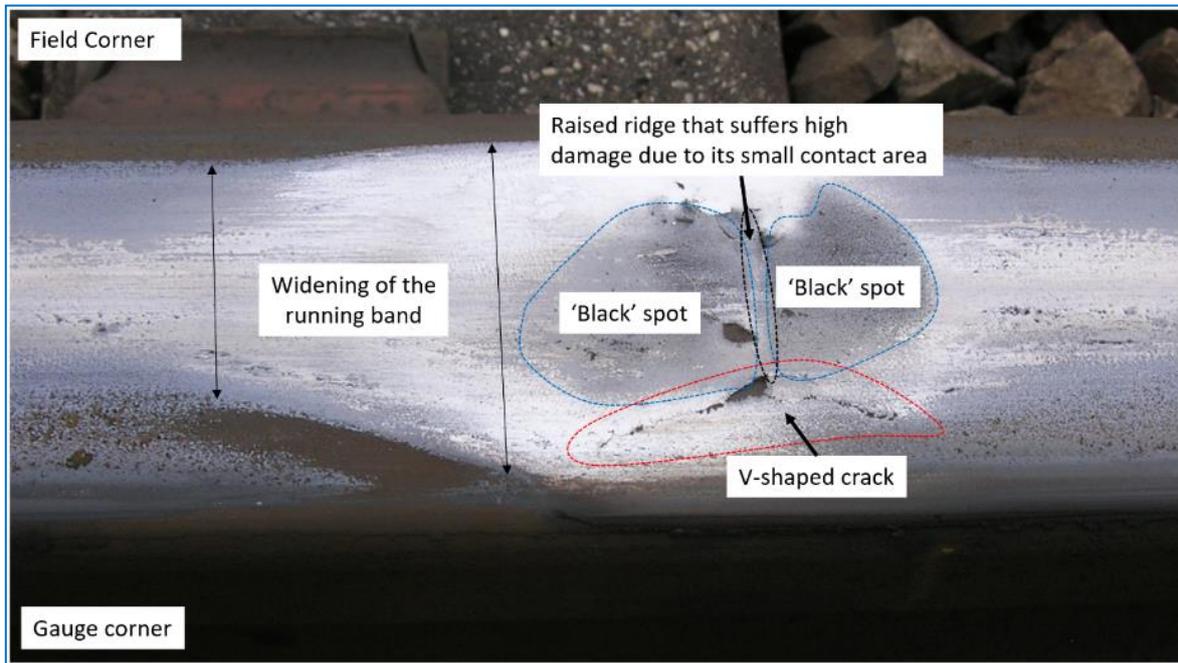

Figure 130: A typical Squat with its usual features highlighted. Original image courtesy of British Steel
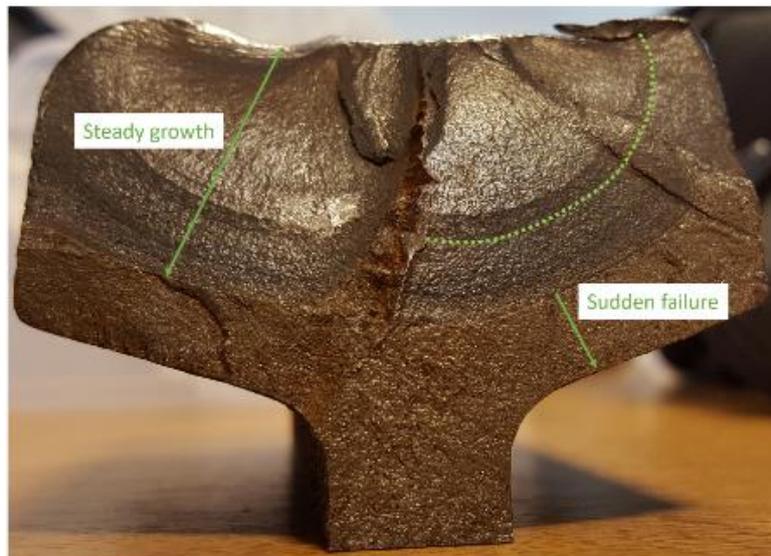
Figure 131: A railhead that has failed due to a transverse defect (TD) that developed from a Squat. The green dotted line highlights the beach-marks that show the crack growth speed increasing with crack length.

6.2 Squat type defect samples

Three defects were compared as part of this investigation in order to establish a link between what can be seen on the surface and the critical features of a sample that make it develop into a dangerous TD. Sample 1 came from a metro line, but little is known about it, other than it being from an inclined track near a station. Defects 2 and 5 came from the same metro line with only one type of rolling stock, with the main difference being the maturity of the defect. Sample 2 was scanned twice because the region of downward branching cracks were difficult to distinguish between within the full defect scan. A smaller region of 25x25x15mm was removed to scan again to get a higher resolution. The defect samples are shown in brief in Table 10.

Table 10: List of defect samples investigated including their reason for inclusion and their size at the time of scanning

Sample	Removed from	Significance	Scans and dimensions (wide, long, deep)
1	Metro, France	Extensive surface cracking due to a sliding event	Entire defect as 70x55x15mm bulk
2	Metro, UK	Surface pitting and a transverse crack	Entire defect as 74x85x15mm bulk and as 25x25x15mm region of interest
5	Metro, UK	Surface pitting and the same railway line as sample 2	Entire defect as 40x40x20mm region of interest

Being ~70mm wide and a similar distance long, railhead samples containing defects are difficult to scan because of the density of steel. The availability of a 450kV μ-CT scanner made the first scans of full railhead defects possible [3]. Inspections of these scans alongside maps of the surface provided clues of why the subsurface features developed. The initial inspection of the scans is covered in Chapter 4 with the supporting evidence in Chapter 5. The identification of surface features provides a practical tool for track inspectors who cannot view inside the rail, making it part of the focus for Chapter 6.

6.3 Results

6.3.1 Surface mapping and roughness measurements

Various features on the rail surface led to the documentation of the surface roughness for all five of the samples. Focus was on grinding marks, pits and any unknown vertical irregularities. The purpose was to determine numerically what was smooth, what a typical roughness was and what could be considered too rough. Roughness was measured from the surface maps made using the Alicona IF Microscope, as detailed in Chapter 3. Some samples were investigated more than others depending on how much the surface varied. Ra and Rz values are referred to, where Ra is the average roughness and Rz is the range between the highest peak and lowest valley. Rz is sometimes given to show the most severe feature in the measurement. Ra is often deceptive in this work as the roughness can change dramatically across defects. The roughness measurements are used in this work to measure a known feature. Table 11 lists the samples with images in Figure 132.

Table 11: List of samples that had their surface roughness measured

Sample	Surface quality
1	Contains a large coverage of microcracks (snakeskin), four clear surface breaking cracks and bright White Etching Layers
2	Pitted and some patches of grinding marks that have not worn away like the others, typical surface breaking cracks on the edges of the running band and a crack connecting some pits together
3	Fairly smooth other than lipped/ folded material on the edge of the break (could be post break damage). Typical surface breaking cracks though some of surface has shelled away.
4	Comparably little cracking. Clear signs of shallow grinding marks that end at the edge of the weld. Very smooth on weld area.
5	Pitted like Sample 2 with some grinding marks visible
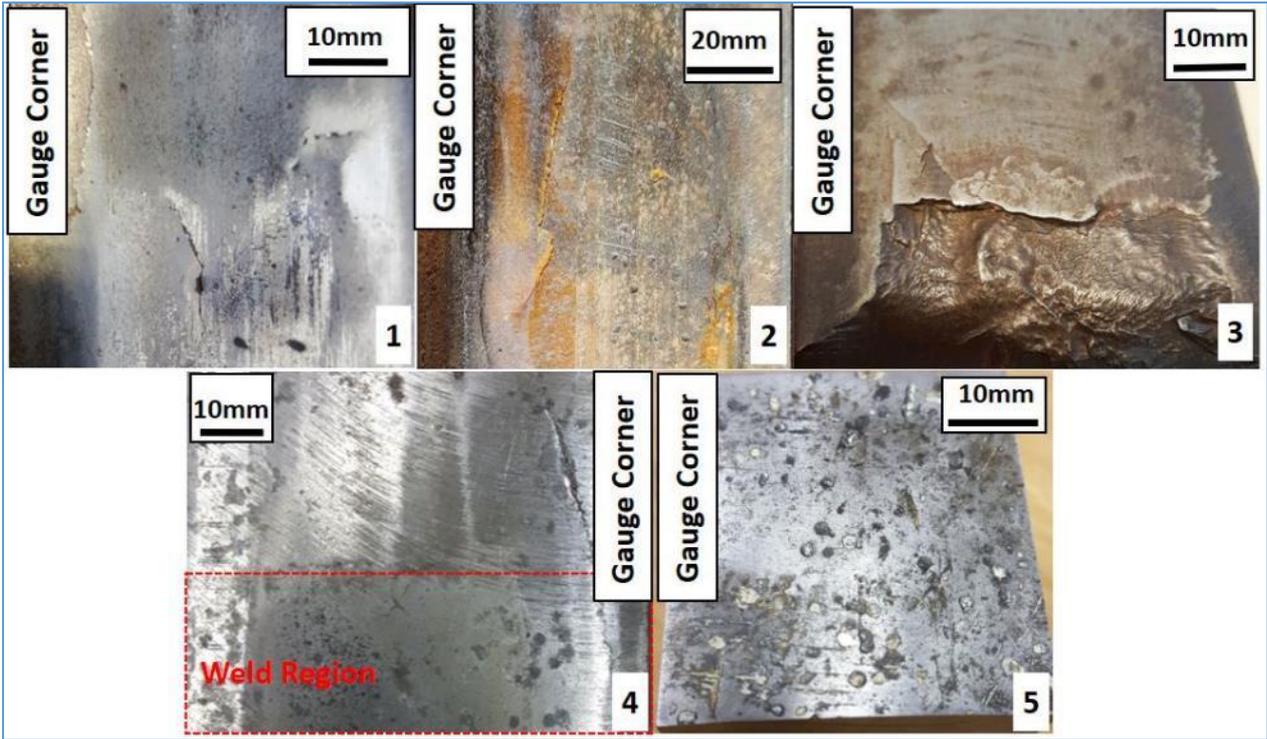
Figure 132: The surfaces of the 5 samples that had roughness measurements.

Due to the multitude of cracks on the surface of Sample 1, the entire Sample 1 region was scanned to produce a large surface map (Figure 133). The maps encompassed all surface cracks, the dark lobes, 'snakeskin' and the two dark spots believed to be 'ballast burn'.
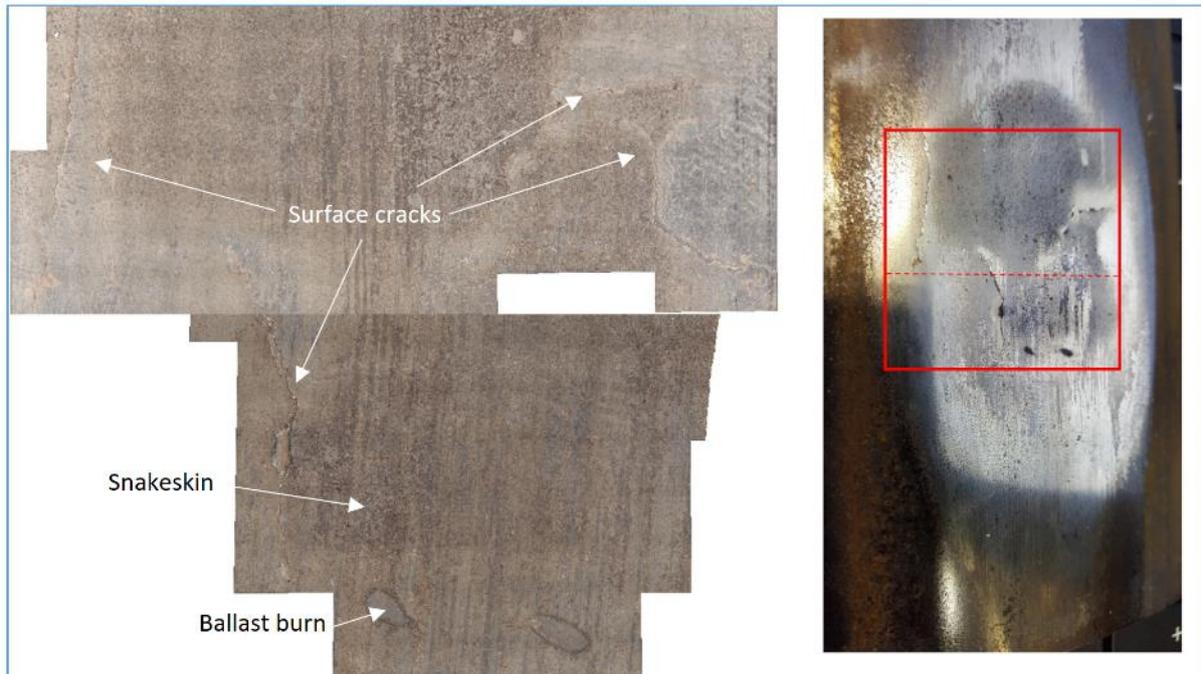
Figure 133: Left) Surface map of Sample 1 comprised of two scans stitched together. Right) The region of the rail scanned with the stitched region marked as a dotted line.

'Snakeskin' (Figure 134) is microcracks on the rail surface, often seen on low rails due to very high contact stress. WEL can often have a similar surface appearance. Ballast burn is when small debris from ballast stones become trapped in the contact and cause damage to the rail.

Figure 134: 'Snakeskin' on the surface of Sample 1

The dark lobes show where the surface has depressed due to developing subsurface cracks. They are dark because oxide can build up, because the wheel passes no longer remove the oxide layer as it has depressed out of contact. The cracks are seen on the boundaries between the darker and lighter areas, the brighter areas experience high contact stresses as they are the few places that the wheel contacts. These bright areas are also much smoother than the dark areas as shown by the roughness graphs below in Figure 135.
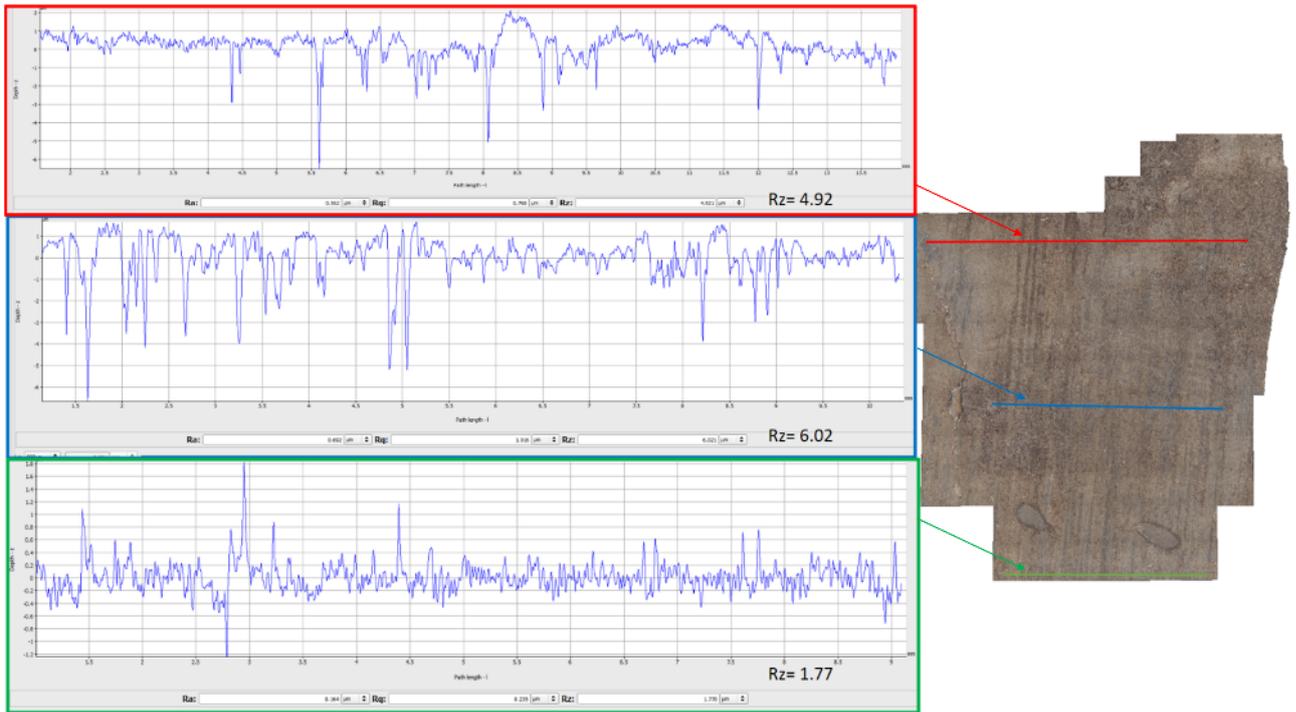
Figure 135: Roughness measurements along varied surfaces on the lower half of Sample 1. The Y-axis reads the depth in microns and the X-axis reads the distance along the line drawn. Rz is the mean roughness depth. Coloured borders around graphs match the coloured lines that the graph was drawn along.

Figure 135 shows that the roughness varies along the wheel's path, and there are some significant valleys (sharp downward peaks) within each measurement line. This could be due to roughness of wheels that passed before the region depressed, third body contaminants, uneven corrosion or most probably, a combination of all three. The oxide makes it difficult to assess the roughness of the surface pre-defect even though it has depressed to below the wheel contact zone.

Rz can be deceptive if the measurement line crosses rough and smooth regions. Figure 136 shows roughness measurements of the upper portion of Sample 1, but only across small areas to give better Rz values for lighter and darker areas. Using the values of Rz from Figure 135 and Figure 136, approximate Rz values are ~2 microns outside the defect, ~1 micron for the 'smooth' or bright areas within the defect and ~5 microns for the oxidised dark region. The value for the oxidised region was more for comparison, because it has little significance to the recent wheel rail interaction as contact is not made. However, the Rz values for the bright area

show that, for this sample, not only is the contact area severely reduced, but also the roughness is lower. This can negatively affect the friction/ traction of the train's wheel. Wheels are known to slip on defects due to poor contact areas, which is why axle accelerometers have been investigated for Squat detection.
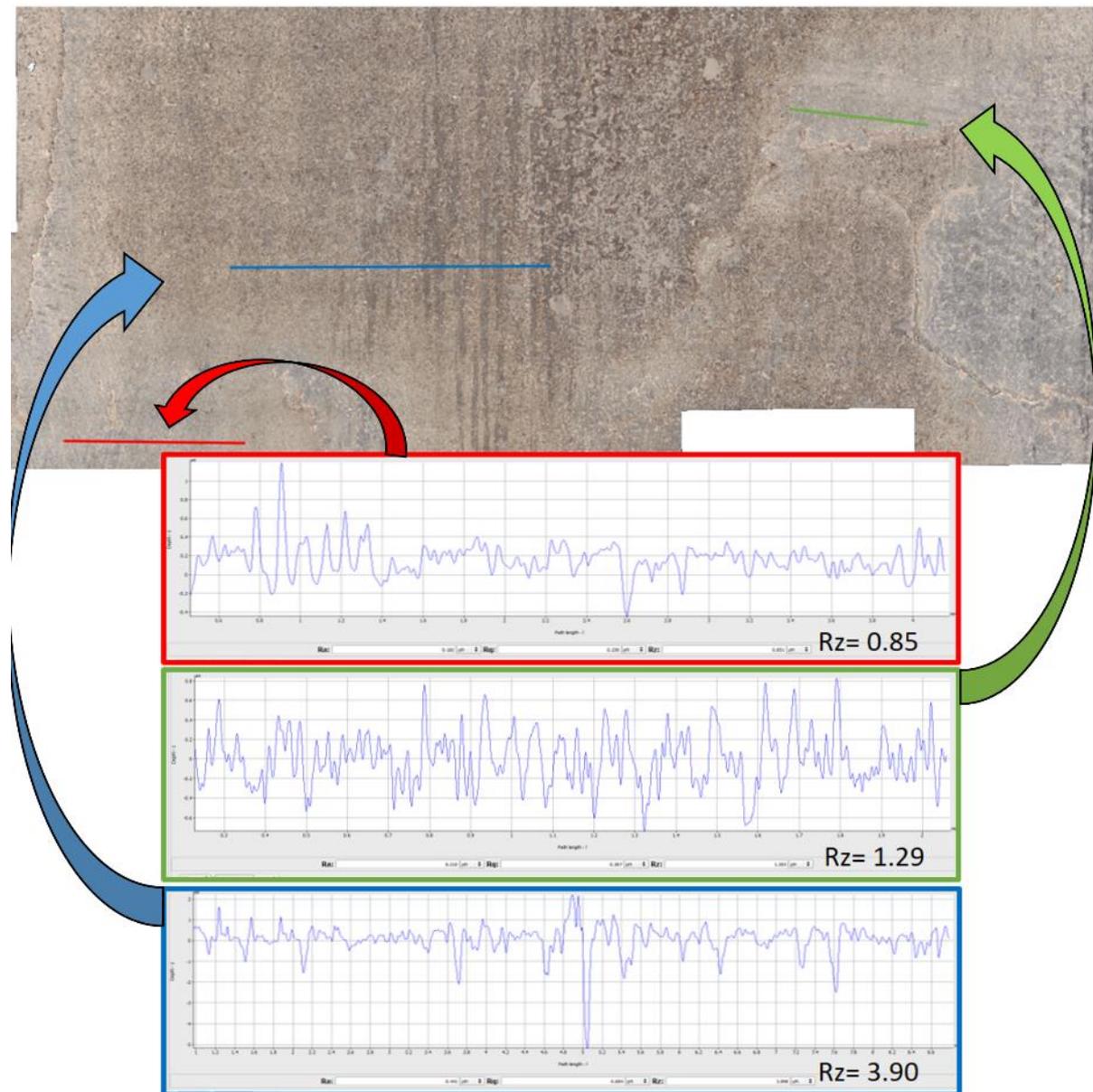
Figure 136: Roughness measurements across shorter areas for more representative Rz values.

The most striking surface features on Sample 2 were the numerous corrosion pits scattered across the surface, both inside and outside the contact band. There were also a few grinding marks still visible within the running band, although these were limited to the area above the defect. Figure 137 shows a photo and surface map of an area of interest in the centre of the defect, which was identified using the CT scan volumes. This area sits directly above a feature that will be discussed more later in this Chapter.
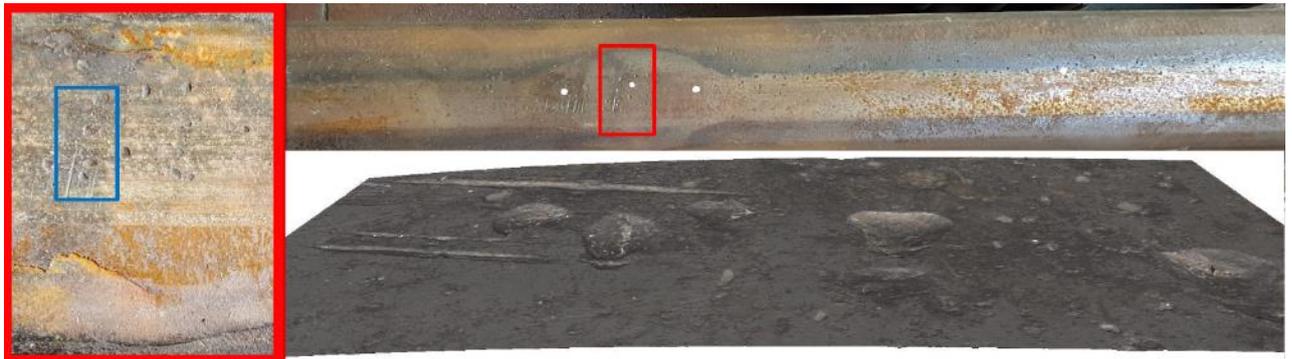
Figure 137: Sample 2. Above) Overview showing how clear grinding marks are. Left) Magnified image showing region of interest. Below) Surface map of region of interest rotated 90 degrees clockwise

Figure 138 shows the depths of the features of interest. The grinding marks are 50 and 30 microns deep. Considering that most of the grinding marks on the rail have been worn away, these marks were either very deep to begin with or the wheel made very little contact in these areas. Alternatively, perhaps some wheels made little contact and some made some contact, but the grinding marks were also deep. The importance of these grinding marks will be discussed later in this Chapter. The Rz value for the smoothest areas around the region mapped in Figure 137 and Figure 138 are ~6 microns, with even the lightest pitting increasing this value to 10 microns.
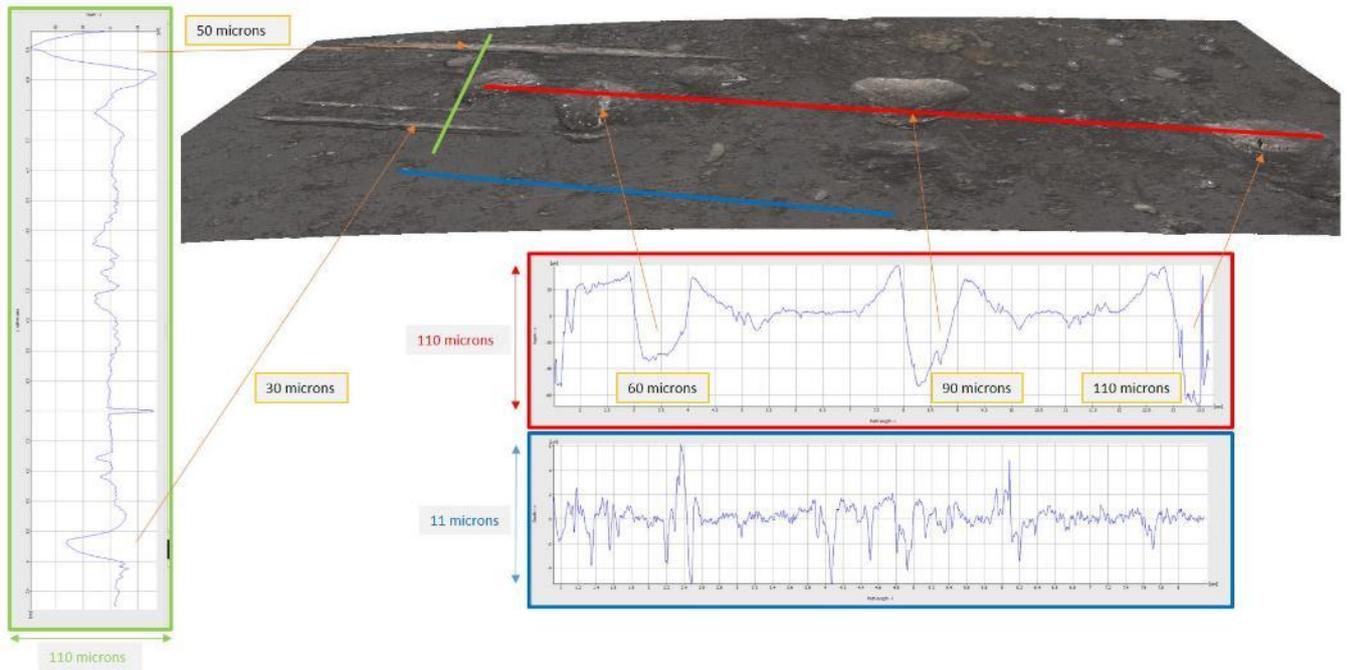
Figure 138: Roughness measurements obtained using an Infinite Focus Microscope. Orange arrows link the graphs to the image along with depth measurements and the double-headed arrows show the range of the depth scale.

6.3.1.2 Sample 3

The in-service breaking and shelling of Sample 3 left only a small section to scan. The focus was on the intact lobe, which had its upper surface removed to expose the crack plane below for scanning. Figure 139 shows the surface map for the crack plan that created the intact lobe in Sample 3. The other lobe was lost due to shelling but can be seen exposed in the inset image. Sample 3 shows the apex crack on the gauge corner that is associated with Squats, shown as a green dotted line in Figure 139. The apex of that crack was probably the initiation site.
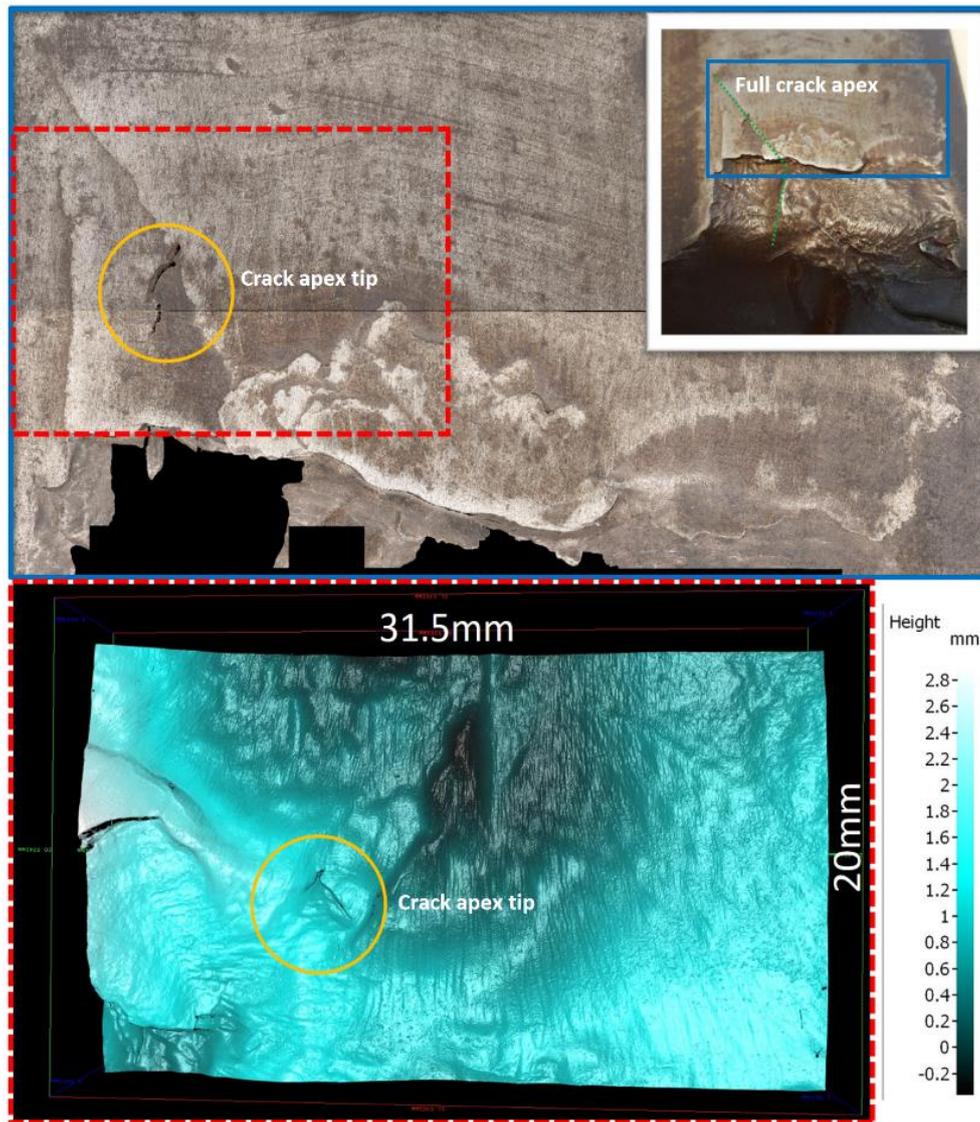
Figure 139: Surface map of the crack plan that makes up the intact lobe of Sample 3. Above) The surface of the rail. Below) The crack plane below the rail surface within the red dotted box. The orange circles aligns a common feature in the two images.

The remnant signs of grinding marks can just be discerned in Figure 139. Figure 140 shows the roughness of the region. Across most of the red line, the surface is smooth, barely deviating more than 5 microns. However, at each end of the red line, just before the two surface cracks, there are two dips. The wider dip seems to have shoulders, meaning the line rises on both sides before in dips. This suggests a third body particle became trapped in the contact. The narrow dip is attributed to the depth of the grinding marks still present, as ~15 microns was found to be the approximate depth of the grinding marks still visible on Sample 4 too. Defect free areas of the surface yielded Rz values of 3 to 4 microns.
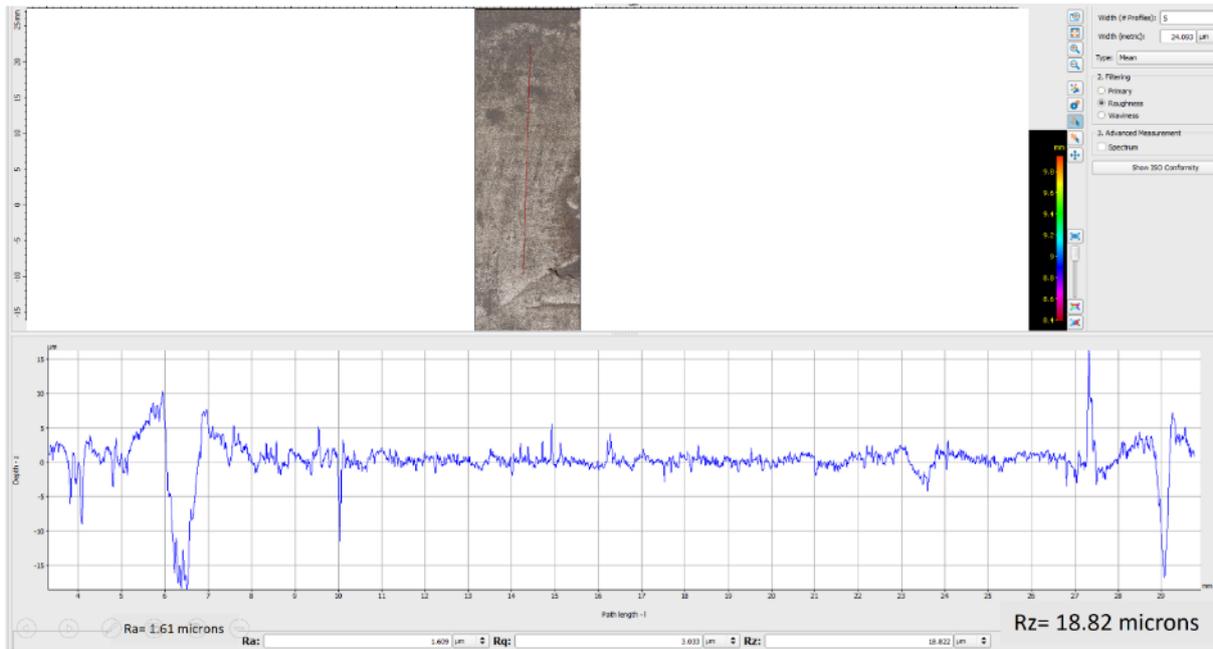
Figure 140: Roughness of Sample 3 just before the shelled region of the defect

6.3.1.3 Sample 4

The surface of Sample 4 consisted of both a rough area displaying grinding marks and a smooth area. The smooth area seemed quite consistent with the area above the aluminothermic weld. The initiation of the defect was on the border of the rough and smooth regions. As with one of the marks in Sample 3, the grind marks in sample 4 seem contained to within a 15 micron range (Figure 141). The transition from rough to smooth may have been a factor in a slip value change for the wheel passing over Sample 4. On the other hand, the smooth area could have occurred during to the development of the defect, perhaps due to decarb from the weld process. With very little information about where the sample came from, and the traffic it experienced, it is very difficult to be certain. The main value in the roughness measurement of Sample 4 is that it gives a reasonable value for the depth of typical grinding marks. Grinding mark depths obviously reduce with wear, more dramatically at first, but some range of numerical values needs establishing to determine if a grind mark is typical, or too deep.
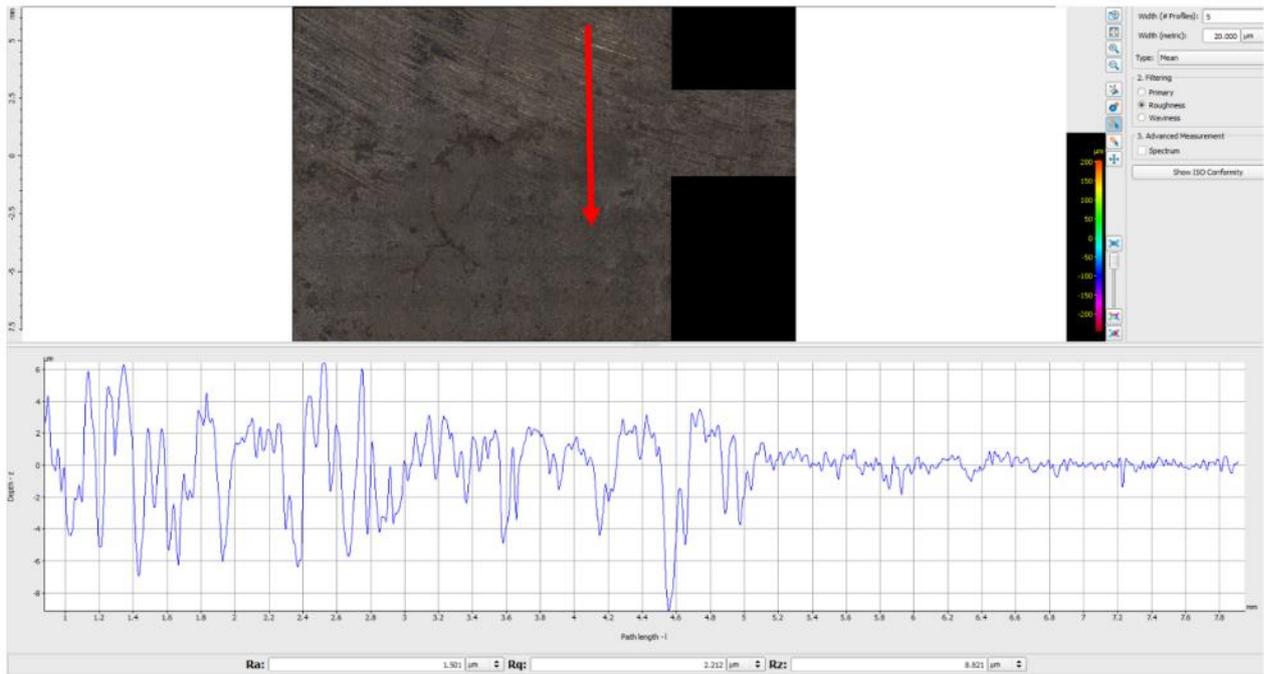
Figure 141: Roughness measurement for the transition from the rough region to the smooth region. Rz is 8.8 microns.

6.3.1.4 Sample 5

Being from the same railway line as Sample 2, Sample 5 shares both the pitting and the grinding marks, but the surface was mapped in more detail. The region mapped can be seen in Figure 142, with the Ra and Rz values in Table 12 and some of the graphs in Figure 143.

The pit's depth dominates Ra and Rz values when the measurement line includes pits. In measurement 5 there are two grinding marks of 20 and 30 microns deep, alongside a pit that seems to be ~35 microns deep, but looks ~65 microns deep if the shoulders are included. The depth of the features is of some interest, because shallow features should wear away faster than a deep one, and deep grinding marks have been associated with defect development in the literature. The width of the pits is also significant because it reduces the contact patch with the train wheel.

219
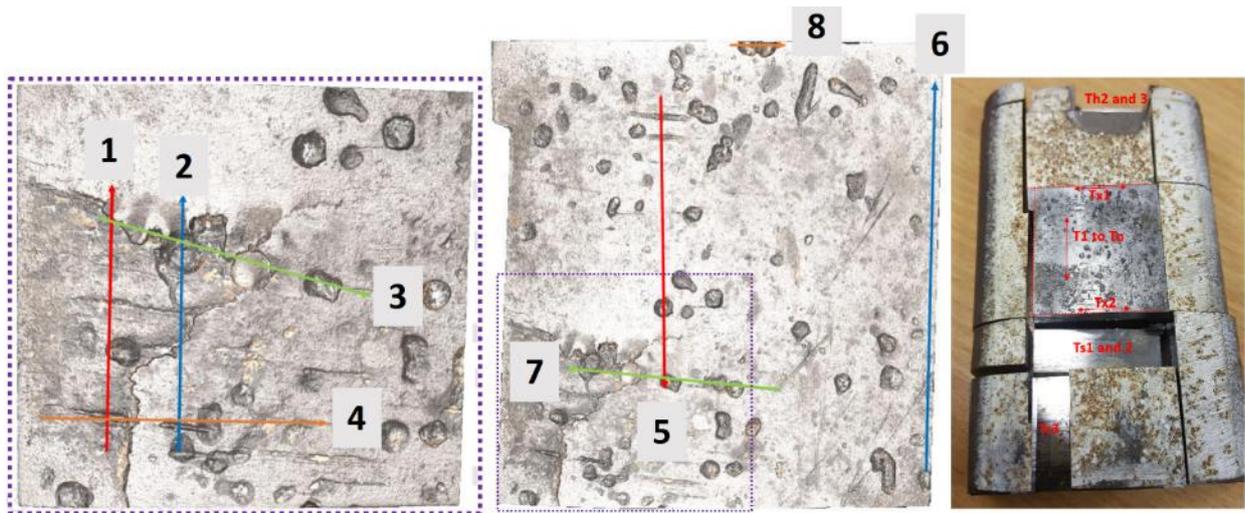
Figure 142: Locations of the measurements taken from Sample 5, with the bulk being from the centre of the rail and the initiation site indicated by the dotted box.

Table 12: Ra and Rz values for various measurements made on Sample 5, along with comparison values from Samples 3 and 4.

Image ref	Sample and features	Ra (microns)	Rz (microns)
1	Sample 5 grind	2.2	13.3
2	Sample 5 pits	10.1	47.6
3	Sample 5 pits	10.5	49.3
4	Sample 5 pits and grind	2.2	17.3
5	Sample 5 pits and grind	4.5	27.5
6	Sample 5 reference	1.2	9.1
7	Sample 5 pits	12.1	49.9
8	Sample 5 pits	7.2	34.0
	Sample 4	1.5	8.8
	Sample 3	1.6	18.8
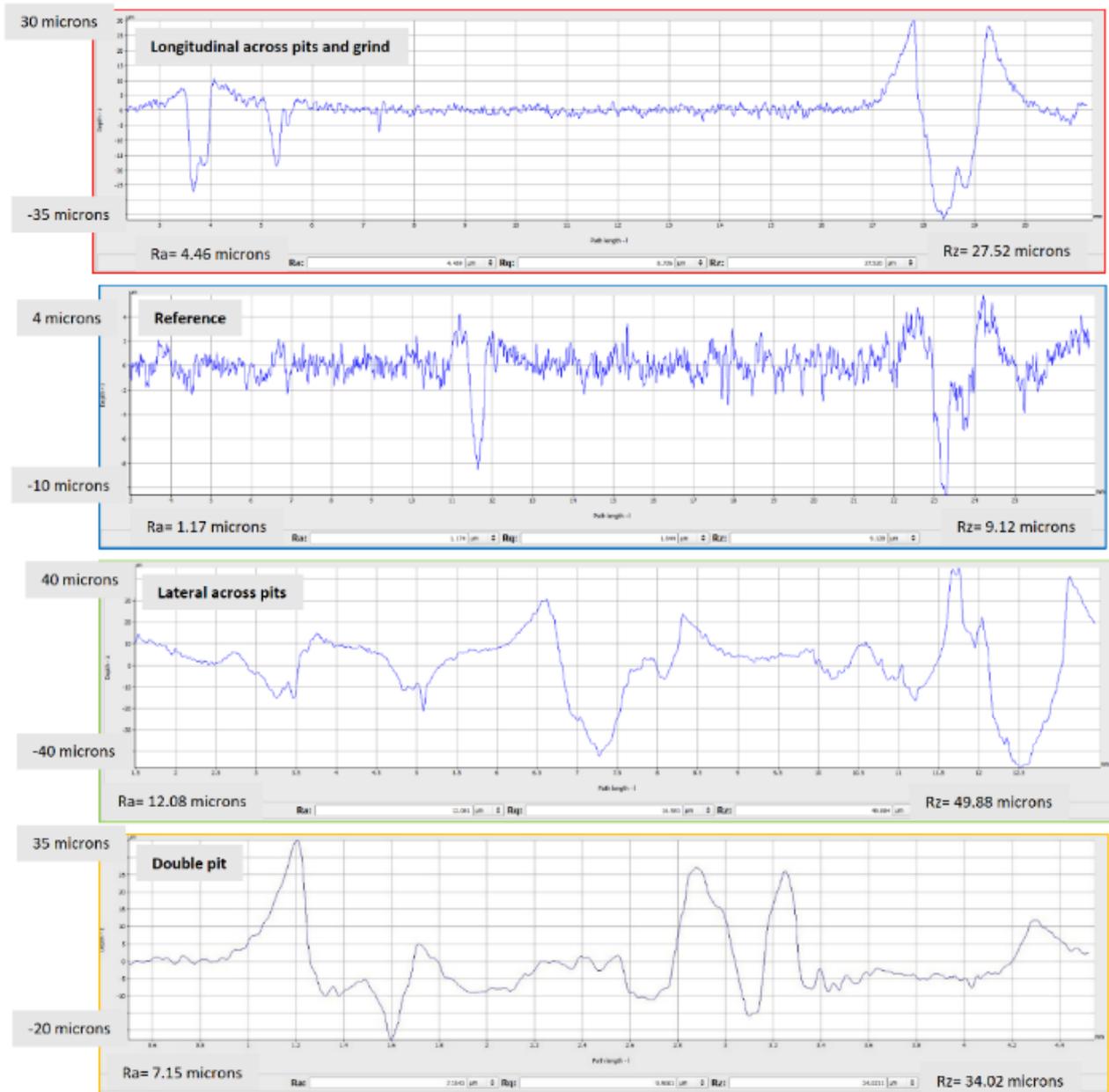
Figure 143: Roughness graphs for measurements 5 to 8, taken from Sample 5 (see Figure 142)

A cross section was taken along the line that measurement 1 (Figure 142) followed. The surprise was that instead of the grind mark being ~30 microns as the graph (Figure 143) had suggested, it was measured optically as 88 microns (Figure 144). An explanation for such as difference is that the process of cutting polishing and etching the sample removed the material that was filling the grind mark, making it appear shallower than it was. This means that the values for depth measured need to be considered as a minimum value, as they are probably also filled with debris that makes them appear shallow.
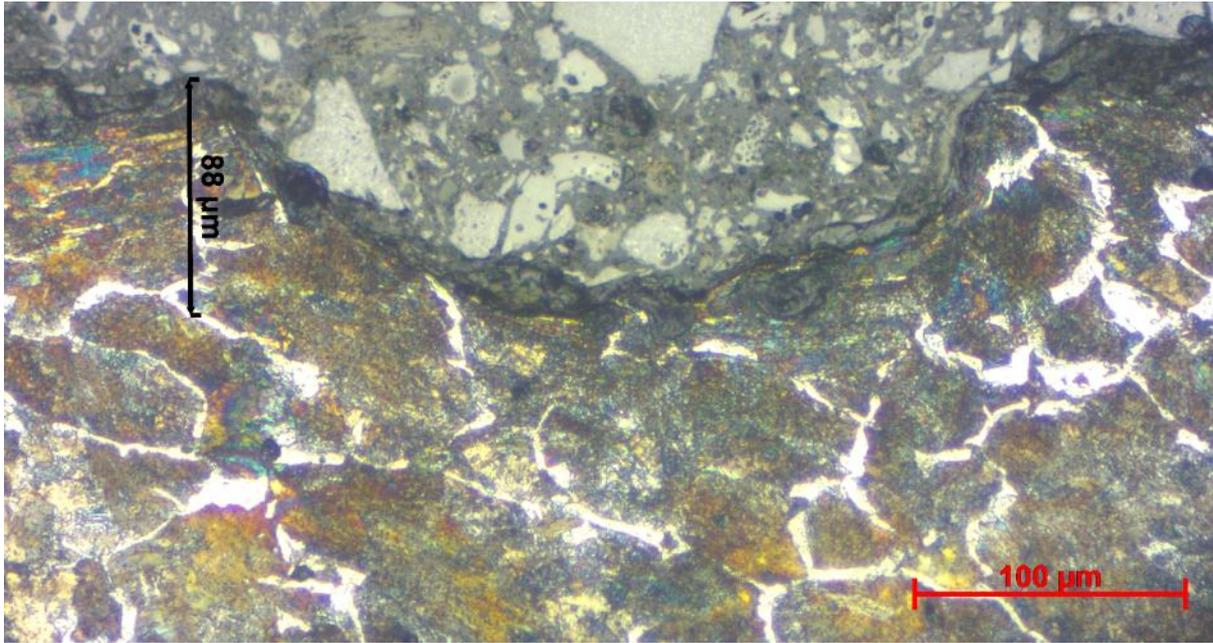
Figure 144: 88 micron deep grind mark found along measurement 1 path, believed to be the same grinding mark measured to be ~30microns deep.

6.3.1.5 Summary for surface roughness

Sample 1 was much smoother than other samples, partly because of the lack of grinding marks, but it is also speculated that a severe wheel slip event, heating the upper surface, would also affect the roughness. A wheel slip or slide is plausible because the sample is from an inclined track. Sample 3 was quite smooth, but still had visible signs of grinding marks, just like part of Sample 4. These grinding marks were found to be ~15 microns deep. In this work 15 microns is being considered a normal grinding depth, although the depth was probably slightly deeper when the defect initiated. Sample 2 and 5 showed deep grinding marks. Sample 2 has grind marks of 30 and 50 microns deep according to the scan. However, given that the 30 micron grind mark in Sample 5 showed to be 88 microns deep after metallographic preparation, the 50 micron deep grind mark could be much deeper. Checking the micrograph of the 30 and 50 micron deep grind marks in the same way as for Sample 5 showed that the grind mark was somewhere between 40-60 and 60-100 microns deep respectively, depending which edge the measurement was taken from. With the two grinding marks associated with the initiation site for Sample 2 and 5 containing grinding marks six times deeper than normal grinding marks, it becomes very likely that the grinding marks are at least partly responsible for the defects occurrence. This concept is explored further as this Chapter progresses.

6.3.2 Subsurface features

6.3.2.1 Sample 1

Sample 1 contained the expected leading and trailing cracks and only one branching crack. The cracks did not branch as much or travel as deep as Sample 2. An immediately interesting feature is the pit on the surface that burrows into the rail slightly, at an acute angle to the surface, in a lateral direction. This pit is seen on some of the orthoslices shown later. The pit may be caused by a 'snakeskin' scale breaking away from the surface followed by corrosion. Snakeskin is micro-cracking of the surface more commonly seen on low rails due to the very high contact stress. The depth of the pit may be due to the interaction with the crack plane.

Another interesting feature is the cylindrical crack that exists where the only major branch from the initial leading and trailing cracks are, seen in the longitudinal view of Figure 145. This cylinder probably began as a branch from the main crack before the close proximity of the two cracks caused a small crack to cross the gap, creating a circle/cylinder. An interesting question is which crack was present first. Looking at the longitudinal image in Figure 145, it seems reasonable that the upper crack existed first and the crack grew downwards. However, ignoring the surface breaking crack, the lower crack lines up smoothly with the other side, as shown by the dotted red line. Therefore, it is plausible that the lower crack developed first as a subsurface crack, and then the surface crack grew later, connecting the pitted region on the surface to the subsurface crack.
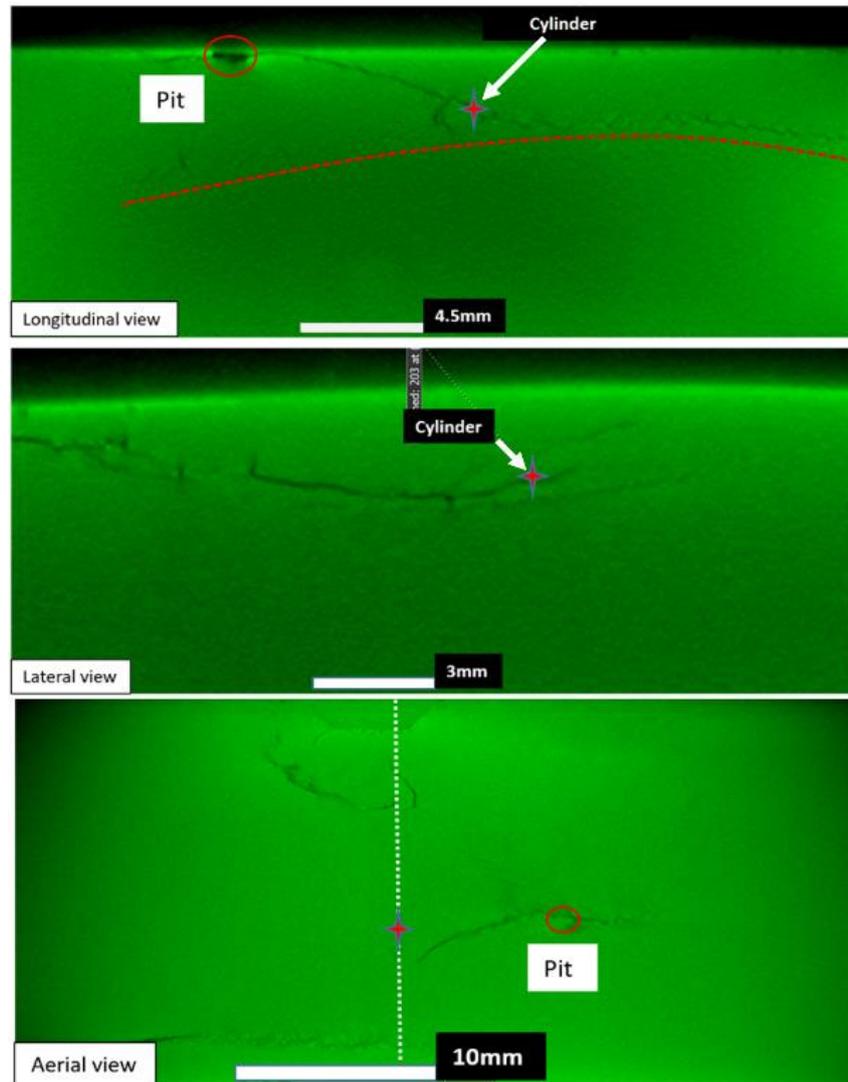
Figure 145: Orthogonal orthoslices from the scan of Sample 1 showing the location of the 'cylinder' feature (red star). The longitudinal view shows the 2D circle of the cylinder with a red dotted line highlighting the faint lower crack. The lateral view shows the cracks that make up two sides of the circle alongside an extra crack above. The aerial view provides a comparison of the location to the photo in Figure 156. The white line is related to the comment in the text.

Given that this defect is believed to be caused by a severe slip event, a subsurface initiation is plausible. Heavy decarburisation and plastic flow of the gauge corner of the rail in the same region as the defect supports the concept of a thermally caused initiation.

The slip will have created a heated upper surface that, when rapidly cooled due to the heat sink effect of the rest of the rail, caused a small crack to form parallel to the surface due to the

224

thermal shock. The lower crack contains a small crack parallel to the surface just to the right of the cylinder feature.

The lateral view shows how there is extensive cracking at various depths, which will allow some movement in the railhead. In the aerial view, it can be seen that if a line was drawn through the cylinder's location, 3 of the 4 surface cracks begin or end at that line. The line is sometimes referred to as the ridge between the two lobes of a Squat.

6.3.2.2 Sample 2

Scanning Sample 2 showed that the full defect was probably two defects that had interconnected, though the deeper defect was not very clear. The higher resolution scan of the deeper Sample 1 revealed its structure in detail. The leading crack, growing in the direction of travel, first seemed to have branched many times subsurface. The identification of two subsurface initiation sites made it more probable that there were two defects in very close proximity that later linked together.

The deeper defect contained a branch that looked like it was growing down into the rail. There was also a dark shadow that needed investigating. The dark shadow is a real feature, a result of multiple cracks that led to fretting: a combination of corrosion and rubbing that erodes the material in the area (Figure 146). It is clear from the upper image of Figure 146 how water runs down the slope into the shadow region, becoming trapped. The water would enter the crack network from the surface breaking cracks on the surface that do not show in the scan due to surface hardening effects. To investigate the region where the downward growing crack was, the sample was trimmed as specified in Table 10, isolating the deeper defect. The rescan of the smaller sample was able to distinguish between the multiple downward growing cracks.

The defect closer to the surface was the typical inverted V-shaped crack typically seen in Squat type defects. The deeper defect had two initiation sites. One was where the transverse defect (TD) developed from, a group of voids more to the gauge corner side, and the other was a vertical crack on the field side of the rail.

The TD crack was traced back to its origin, which was the dark shadow mentioned earlier and shown in Figure 146. The void shown was part of a cluster of small voids in a localised area.

The clusters produced a laterally orientated 'cylinder' of damage ~3mm under the surface and ~1.2mm across at its largest point, from which the other major crack planes propagated. The grind mark above the voids was first noticed when the sample was delivered. When comparing the voids locations to the grind mark, a correlation is noted. Figure 147 shows how the grind mark, noted as being deep in the previous section, has the same orientation and length as the voided region 3mm below it. The feature is offset in the direction of travel, but this can be explained by the distortion of the contact stress field for a rolling contact.
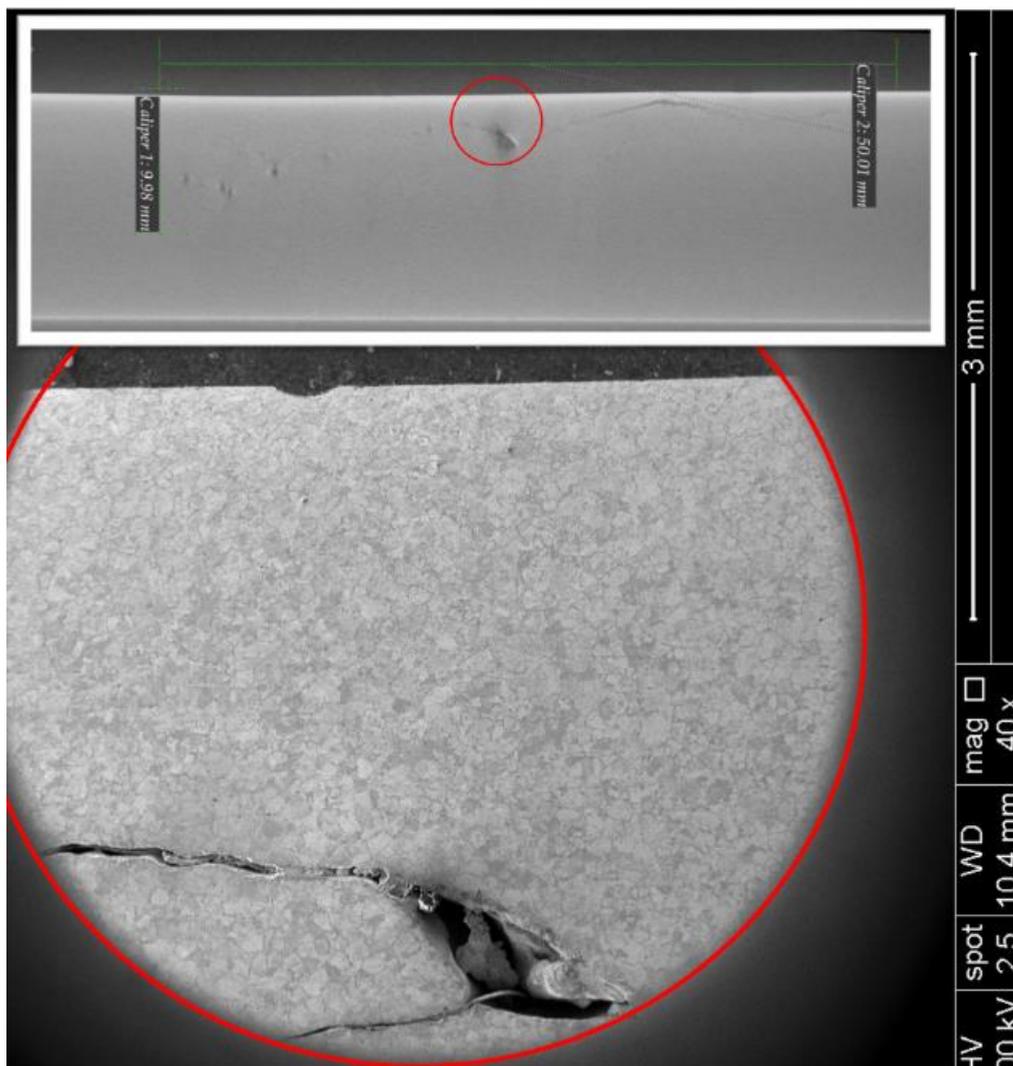
Figure 146: The cause of the shadow in the CT volume. Top) The CT orthoslice showing the shadow and faint downward branching cracks. Bottom) SEM image of the site showing the void caused by rail fretting and a grinding mark can be seen on the surface. The beginning of the TD crack can be seen leaving the bottom left corner of the void.
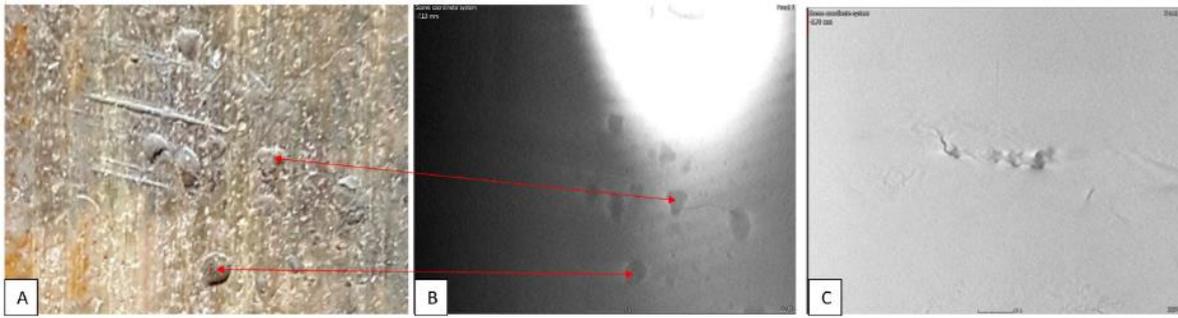
Figure 147: a) Photograph of surface, b) same surface as 'a' but from the scan with examples of the same pits linked by arrows, c) the same region as 'b' but ~3mm deeper into the rail

The vertical crack initiation site is believed to have occurred earlier, assuming both crack planes grew at a similar rate, because the crack planes meet at a distortion closer to the voids than the vertical crack. This initiation is a vertical crack that developed longitudinal cracks from its upper and lower crack tips (Figure 148). The vertical crack was identified as an initiation site because the region around the vertical crack was chaotic, and the crack path becomes more continuous and smoother further away from either of the two initiation sites. Vertical cracks can occur, although they are usually seen surface breaking. Therefore, cracks growing from the top and bottom of a vertical crack is much more plausible than a crack making two ~90 degree turns in quick succession, if at all.
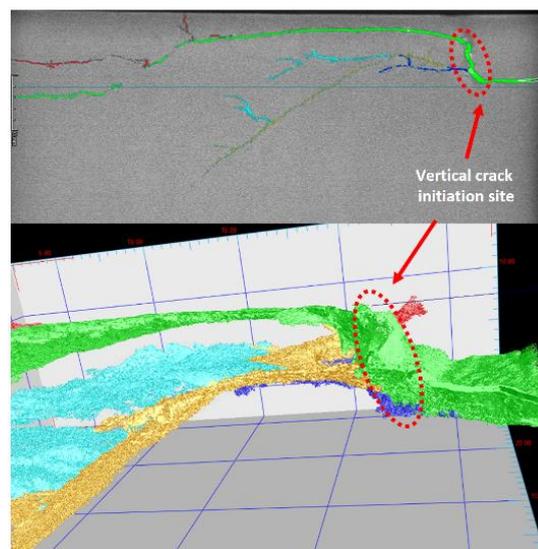
Figure 148: The vertical crack initiation site as a 2D and 3D image. The 3D image is isolated from the rest of the defect for clarity, with a scale in mm on the grid.

The vertical crack is close to the field side of the rail, suggesting a hollow wheel creating high contact stress. Hollow wheels were confirmed as an issue at the site where this sample was extracted. As for why the crack is vertical, there must have been a point in the microstructure that failed first.

In some areas, grains were noted to align their borders so that there is a line of ferrite running vertically into the material. Given the thicker ferrite layers on grain boundaries seen in this grade (R220) and the normality of cracks appearing on grain boundaries, it seems that the occurrence of this subsurface vertical crack was the result of a 'perfect storm'. That is, the high contact stress, probable slip event and the favourable orientation of the grains in a grade with softer grain boundaries.

Figure 149 demonstrates how the main crack plane (green) is still present in one region, but has worn away 1.5mm closer to the field corner, probably where the hollow wheel was contacting. Because the rail has worn more in this area, it makes it look like the original green crack is rising to the surface. Consider the 3mm depth that both the voids and the bottom of the vertical crack reside at. When looking at other areas of the rail where the crack is closer to the surface, the 3mm depth now coincides with the location of the transverse crack. This suggests that the first crack developed at a depth where the peak subsurface contact stress was, and then as the rail wore, the depth shifted with the new surface location, causing another crack to initiate at the same depth, but below the original crack plane. This situation was only observed on the field side, and is the explanation for the yellow downward growing crack (Figure 150), which was not as steep as the blue one, but was also a separate crack plane until the two eventually linked up in the later stages of growth.
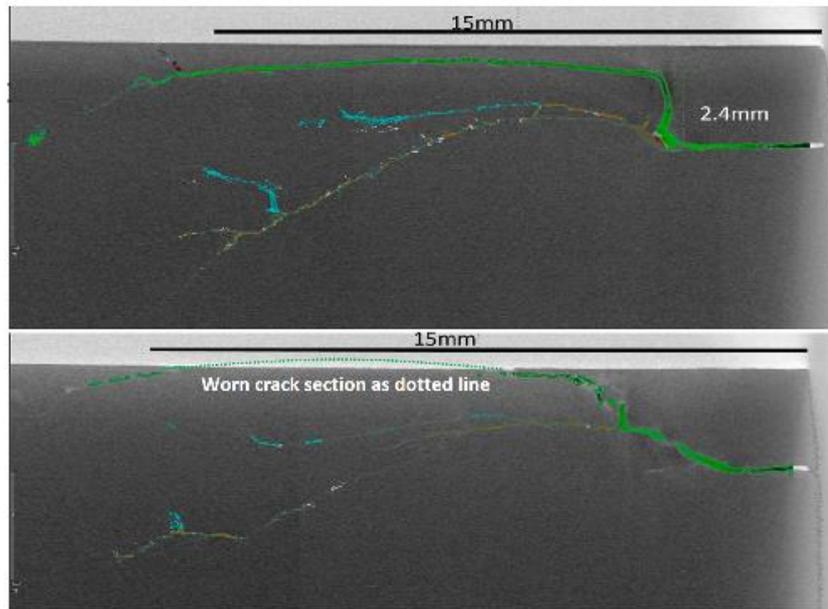
Figure 149: Above) CT orthoslice of the vertical crack that initiated the orange downward crack. Below) 1.5mm closer to the field corner, showing how the upper crack has been worn away
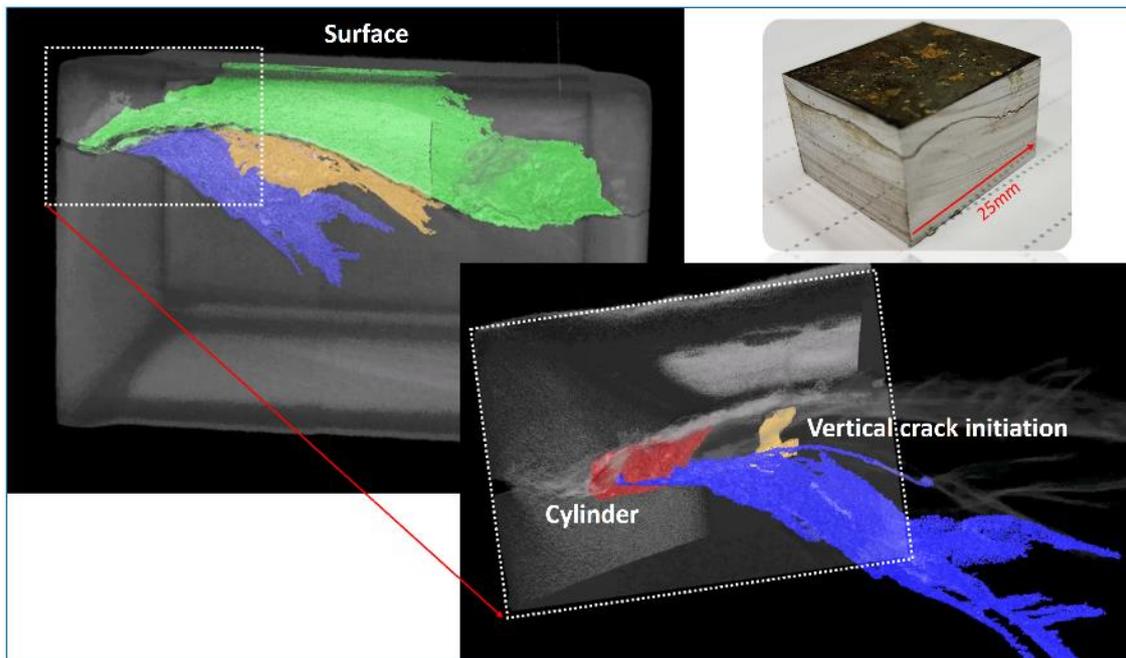
Figure 150: The red 'cylinder' of damage where the blue TD branches from the longitudinal green crack. The yellow crack grew from the vertical crack, which is also yellow in the lower image. The lower image shows the initiation sites in relation to the TD crack plane.

Figure 150 contains an inset image that shows the red and yellow initiation sites in relation to the blue TD crack plane. The voids are red and the vertical crack is yellow. The yellow crack plane can also become a transverse defect, but is of less focus because it is a similar depth despite probably initiating earlier, and is shallower in growth angle from the surface. The growth rate of the blue crack made it a focus.

The grinding mark has a strong correlation with the voids that the TD originated from, so it is likely that it is a strong factor in the initiation of the TD. The pitting found all over the surface seems to have had an influence on the growth of the overall defect.

The width of pits make them reduce the contact patch more than grinding marks. They are also deeper. This is expected as the corrosion in the bottom of the pits allows them to keep growing deeper whilst grinding marks are slowly worn. This means that in the early stages of the defect development, the grinding marks will have been more significant than the pits in terms of depth. Figure 151 shows the location of some of the pits that were above the high-resolution scan region. These pits seem to correlate with the crack growth route. This is believed to be because the pits caused subsurface stresses, damaging the subsurface material, making it an easier path for crack growth.
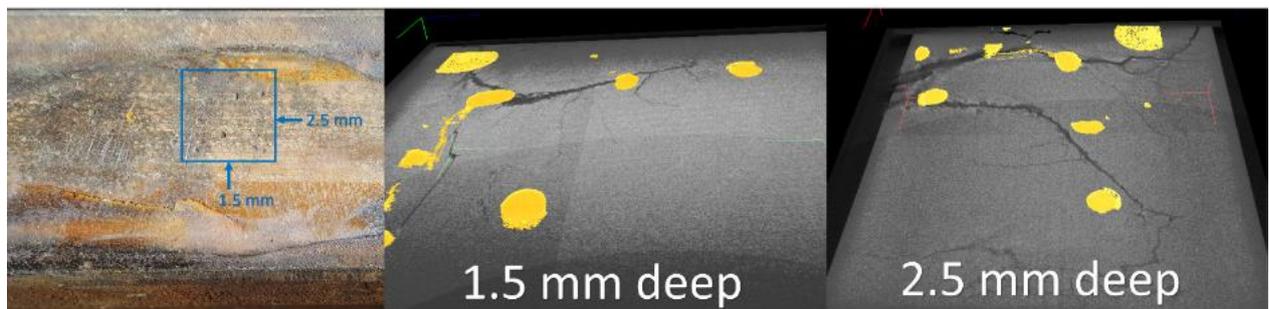
Figure 151: Pits are shown in yellow from the high-resolution scan of Sample 2, showing a correlation between the pits locations and the cracks growth path. The blue arrows show the direction of view for the two other images.

Sample 5 was taken from the same line as Sample 2 but is not as mature, providing an insight as to the development of Sample 2. Sample 5 has only developed the leading crack, the trailing crack occurs when the defect matures [9]. Initially the crack seems to branch out laterally and longitudinally from the same initiation but the CT volume shows that the cracks pass each other at slightly different depths, although they do link up eventually (Figure 152).
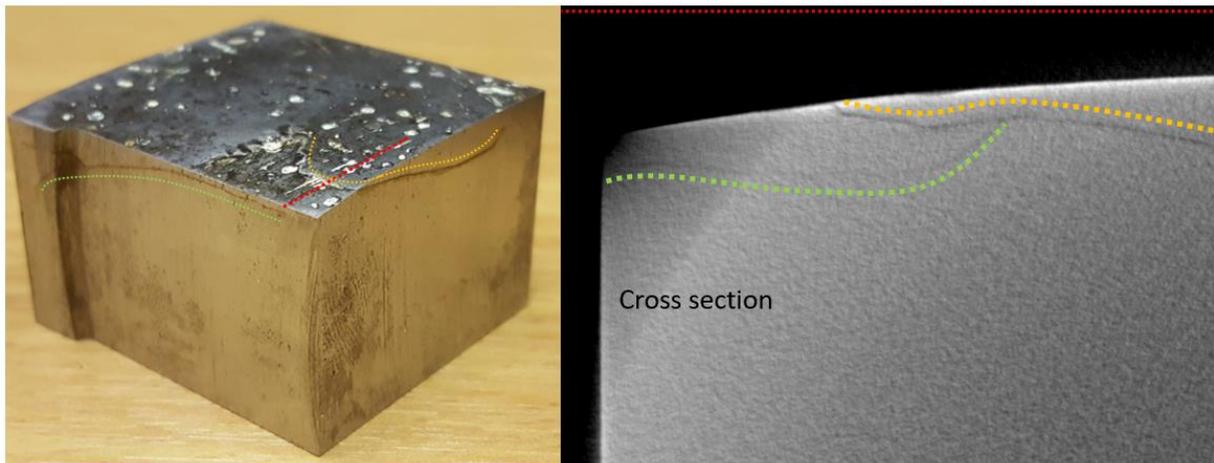
Figure 152: A comparison between the two cracks running longitudinally (green) and laterally (orange) with a CT cross section orthoslice. This shows that they are not the same crack plane even though they link up eventually.

Two initiation sites were present, as the typical conical growth of a defect in its early stages became distorted into a 'figure of 8', when the two cones meet as they grow down into the rail (Figure 153). One of these initiations is where a grinding mark and a surface crack cross (Figure 155). The other initiation is at the apex of the V-shaped crack that is often seen in Squats, although this one is on the field side rather than the gauge side. This has been seen before, although that defect was identified as a spalling defect [10]. The circular surface crack means that some wear has occurred since the apex crack initiated, as the tip of the cone has been truncated [3].

The typical growth from the apex initiation is distorted slightly by the initiation on the grinding mark, the same distortion that causes the 'figure of 8' in Figure 153. However, one of the crack planes (green in Figure 152) appears as a subsurface crack parallel to the surface like in sample

1, which was caused by thermal input. There could be a combination of subsurface and surface initiated cracks in the same defect. Just like Sample 2 but much closer together.
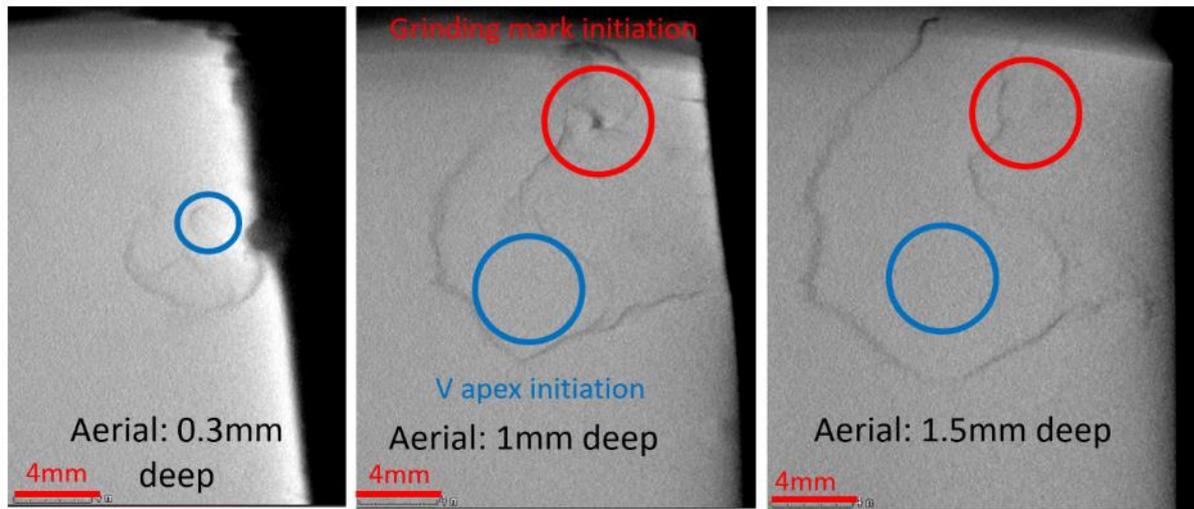
Figure 153: The two initiation sites in volume T, at different depths in aerial view showing the 'figure of 8' forming. The cross section shows how the cracks link up as they get close. Left image does not show all of rail due to the rail surface curvature.

Sectioning Sample 5 in a region where a shadow was found in the CT scan revealed a familiar feature. The shadow was caused by the branching of the crack network in multiple directions, resulting in the release of material that was lost when sectioning the sample (Figure 154). A crack can already be seen growing vertically from this cluster of damage. If the vertical crack breaks the surface, corrosion can occur in this void, allowing the same rapid crack growth seen in Sample 2. Sample 2 growth is considered rapid because it is clearly much more developed, despite experiencing exactly the same traffic conditions as Sample 5. This branching feature was found directly below an elongated pit and a circular pit, i.e. a severely reduced contact area (purple circle in Figure 155).

Figure 154: Branching of the cracks and loss of material in Sample 5. Probable precursor to voids seen in Sample 2.

Figure 155: IF microscope map of the surface of Sample 5 showing the two circled initiations. The image is upside down compared to the CT viewer in Figure 153. The purple circle shows the location of subsurface damage shown in Figure 154. Scanned area = 40x40mm.

6.4 Discussion

Trying to determine the cause of the origin for the transverse crack in Sample 2 was the reason for investigating connections between the subsurface and the surface features. Subsurface features were noted and linked to surface features, made possible by the CT volumes produced for Sample 1 and 2. Sample 2 shared many similar features to Sample 5, which was expected because they are from the same railway line. There was a similar feature to the voids in sample 2 found in Sample 1. This feature was similar in width and length but was found in a Stud. Sample 2 and 5 are believed to be grinding induced Squats (GIS) due to their strong connection to grinding marks at their initiation sites.

6.4.1   'Cylinders' of damage

It is possible that the 'cylinder' in Sample 1 was simply caused by the branching cracks passing close enough to each other that a smaller crack connected them during the defects growth. The 'cylinder' is not believed to have been the first initiation point in Sample 1: the first initiation is believed to have been on the field side of the defect for reasons explained previously (Figure 9 in Chapter 4). Looking closer at the cylinder after metallographic preparation shows that one of the upper cracks of the cylinder has experienced high shear forces. The crack faces had rubbed together with enough intensity to cause plastic deformation as documented in Chapter 5 using nano-indentation. The area of deformation is on the upper edge of the cylinder.

The cylinder of damage in Sample 2 was made up of clusters of voids. These voids were created by rail fretting, caused by corrosion getting into the crack network and oxidising the branched region. It is believed that this rail fretting is responsible for the accelerated crack growth and TD. Rail fretting is similar to a phenomenon seen in other metals, called Stress Corrosion Cracking (SCC). SCC is when a material would not fail under fatigue or corrosion separately, but fails when they exist together. In this case, the corrosion seems to have accelerated what the fatigue alone was causing at a slower pace.

Comparing the cylinder feature in sample 1 against Sample 2. The Sample 1 'cylinder was ~2mm (to its upper edge) below the surface and ~2mm across at its largest point. The sample 2 'cylinder' was ~2.5mm (to its upper edge) below the surface and ~2mm across at its largest point. Differences in depth could be due to differences in wear rate. Sample 1 is a slightly

harder grade at ~260HV (R260 grade) whereas sample 2 is ~220HV (R220 grade). Both samples are known to have been from a downward incline so wheel slip and braking are possible in both cases, which can produce very high temperatures in the wheel-rail interface (WRI) [11]. Unfortunately the amount of traffic that passed over Sample 1 is unknown as its location was not recorded when it was extracted and delivered. Sample 2 had a very low traffic total of 11MGT. A classic Squat would have taken at least ~100MGTs to become cause for concern, which is still a reduction of a rail's ideal lifespan. Given enough time, the cylinder in sample 1 may have broken down to resemble the void in Sample 2, although the cluster of cracks in Sample 5 is a more likely predecessor. Sample 5 shared a similar feature that resembles the voided region in Sample 2, but in a less degraded state. There was only the leading crack present and water could not enter the branched region easily due to a lack of surface breaking cracks in the area.

It is difficult to be certain which of the initiations in Sample 5 came first, although the larger spread of the V-shaped initiations cracks apex (Figure 153) sways opinion towards it being first. The interest in the cylinder is its similarity to the cylinder/ fretted region of Sample 2 that was the source of a potentially dangerous TD. Both features are ~3mm deep in the rail (variation could be attributed to different wear rates of the surface since development) and both cylinders are ~1.2mm across at their largest point. This could be coincidental or it could be a sign that the development of enough cracks (i.e. the cylinders) in a defects growth can lead to branching that becomes a TD. Both cylinders sit below a damaged surface that results in higher subsurface stresses. Therefore, if these cylinders are responsible for TDs developing, surface quality is a crucial component of preventing TDs.

6.4.2   Surface damage

Figure 156 compares the surfaces of Samples 1 and 2, showing the common widening of the band due to the subsurface crack network and the surface breaking cracks on the field and gauge sides of the rail. Both samples have surface roughness, but from different causes. Sample 1 has 'snakeskin', which is a network of shallow microcracks that makes the surface look scaly. The small pit that sits on a crack on Sample 1 is believed to have started from the flaking away of one of these scales. Sample 2 and 5 have similar pits and grinding marks as they experienced the same traffic environment and maintenance.
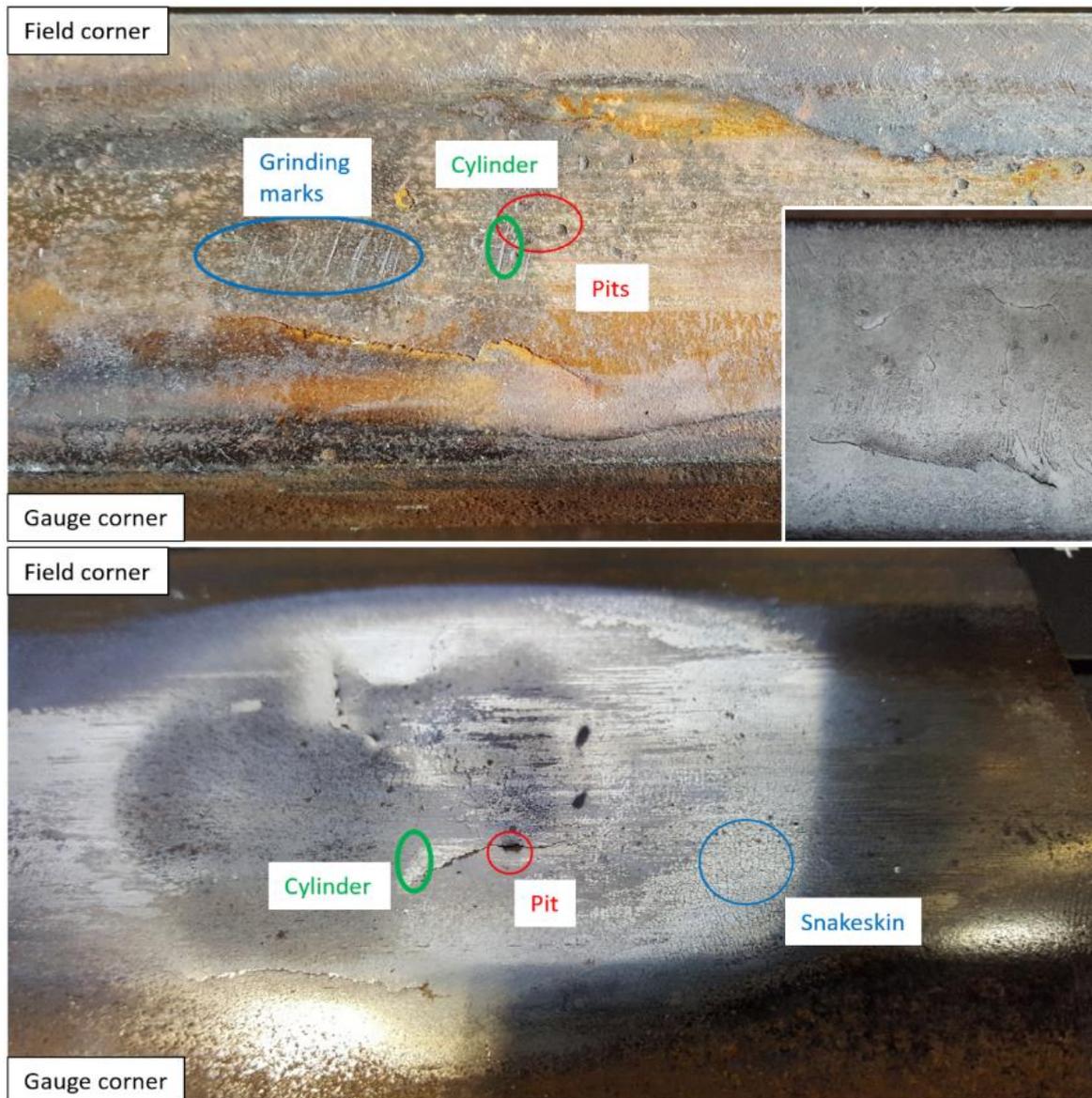
Figure 156: Above) Sample 2 with surface damage highlighted and a contrasted image inset to show the cracks. Below) Sample 1 with 'snakeskin' cracks and the pit visible subsurface in Figure 145.

6.4.3   Grinding damage

The roughness maps presented show that some grinding marks and pits are quite deep, even though it seems wear has removed many of the other grinding marks. This reinforces the principle that a good surface finish after grinding is important. Samples 2 and 5 show grinding marks that were deep compared to the others that have worn away (Figure 147 and Figure 155). Deep grinding marks were in the vicinity of an initiation site in Samples 2 and 5. Deep refers

to in the range of 50-100 microns compared to many grinding marks being measured in the ~15 micron range.

The presence of pits and grinding marks increase the roughness of the surface and a rough surface can increase the subsurface stress in a contact (such as that between a wheel and a rail) by a factor of 8 [12]. This could explain how the 'cylinder' of damage in Sample 2 occurred, as one of the deep grinding marks is directly above and is in the same orientation as the 'cylinder' of damage (Figure 147). This means that the deep grinding mark is probably responsible for the extensive subsurface cracking, which leads to further downward branching of cracks, one of which may have broken the rail if not detected and removed. Sample 2 also had another initiation, believed to have occurred first due to the hollow wheels that were acknowledged as being a problem on that railway line. This initiation developed its own downward growing crack but at a shallower angle with a slower growth rate. The two downward growing cracks linked together at some point in their development.

Grinding does not just produce valleys that reduce the contact area, they can also produce hills that are higher than the rail profile (Figure 157). These raised hills are compressed into the rail by the high pressure of a passing wheel if it were in the running band. As discussed in Chapter 2, this pressing of ground material into the surface of the rail creates residual stresses and initiation sites for cracks.
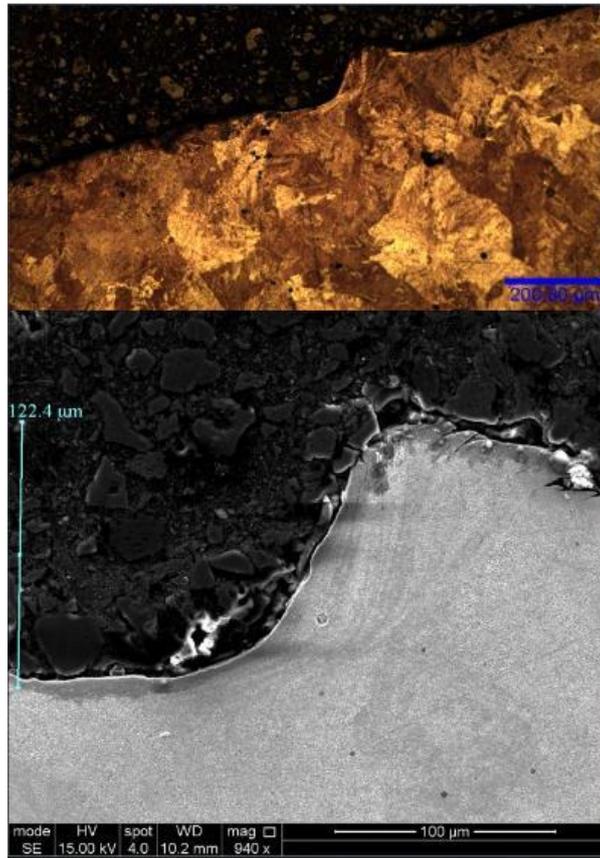
Figure 157: The hill created by grinding on the field corner of Sample 4. The peak of the hill has multiple cracks within it.

6.4.4 Pitting

Unlike grinding marks that are expected to become shallower as wear occurs, a pit can become deeper if the corrosion rate outpaces the wear rate. It is most probable that the pits occurred after the grinding marks were created, as grinding would remove any pre-existing pits. How much later the pits appeared after the grinding and at what point in the lifetime of the defect's development is uncertain. The presence of the pits over the fretted region subsurface in the rail is the reason behind their investigation, to try to determine what caused their presence. Because the distribution of the pits seemed to align with the crack growth path. For many of the pits in the CT volume, branching was noted below the edges of pits (Figure 158). Pits being above branches may be coincidental. More samples with pits would need investigating. A full volume of Sample 5 was not available but the worst subsurface damage was below quite a long pit (Figure 154). The detailed work on the features of the pits is presented in Chapter 5.
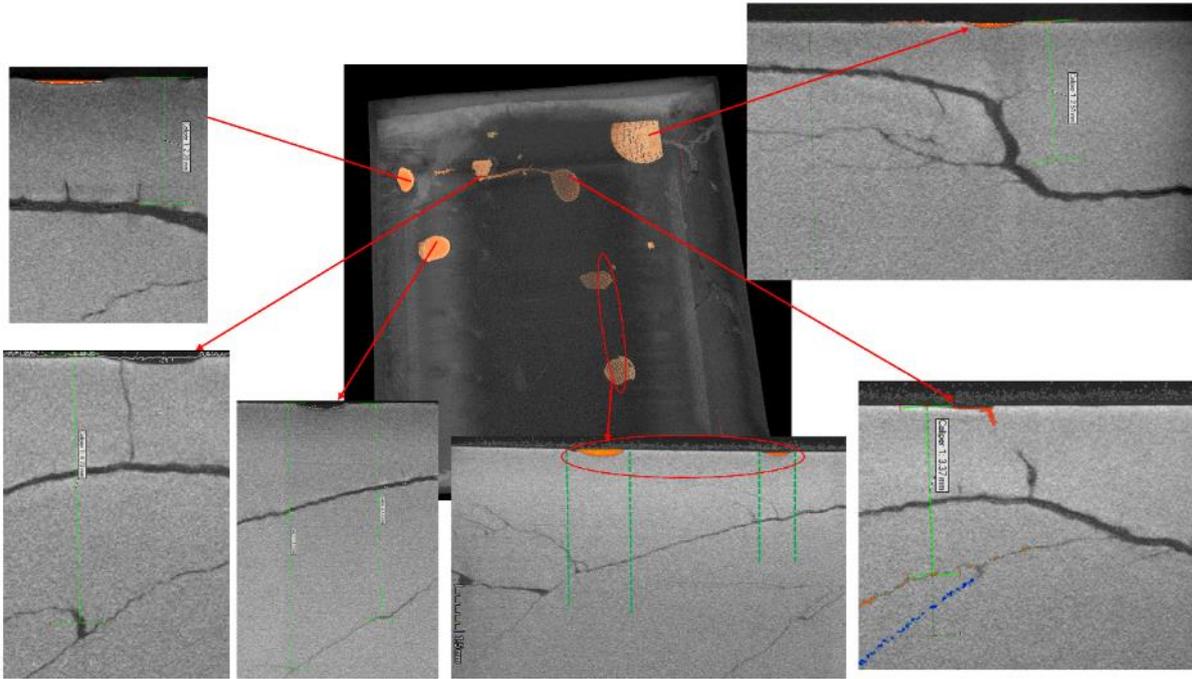
Figure 158: Locations of pits and the corresponding longitudinal view from Sample 2.

6.5 Conclusions

The use of micro-CT scans provide a rare opportunity to effectively make a comparison between features on the surface to what lies directly beneath them non-destructively. In this Chapter, grinding marks that are still visible, even after most other grinding marks had worn away, were identified.

A region of cracking and fretting, that created small voids, was the source of the cracks branching in Sample 2. These voids correlated to a deep grinding mark (60~100 microns) on the surface of the rail directly above the source. The grinding mark was the same length and orientation as the voids, suggesting that the grinding mark was a likely cause of their initiation. The voids were the origin for a dangerous TD. Sample 5 also had a grinding related initiation (88microns deep) and Sample 1 had a similar but less corroded subsurface feature to Sample 2. There were also many pits on Samples 2 and 5. These would have developed after the grinding marks but are seen to have influenced the crack morphology of the defect, possibly being a contributor to the accelerated growth of Sample 2. Sample 2 and 5 are GIS defects, as opposed to the initial categorisation of Squat defect.

There were also a number of surface pits just before the grinding mark. These pits lined up with each other in such a way that they would have drastically reduced the contact area of the wheel. This reduction in contact probably allowed a wheel slip to occur in the already damaged region. The reduced area will also increase the contact pressure as shown by the surface cracking between two of the pits within the running band. The wheel slipping as it passed over the stress intensifying grinding mark led to high subsurface stresses, subsurface cracking, fretting and eventually branching of cracks into a TD. The position of the pits correlate with some subsurface crack branches and the general growth pattern of the crack plane. Between the deep grinding marks and the pits, the importance of a smooth running surface is demonstrated, via the damage that occurs when the surface is rough.

Sample 2 and 5 had similar surface features but Sample 5 was less developed. Sample 5 had two initiation sites very close together whereas Sample 2 also had two initiations, but on opposite sides of the railhead. The vertical crack initiated first on the field side in Sample 2, as shown by the middle of the crack having worn away. Alternatively, the wear rate on the field side may have been higher than other areas of the running band. These scenarios suggested the presence of hollow wheels, a factor that was confirmed by the track operator. This demonstrates the importance of good wheel-rail profile conformity.

Sample 1 did not show a grinding mark above the 'cylinder' feature but there was a pit and surface breaking crack in close proximity. The pit is believed to have developed from the presence of 'snakeskin'. Snakeskin is usually found on low rails, but it is unknown if the sample was from a high or low rail. Given that the initial crack plane in Sample 1 was subsurface, with a surface breaking branch connecting the surface to the cylindrical feature, it is very likely that this defect was a Stud defect.

The 'cylinders' in both samples were both ~2mm wide at their widest point despite having different surface features. The depths of the features both lay between 2 and 2.5 mm deep below the surface, though this is a hard value to legitimise because wear of the surface affects this value. Considering the new information gathered and discussed in this Chapter, some of the defect identities have changed from the initial table in Chapter 4, as shown in Table 13.

Table 13: Defect classifications

Sample	Classification	Cause
1	Stud	High contact stress
2	GIS	High contact stress
3	Squat	Inclusions
4	Potentially Squat or Stud	Contaminated weld
5	GIS	High contact stress

Grinding is necessary as a safety critical maintenance tool for railways in order to ensure failure occurs by wear rather than fatigue (the Magic Wear Rate). Nevertheless, the surface finish needs to be as consistent as possible to avoid high surface and subsurface stresses from arising. The grinding marks that are believed to have been directly related to the defects initiating in Samples 2 and 5 were both ~60+ microns deep compared to ~15 microns on Sample 4. Pits must have developed later than the grinding marks, but they also seem to have a correlation with the crack growth in its later stages. So pitting can continue what grinding marks initiate. New grinding trains were recently put into service in the UK, so this may improve the situation. Milling provides a very good surface finish but the cost is higher.

6.6 References

[1]   M. Naeimi, Z. Li, and R. Dollevoet, "Nucleation of Squat cracks in rail, calculation of crack initiation angles in three dimensions," *J. Phys. Conf. Ser.*, vol. 628, p. 012043, 2015.

[2]   Y. Zhou, X. Zheng, J. Jiang, and D. Kuang, "Modeling of rail head checks by X-ray computed tomography scan technology," *Int. J. Fatigue*, vol. 100, pp. 21–31, 2017.

[3]   S. Earl, K. E. Rankin, R. Lewis, L. Smith, and W. M. Rainforth, "Comparison of Squats and Studs from different traffic environments," in *Proceedings of the 11th International Conference on Contact Mechanics and Wear of Rail/wheel Systems, CM 2018*, 2018.

[4]   Z. Popovic Radovic, V., "Analysis of cracking on running surface of rails," *Gradevinar*, vol. 65, no. 3, pp. 251–259, 2013.

[5]   S. L. Grassie, D. I. Fletcher, E. A. Gallardo Hernandez, and P. Summers, "Studs: a Squat-type defect in rails," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 226, no. 3, pp. 243–256, 2011.

[6]   S. L. Grassie, "Studs and Squats: The evolving story," *Wear*, vol. 366–367, pp. 194–199, 2016.

[7]   A. Ekberg, "D3.1- Enhanced track structure- Status, key influencing parameters and prioritised areas of improvement v.6," 2018.

[8]   M. Steenbergen, "Rolling contact fatigue in relation to rail grinding," *Wear*, vol. 356–357, pp. 110–121, 2016.

[9]   M. Steenbergen and R. Dollevoet, "On the mechanism of Squat formation on train rails – Part I : Origination," vol. 47, pp. 361–372, 2013.

[10]   M. Steenbergen, "Rolling contact fatigue: Spalling versus transverse fracture of rails," *Wear*, vol. 380–381, pp. 96–105, 2017.

[11] C. Bernsteiner, G. Muller, A. Meierhofer, K. Six, D. Kunstner, and P. Dietmaier, "Development of white etching layers on rails: simulations and experiments," *Wear*, vol. 366–367, pp. 116–122, 2016.

[12] A. Kapoor, F. J. Franklin, S. K. Wong, and M. Ishida, "Surface roughness and plastic flow in rail wheel contact," vol. 253, pp. 257–264, 2002.

Chapter 7: Discussion

The overall aim of this project was to enhance the understanding of Squat defects using micro-CT scanning. This approach was used because it was uncertain where the best information was held within a defect, so a full visual representation of the defect was needed to highlight areas of interest. This then unravelled some of the crucial and sometimes unique features of each defect. Coupling the CT volumes with traditional metallographic methods allowed the classification of each defect, which was often not the original defect classification before the scan and sectioning.

Big data analysis can be a powerful tool for bringing together many factors and making use of the large amounts of track data that inspection trains and teams gather. However, if there are various causes of Squats, more precise terms are needed for the defects, such as Studs and GIS, to ensure the data is representative. A strong agreement as to the nature of a Squat type defect and their classification needs to be established and implemented on track in order for the main influences on Squat development to become clear.

Some strong correlations have been found in the literature, such as 70-75% of Squats being found on sleepers rather than between, 70-75% of Squats being associated with short pitch corrugation, either as a result or a cause [1], and Squats being more likely at inclines above 1% due to traction and braking behaviour [2]. These correlations account for a majority of the defects, so perhaps the other quarter of the defects investigated have slightly different causes. This work looked at the causes of five different defects that were all initially identified as Squats by the operators that kindly sent them. Only one of them was believed to be a traditional Squat after closer inspection.

The determining of the five defects classifications was initially tackled by gaining a volumetric representation of the defect, through μ-CT scanning. Due to some drawbacks of the μ-CT technique, as discussed above, this had to be coupled with complimentary techniques such as IF microscopes and traditional metallography to ensure what was being seen was trustworthy.

7.1.   Can μ-CT scanning accurately detect crack networks on a large scale?

X-ray penetration of thick steel is a challenge. To map a full rail defect between 7cm and 11cm thick steel needed to be scanned. The thicker the sample is, the lower the resolution will be and the more surface distortion will occur due to more of the photons being attenuated or blocked. This is partly scattering effects that occur across boundaries. It is also because if a photon does not make it all the way from source to detector, it is unknown how far the photon reached, which results in the surface appearing more dense than it actually is. The fewer photons that make it through the length of the object, the worse the image will be, with no image if no photons get through. This is why scanning thick steel is difficult.

From the techniques employed in this work, the combination of CT scanning with the mapping capability of an infinite focus microscope proved very effective. Overcoming the main drawback of CT scanning, other than not being a mobile technology, which was not being able to see the surface features clearly. This was compensated for by the surface maps produced by the infinite focus microscope. The CT scanning allowed a view into the rail and the mapping of where the crack network originated and then how it developed. The maps also allowed a referral back to the surface of the rail after sectioning has begun to check what was above any other features found.

7.2.   Are all of the features captured during scanning real, or do artefacts make the results deceptive?

Section 4.3.1.1 discussed the verification of the initial scans that were made of three samples. The images were of varying resolution due to the variation in size of the samples scanned. In all three scans the main crack networks could be distinguished from the bulk steel, but there were clear artefacts of various types, the most curious of which were the shadows that appeared on crack networks.

These artefacts were found to be caused by real features, namely branching cracks and voids (which may have been filled with debris). The presence of two cracks very close together created a shadow that appeared the same as the shadow seen when a real void was found.

Therefore, care needs to be taken when observing what may look like porosity, especially if there are cracks or some other form of boundary very nearby.

As mentioned above in Section 7.1, the surface appears much denser that it actually is due to the lower energy photons that do not make it through the sample. Photons are created as a packet with varying energies, some are high enough to penetrate whereas others are not. Not only is this deceptive when trying to determine any differences that may be present in the steel, but coupled with the scatter that occurs at the surface, it makes the near surface images harder to trust. This was alleviated through the use of the IF microscope and the surface maps it produced. If the surface is of extreme interest, reducing the sample size as much as possible would help.

In Sample 4, which was larger than the other samples, there were parts of the scan that were not very clear. These less clear parts could be seen by the human eye and highlighted manually, but suffered from too much scatter for any automatic detection methods. Even when using the segmentation technique that usually helped the automatic modules, by reducing the range of voxel values that it had to consider when trying to detect cracks. Sample 4 also suffered from an artefact that was caused by a partial cut that was made into the sample, creating a streak of bright voxels from the tip of the cut into the material. This was not covered earlier in the work, but is worth mentioning to show that a partial cut into a sample can also distort the final image. Luckily in this work, the area affected was not of interest.

7.3. Are there differences between the crack morphologies of defects?

Some defects had separate tiers of cracks growing further down from a branch on the upper crack plane. Sample 1 looks like its lower branch developed first with the upper branch appearing later. In contrast, Sample 2 was particularly bad with many branching cracks. So many that the some of the finer cracks in the models shown in this work had to be omitted for clarity. The main crack developed fairly parallel to the surface, with branches growing below it, sometimes deeper down into the rail. Sample 3 was not scanned, but had a transverse defect like Sample 2, except it grew all the way through the railhead whilst in service. Sample 4 has a simple planar crack structure that did not branch like the others. It initiated at a sphere of contaminant, dipped down into the rail slightly before running parallel to the surface in all

directions. Sample 5, being from the same line as Sample 2, looked like it could have developed into a similar defect to sample two because it also had already developed a crack cluster underneath a severely pitted region.

Why some Squats develop transverse defects whilst others resurface and shell is curious. The safety risk of Squats derives from this unpredictability, in the form of which ones will break a rail versus which ones will not. This was demonstrated by Sample 3, which broke due to a transverse defect whilst a train was passing over it. Luckily the train still passed safely.

The rapid development of the transverse defect in Sample 2, within 11MGTs of service, shows how quickly a transverse defect can seed. The fretted region was found below and in the same orientation as a deep grinding mark on the surface. The deep grinding mark found on the surface of Sample 2, in the same orientation as the fretted region/voids, is very likely to have influenced the transverse defects development. However, it cannot be the only reason otherwise all deep grinding marks would create defects. It should be noted that Sample 2 was taken from a less used section of the network, so although being only 11MGTs of traffic, the age of the rail may be older than expected.

The age of the rail may factor into the determination that the transverse defect developed from a fretted region of the rail, because fretting requires corrosion, which is usually time dependant. Therefore, the presence of the fretted region is probably just as much a result of the age of the rail as the deep grinding mark found to coincide with it. Perhaps if only one was present, the damage would be less significant. As mentioned above, an earlier stage version of the fretted structure in Sample 2 was also found in Sample 5, which was from the same section of track. However, corrosion and fretting was not found, though the cluster of cracks found in Sample 5 were at the farthest point of the crack structure, so fretting may have occurred given more time. The temperature history of the rail also plays a role in the growth of transverse defects, as cold rails create tensile stresses as they contract the rail.

Although information about track parameters and traffic history for the rail is very limited, the following general observations were made regarding the defect crack morphology as grouped by traffic type. Metro traffic produced more complex crack structures in the form of more branches. The high-speed sample had the simplest but longest morphology having no branches, but this was on a weld and passed through a varying microstructure. The mixed traffic sample

branched and developed a transverse defect, but the microstructure was compromised by many long inclusions, affecting its formation. Sample 2 and 5 had signs of at least 2 initiation sites, producing cracks that linked up later in development. Sample 2 and 3 had multiple transverse cracks that also interconnected during development.

The presence of Y-shaped surface cracks on the field side of Sample 1 and 2 matched an early stage Squat from the literature (Section 4.5.1, Figure 83). A crack developing early on the field side of the rail can be explained by hollow wheel profiles. Hollow wheel profiles were confirmed as being present for sample 2, highlighting the need for good wheel-rail profile conformity to distribute the load evenly.

7.4.   Is there a correlation between the microstructure and the defect?

Sample preparation was still needed in conjunction with the scanning to check the microstructure and issues such as inclusions and WEL. These checks were focussed around the initiation site once it had been identified. Being a Stud, Sample 1 had clear bright WEL streaks on the surface and some localised grain refinement. Signs of thermal damage such as excessive WEL, grain refinement and tempering of the surface would be suggestive of a Stud rather than a classic Squat. Studs from the literature showed tempering colours on the surface, but were not seen on Sample 1.

The sample microstructures were different because of the different steel grades. Microstructure had an influence in the softer grades, providing a preferable crack route through the ferrite on the grain boundaries. The microstructure of Sample 4 varied longitudinally due to being across a weld. The defect initiated at a flaw in the microstructure on the edge of the weld, but after the initiation site the crack plane maintained its direction parallel to the surface regardless of the change in microstructure. The microstructure varied with depth in Sample 1 due to the thermal damage, which added to its categorisation as a Stud. Sample 3 had many inclusions in the microstructure and their detrimental effects are already known. Other than Sample 3 and the initiation site in the weld of Sample 4, the overall microstructures of each defect was normal.

7.5. What influence does surface quality have on the defect growth behaviour?

The severity of Sample 2, in terms of being two interconnected defects with a transverse defect that reached 9mm into the rail after only 11MGTs of traffic, was linked to the quality of the surface. A specific deep grinding mark was connected to the origin of the transverse defect and the extensive pitting of the surface correlated with the crack growth directions. Sample 5 was less developed, but from the same location as Sample 2 and also had a subsurface crack that suddenly branched multiple times (section 6.3.2.3). Directly above this branching location was a long pit that reduced the contact area, increasing the contact stress per unit area. It is believed that this pit contributed to the sudden branching of the crack. The branched crack looks like a precursor to the fretted region in Sample 2 that led to the transverse defect developing. This connection between surface roughness and subsurface damage is already known from the literature, which describe rough surfaces increasing subsurface stress by approximately 8 times.

The affect that a rough surface can have on the material was also noted in the transition from a smooth area to a rough area (uniform nominal grinding marks) on Sample 4. The cracks in the smooth area where typical RCF cracks that ran at an acute angle to the running surface whereas the cracks in the rough region were vertical/perpendicular to the running surface (section 5.5). The exact reason as to why the weld contained a transition from a rough to a smooth section was not determined, so this is not factored into the connection. This is more of an observation than a specific cause and effect.

Grinding marks may be indicative of a GIS, though care should be taken to assess if the grinding marks are considered excessively deep, as clearly not all grinding marks cause a defect, and if the grinding mark has a clear influence on the initiation site. In this work, the CT scans combined with the surface maps were used to determine that the deep grinding marks were in the same location as the defect origin.

Samples 2 and 5 were both GIS, with Sample 2 looking like a double defect in the form of what appears to be an interconnected surface defect and subsurface defect. The surface defect may have originally been a subsurface defect, with the surface worn away from grinding and natural

wear. The deeper of the two defects may have initiated at the same depth as the first, but later. The close proximity of the two defects could be a factor in the appearance of a transverse defect, as water seems to have been transferred down the cracks from the breached surface, down into the origin of the transverse defect, allowing fretting. Perhaps double/interconnected defects should be considered more dangerous due to a larger portion of the rail being compromised. To be sure how this defect developed, track observations were needed to map its development. It is so complex that it is difficult to be sure what cracks grew first.

7.6.   Classifications of Squats

As Squats research has developed, both the details of the defects themselves and the awareness of variations in the defects have been appreciated, in the once singularly named Squat defect (aka tache noir, Shinkansen shelling, black/dark spot). This has led to new categories within the Squat defect: the classic Squat, the Stud, the Grinding Induced Squat (GIS) and the Belgrospi (not covered in this work). There is even possible scope for a category that covers Squats related to welds, such as Sample 4, and perhaps Squats that are related to unexpected microstructural flaws, such as Sample 3, should also have a category. Although these extra categories may seem to make the subject more complex and less clear, their introduction into big data analysis on where Squats appear, could provide better results by removing anomalous results.

The classic Squat takes ~100MGT of traffic to develop through ratchetting, like other forms of RCF. This relatively slow and incremental development makes this defect less concerning compared to some of the defects Studied in this work. Sample 1 was a Stud with unknown MGTs, but the literature mentions Studs develop at ~30MGTs, making them more time sensitive than classic Squats. Studs could be considered more dangerous because they form faster than classic Squats, but they have not been observed to develop transverse defects without the influence of other defects.

For the five defects investigated in this work, three different causes were identified: thermal damage, microstructural flaws (MnS or weld contamination) and deep grinding marks. The causes are tabulated in Table 13 of Chapter 6. These three causes, all different to the expected classical Squat cause of RCF, segment the five defects into different categories. The following are the distinctions made between defects in this work based on the information available,

therefore it may be expanded upon if track inspections are included. The thermal damage, along with other factors such as a very low MGT, indicates the defect is a Stud, whereas the grinding marks on two defects make them highly likely to be GIS, a newer category of defect. Sample 2 was initially believed to be a Stud, primarily due to the lack of plastic deformation and low MGT, as tabulated in Chapter 4. However, using the new information shown in Chapter 6, following the trace of the subsurface voiding back to its cause, it was realised that Sample 2 was actually a GIS. The low MGT makes the GIS a rapidly developing defect like the Stud, but with a stress intensifying surface flaw rather than heat flux as an initiating factor.

From this work, the identification of two GIS defects is a key finding. Especially because they were identified using NDT by tracing the crack growth back to its origin, which is new and was not how GIS have been identified in the past. Identifying GIS usually needs careful inspection of the WEL, knowledge of the grinding history of the rail and inspection of a reasonable length of the rail. For a defect where none of this is possible, as with all of these samples, µ-CT scanning is very effective for tracing a 3D crack structure back to its origins. The effectiveness of this method proved itself further by identifying more than one initiation site for both GIS defects. This can be very difficult to do if a sample is sectioned into 2D slices due to the need to reconstruct the slices back together, needing interpolation between slices that may not be accurate if the crack path is erratic.

The defects that developed from microstructural flaws probably should not be grouped with classic Squats, even though they are called Squats in table 5 of Chapter 6, due to a lack of a more suitable term. Classic Squats develop as RCF from ratchetting of the surface and cracks developing in the exhausted surface material. Classic Squats do not need a microstructural flaw to develop as the ratchetting provides the opportunity for initiation. Therefore grouping defects that occur due to microstructural deviations, such as inclusions or weld contaminations, could make the data misleading. Squats due to inclusions should become a legacy issue as continuous casting and secondary steelmaking has eliminated the abundance of inclusions seen in Sample 3. Inclusion Squats should be identified as such so that classic Squats, caused by RCF, poor track parameters, indentations etc., are also clear in their identity. Squats on welds could be caused by the differential wear of the varying hardness within the rail, or by a contamination as found in this work, so they could be separated out too. However, the author is aware of the opinion that weld Squats lead to further Squats down the rail through excitation of the rolling stock suspension, so they may need to be grouped with in-service Squats for this reason.

Grouping defects by their cause would allow the elimination of any manufacturing or installation causes, leaving the in-service flaws for root cause analysis.

The reality in trying to separate out different categories of the Squat defect is that, without a clear surface indicator such as a grind mark, tempering colours or plastic deformation, categorising these defects on a track inspection is difficult. For the inclusions, time will eliminate this issue as legacy rail reaches the end of its life and is removed. For issues with welds, this adds gravitas to the idea of separating them from other Squats in case weld related issues caused the Squat. Classic Squats, Studs and GIS should be distinguishable from track inspection, although grind marks may have worn away between a GIS initiation and its identification. Inspection trains have a larger scope for detecting flaws as they utilise non-destructive techniques such as eddy current and ultrasound. Through developments in these areas and new areas such as magnetisation, it may be possible to detect differences in Squat types. Effectively replacing the role CT scanning played in this work with a more train-mountable option.

7.7.   Unique features within defects

Sample 1: Cylindrical crack feature beneath defect ridge, probably beginning as a branched crack before another small crack completed the 'circle' by growing vertically. Effectively breaking the tip off the material between the higher and lower longitudinal cracks. This seemed to be a source of more cracks forming due to increased movement of the broken off tip. Shared some similarity in morphology to the defect in Sample 2, but without a transverse defect.

Sample 1 and 2: Y-shaped crack. Seen in a very early stage defect in the literature. Could be a precursor to a Squat defect and may indicate hollow wheels. Hollow wheels were confirmed for Sample 2. It should be noted that Samples 1 and 2 were different defects, so Y-shaped crack may not guarantee a type of defect if any defect at all.

Sample 2: Multiple initiations, one from a grinding mark on the gauge side, one from hollow wheels and the field side, both connected closer to the gauge side to form a much larger defect.

Sample 2: At least one transverse defect stemming from a fretted crack cluster, caused by a deep grinding mark directly above it. There was a second crack with a shallower growth path.

Sample 3: Long MnS inclusions. A known cause of steel failure. A legacy issue that should show up less where steel is manufactured using continuous casting rather than ingot casting.

Sample 4: Bubbles/spheres just under surface in weld. Not necessarily new but a strong case for cleanliness during welding.

7.8.　　New findings

1. Full Squat defects can be scanned using a 450kV polychromatic beam source. The results are reasonably representative as long as artefacts are understood.

2. The source of a transverse defect was found to be a fretted crack cluster that became a void in sample 2. Caused by a deep grinding mark directly above and in alignment with the void.

3. The identification of two GIS defects (Samples 2 and 5) using an NDT technique..

4. A Squat (possibly Stud) that was caused by weld contamination in sample 4.

5. Multiple initiation sites in one defect, Sample 2 and 5. One was a grinding mark in Sample 2, whilst the other was probably hollow wheels (confirmed as being present). Both were grinding marks in Sample 5.

6. Y-shaped cracks on the field corner seem to be an early indicator of Squat type defects caused by hollow wheel profiles. They might develop into the full V-shaped cracks common with Squats, as seen in the literature.

7.9. Limitations of work

A strong limiting factor within this work was the unknown origin of most of the samples. Samples are difficult to obtain from track and every defect received was greatly appreciated, because work like this would be impossible without such generosity. Some information was provided with the samples, such as estimated MGTs and rolling stock type However, it was generally limited and could not necessarily be compared to another sample. Not knowing the origin of the samples meant that the traffic conditions were unknown, such as actual axle loads, curvature of track, inclines, locations of lubricators etc. Not having the manufacturing details from the side of the rails also limited how much information could be gathered about the steel.

Not knowing the origin of the sample leads to another limitation, not being able to observe the defect on site. This would have allowed the opportunity to gather some of the details mentioned above as well as observe factors such as periodicity of defects, track deflection, crushed ballast, track alignments and local environment (tunnel, trees, bridges, level crossings etc). It would also allow long term studies to document the evolution of the defect and the wear behaviour of the railhead.

As with most work, there was a cost limitation, especially regarding the μ-CT scans. Ideally, many more scans would have been performed from the sample being as large as possible, slowly trimmed down through multiple stages, to capture the finer details for all of the samples. To be effective within the budget, Sample 2 was selected as being the most interesting sample to progress with and only one more trim and scan was performed. A large scale scan allows the overall morphology to be captured. Then areas of interest, such as the origin of a transverse defect or initiation site to be identified for higher resolution scans.

The authors limited skill with μ-CT reconstruction software was also a factor, not having any experience of using voxel based reconstruction software before this study. The defect volumes could not be reconstructed using the usual methods/modules that could be used in palaeontology, for example. This was due to much more interference because metal was being scanned rather than rock. The author developed a method of constructing large models using a segmentation technique, as described in this work. More experienced users may be able to refine the segmentation technique further and deploy techniques that are more advanced, allowing more information to be gained from the reconstructed models.

7.10. References

[1]   Z. Li, X. Zhao, C. Esveld, R. Dollevoet, and M. Molodova, "An investigation into the causes of Squats-Correlation analysis and numerical modeling," *Wear*, vol. 265, no. 9–10, pp. 1349–1355, 2008.

[2]   Z. Li, "Squats on Railway Rails," in *Wheel-Rail Interface Handbook*, R. Lewis and U. Olofsson, Eds. Woodhead Publishing Limited, 2009.

Chapter 8: Conclusions

8.1.   Can μ-CT scanning accurately detect crack networks on a large scale?

Yes, if the source strength is high enough to penetrate the thickness of the specimen. The source used in this work was 450kV. There is a limit to how fine the cracks that the scan will detect. This limit is determined by the size of the scanned sample, the size of the detector being used and the thickness of the crack being detected. The main crack structure was detectable in each of the samples, with only the finest cracks being omitted from the scan. This was improved upon in sample 2 by trimming the sample down to the unclear region and scanning again to pick out the very complex network of cracks that were underneath the main crack plane, some of which were growing down into the rail.

8.2.   Are all of the features captured during scanning real, or do artefacts make the results deceptive?

Artefacts do exist within the reconstructed volume but most of them are obvious enough to be ignored. The shadows that were seen along the crack path proved to be caused by real features such as branching cracks and voided regions. The surface of the sample was not particularly clear due to size and density of the sample. This can be complemented by mapping the surface with photographs and surface mapping technology such as an infinite focus microscope.

8.3.   Are there differences between the crack morphologies of defects?

The crack morphologies of the samples all varied, whether that be through the depth that they reached, how often they branched or how far along the rail they stretched. The number of clear branches between defects, their overall size and the depth they reached all varied. This may have been due to traffic and track factors, but these details are unknown for most of the samples, so conclusions cannot be drawn.

Sample 2 became a strong focus due to the presence of a transverse defect with a steep growth angle to the surface. Sample 3 also contained a transverse defect , but because it had already failed in track, most useful information had already been lost. Identifying a transverse defect in track using ultrasound is difficult due to crack shielding; a problem not present in CT scanning because the signal is not reflected back by a crack and the sample is rotated during the scan.

8.4.   Is there a correlation between the microstructure and the defect?

Sample 1 showed various microstructural signatures of high thermal variations, which is consistent with the theory of Stud initiations. It contained WEL consistent with previous Stud investigations, vertical cracks as seen in rail gun metal damage (high surface temperatures) and some localised grain refinement. Sensitivity for microstructure was seen closer to the surface, probably earlier in the defects development, but as expected, this sensitivity dissipated as the crack grew longer. Flaws in the microstructure where found in samples 3 and 4, one being a legacy issue of large MnS inclusion with the other being contamination during a weld. These flaws then developed into small crack structures and eventually a full defect. MnS will be less of an issue in the future but weld cleanliness is still very important to prevent contaminations that can lead to defects developing.

8.5.   What influence does surface quality have on the defect growth behaviour?

The transverse defect found within sample 2 became a strong focus, with its origin being a fretted void from multiple crack branches, located in the same orientation and size as a deep grinding mark above it. The likelihood of these two distinct features being in alignment with each other by pure coincidence seems very low. This demonstrates the effects that a rough surface feature can have on the subsurface. This was calculated in the literature as increasing stress by eight times compared to a smooth surface. The same situation was also found in sample 5, where both initiation sites were traced back to grinding marks.

Pits were noticed in abundance on both sample 2 and 5, but any role they may have played in the development of the defect is not clear. There is some evidence that the pits may have

influenced the morphology of the overall defect, but this may be coincidental. Pits would reduce the contact patch between the wheel and rail, increasing the pressure on the contacted surface, so it is reasonable to consider them as undesirable. Between the grinding marks and the pits, there is a strong case for trying to keep the rail surface as smooth as possible.

8.6.   Squat defect classification

Some Squat defects were reclassified to Studs in the last 20 years, after more than 60 years of being aware of Squats by some name. A much more recent group was identified much more recently, Grinding Induced Squats (GIS), although they were not called this initially. In order to utilise big data analysis effectively, these different defects need to be identified correctly because they have different causes. Even more groups may also need to be established for weld related defects and defects due to inclusions rather than in service parameters. Just within the five samples studied in this work, only one Squat retained its original classification. The others were due to a microstructural flaw from manufacturing or installation or a grinding flaw from maintenance. This is compared to traditional Squats that are from fatigue during service.

The identification of the two GIS defects is a key finding. Especially because they were identified using NDT by tracing the crack growth back to its origin, which is new and was not how GIS have been identified in the past. Identifying GIS usually needs careful inspection of the WEL, knowledge of the grinding history of the rail and inspection of a reasonable length of the rail. For a defect where none of this is possible, as with all of these samples, µ-CT scanning is very effective for tracing a 3D crack structure back to its origins. The effectiveness of this method proved itself further by identifying more than one initiation site for both GIS defects. This can be very difficult to do if a sample is sectioned into 2D slices due to the need to reconstruct the slices back together, needing interpolation between slices that may not be accurate if the crack path is erratic.

Knowing more about the origin of the samples would alleviate the biggest limitation of this work. Squat research is normally conducted in conjunction with track visits to evaluate the defects environment fully.

8.7.   Publications from this work

Title: Comparison of Squats and Studs from different traffic environments

Authors: Earl, S., Rankin, K.E., Lewis, R., Smith, L., Rainforth, W.M.

Source: Proceedings of the 11th International Conference on Contact Mechanics and Wear of Rail/wheel Systems, CM 2018

Title: Verification of the use of Micro-CT scanning to assess the features of entire Squat type defects

Authors: Earl, S., Rankin, K.E., Lewis, R., Smith, L., Rainforth, W.M.

Source: WEAR journal, Volumes 438–439, 15 November 2019, 203074

https://doi.org/10.1016/j.wear.2019.203074

8.7   References

[1]   J. M. Cookson and P. Mutton, "The role of the environment in the rolling contact fatigue cracking of rails," *Wear*, vol. 271, no. 1–2, pp. 113–119, 2011.

Chapter 9 : Future work

9.1 X-ray scanning related

Using a μ-CT scanner with a monochromatic beam can allow the detection of microstructure. This was not the focus of this work, but considering the variation in microstructures observed in some areas, especially in a weld, this type of setup could prove very interesting when looking at a defect. If WEL or welds are of interest within the defect, it would definitely be a good option.

A catalogue of 3D scan data, detailing Squat defects should be accumulated and made available. This would allow a reasonable comparison of defects to previous investigations and the re-evaluation of defects in the future, if new information arises.

9.2 Defect related

The link between deep grinding marks and defect occurrence warrants further investigation into this connection (GIS). The differences in finish from a high-speed grinder needs comparing to a traditional grinder, given that one has driven stones whereas the other does not. It may be that there is a connection between the increase in defects and the introduction of high speed grinding trains. The presence of grinding marks brings another possible scenario: corrosion fatigue. In materials science, the introduction of dislocations from grinding and the naturally wet (corrosive) environment of railways could cause corrosion fatigue. Given that the grooves from the grinding are still present in sample 2, there have been locations of little wheel contact where corrosion could develop, perhaps as crevice corrosion. Cracks initiate more easily under corrosion conditions. This could add to the understanding that wet and dry cycles in twin disc testing can produce Squats and RCF. A good start for this branch of work is a paper that lays out the fundamentals of the idea very well [1].

Long term track observations of freshly ground sites that are Squat defect hotspots, to see how many grinding facets develop into defects and see if there is any periodicity to them.

Investigations into the effect of traffic type on defects from the same rail grade would lead to more comparative results. It would be ideal to include Squats from freight traffic in the comparison. In contrast, using the same traffic type on multiple rail grades could also be illuminating, though this has been shown in the literature. Perhaps a test track with multiple grades laid out in series with exactly the same traffic would be a more controlled way of approaching the task.

Bainitic rails should be investigated to see how they perform against Squat development. Bainitic rails have been used in the channel tunnel with positive results, regarding defect mitigation and wear properties. The author is aware of one Squat found on a bainitic rail, but it was on a switchblade, where the contact is much smaller.

Big data analysis should be investigated to look for trends in defect appearance, especially regarding track layout, rail age and maintenance regime. This will be subject to the defects being classified correctly, so evaluation of anomalous results will probably be necessary to compensate.

Knowing more about the residual stresses present in new, defect free and defective rail may also give clues as to why the morphology of Squats vary so much. Residual stress could explain why hotspots occur and why some cracks grow into defects whereas others do not. A consideration of this approach is that the act of cutting a rail can release residual stress. It may need to be done in situ or on as long a length of rail as possible.