I. INTRODUCTION

Tribology, the science and technology of friction, wear, and lubrication, is an interdisciplinary subject. It can therefore be addressed from several different viewpoints. This chapter focuses on the friction, wear, and lubrication of the tiny contact zone (roughly 1 cm$^2$), where steel wheel meets steel rail, from a mechanical engineer’s viewpoint. In contrast to other well-investigated machinery, such as roller bearings, the wheel–rail contact is an open system. It is exposed to dirt and particles and natural lubrication, such as high humidity, rain, and leaves, all of which can seriously affect the contact conditions and the forces transmitted through the contact. In contrast, in roller bearing the ball-cage contacts are sealed away. The steel rail meets a population of steel wheels from a number of different vehicles and the form of both the wheels and the rail can change due to wear. In contrast, a roller bearing meets the same rollers without any form change of the contacting bodies.

A comprehensive overview of the science of tribology is presented in the ASM handbook, while a closer examination of the material science field is given by Hutchings. The mathematical modelling aspects of tribology, i.e., contact mechanics and fluid film lubrication, are presented by
In the contact zone between wheel and rail, normal and tangential loads are transmitted. How the steel wheel meets the steel rail and the size of the forces transmitted in the contact zone influence damage mechanisms, such as wear and surface cracking, are discussed. The contact conditions of the wheel–rail contact are discussed in Section II (Contact conditions at the wheel–rail contact).

When two surfaces under load move relative to each other, wear will occur. Wear is often defined as damage to one or both surfaces, involving loss of material. Wear and other surface damage mechanisms are discussed in Section III (Wear and other surface damage mechanisms).

The friction force can be defined as the resistance encountered by one body moving over another body. This definition covers both sliding and rolling bodies. Note that even pure rolling nearly always involves some sliding and that the two classes of motion are not mutually exclusive. Any substance between the contacting surfaces may affect the friction force. The contact conditions may cause the substance to be wiped away quickly and its effect will be minimal. On the other hand, surface films formed between interposed substances have a major effect on the frictional behaviour. The friction of the wheel–rail contact is discussed in Section IV (Friction), as well as causes of friction loss and methods for increasing the friction.

Lubricant application to the wheel–rail contact as well as surface coatings are used to reduce friction and damage due to wear etc. This is discussed in Section V (Lubrication and surface coatings).

What one always should bear in mind when studying and using tribological data is that friction and wear are system parameters and not material parameters like modulus of elasticity or fracture toughness. This means that frictional and wear data taken from one system, such as a roller bearing, cannot be directly applied to another system such as the wheel–rail contact. This also highlights the need for a special study of the tribology of the wheel–rail contact.

II. CONTACT CONDITIONS AT THE WHEEL–RAIL CONTACT

In the contact zone between railway wheel and rail the surfaces and bulk material must be strong enough to resist the normal (vertical) forces introduced by heavy loads and the dynamic response induced by track and wheel irregularities. The tangential forces in the contact zone must be low enough to allow moving heavy loads with little resistance, at the same time the tangential loads must be high enough to provide traction, braking, and steering of the trains.

The contact zone (roughly 1 cm$^2$) between a railway wheel and rail is small compared with their overall dimensions and its shape depends not only on the rail and wheel geometry but also on how the wheel meets the rail influence, i.e., lateral position and angle of wheel relative to the rail, as shown by Le The Hung$^6$.

It is difficult to make direct measurements of the contact area between the wheel and the rail. An interesting approach for measuring the contact area for full-scale worn wheel and rail pieces is presented by Marshall et al.$^7$ They used an ultrasonic reflection technique and the results were compared with calculated contact areas showing good agreement, as shown in Figure 5.1. The surface topographies of the ultrasonic measured surfaces were measured with a stylus instrument and used as input to a contact mechanics method for rough surfaces (for details see Björklund et al.$^8$). Poole$^9$ used low-pressure air passing through 1 mm diameter holes drilled into the rail head to measure the contact area as the holes being blocked by the passing wheel. Measurement of these pressure variations allows studying of the contact area shape under dynamic conditions.

The size and shape of the contact zone where the railway wheel meets the rail can be calculated with different techniques. Traditionally, the Hertz theory of elliptical contacts$^3$ has been used implying the following assumptions: the contact surfaces are smooth and can be described by second degree surfaces; the material model is linear elastic and there is no friction.
between the contacting surfaces; and the contacting bodies are assumed to deform as infinite half spaces. The half space assumption puts geometrical limitations on the contact, i.e., the significant dimensions of the contact area must be small compared with the relative radii of the curvature of each body. Especially in the gauge corner of the rail profile, the half plane assumption is questionable since the contact radius here can be as small as 10 mm. Due to its simple closed form solutions, the Hertz method is the most commonly used approach in vehicle dynamics simulation. However, other methods are used for simulation of wear and surface fatigue due to the overestimation of the contact stresses attributed to the nonvalidity of the half plane assumption and nonlinear material behaviour. Kalker’s numerical program Contact still depends on the half space assumption, but is not restricted to elliptical contact zones. The contact surfaces are meshed into rectangular elements with constant normal and tangential stresses in each rectangular element. Telliskivi and Olofsson developed a finite element model, including plastic deformation, of the wheel–rail contact using measured wheel and rail profiles as input data. They compared the traditional methods (Hertz and Contact) with their detailed finite element solutions of the wheel in contact with the rail gauge (Case 1 in Figure 5.2) and the wheel in contact with the rail head (Case 2 in Figure 5.2). The results in terms of contact zone shape and size, as well as stress distribution, are presented in Figure 5.3. The results from two test cases show that the difference in maximum contact pressure between Contact/Hertz and the model was small for test case 2 when the minimum contact radius is large compared with the significant dimensions of the contact area (half space assumptions valid). However, in test case 1 where the minimum contact radius was small compared with the significant dimensions of the contact area, the difference between the model and Contact/Hertz was as large as 3 GPa. Here, the difference was probably due to both the half space assumption and the material model.

The Stockholm local network has been the subject of a national Swedish transport programme (the Stockholm test case) in which the wear, surface cracks, plastic deformation, and friction

**FIGURE 5.1** Contact pressure maps for a load of 80 kN: (a) ultrasonic measurement; (b) Hertzian; (c) elastic model; (d) elastic-plastic mode (from Marshall et al.7).
of rail and wheel have been observed for a period of 2 years. The data from the Stockholm test case has been used for validation of different wear models, see\textsuperscript{16–18} and also surface crack models.\textsuperscript{19}

Furthermore, the trains used in this study have been modelled with train dynamic simulation software such as GENSYS\textsuperscript{17} and Medyna.\textsuperscript{20} A parametric study\textsuperscript{17} was performed on curves with different radii representative of Stockholm local traffic. The results are presented here in the form of a contact pressure sliding velocity diagram (Figure 5.4). A clear difference could be found between the rail head-wheel tread contact and the rail gauge-wheel flange contacts in terms of sliding velocity and contact pressure. For the rail head-wheel tread contact, the sliding velocity and the contact pressure was never above 0.1 m/sec and 1.5 GPa, respectively, but for the rail gauge-wheel flange the maximum sliding velocities reached 0.9 m/sec, and maximum contact pressure was observed up to 2.7 GPa. Also shown in Figure 5.4 are simulation results from a curve with a 303 m radius, for the Stockholm test case using the software Medyna. This is a sharp curve with one of the smallest radii in the network and one can note a very high contact pressure for the first wheel on the

![Contact mechanics analysis methods comparison](image)

**FIGURE 5.3** Comparison, with respect to maximum contact pressure and the contact area, between three different contact mechanics analysis methods.\textsuperscript{11}
leading bogie in contact with the rail gauge. Other examples of modern railway operation which led to high contact stresses that are significantly over the yield strength of the material are presented in Kumar and Cassidy.

III. WEAR AND OTHER SURFACE DAMAGE MECHANISMS

The profile change of rails on curves makes a large contribution to track maintenance cost. The profile change on wheels can also be significant, especially on a curved track. Damage mechanisms such as wear and plastic deformation are the main contributors to profile change. Another growing problem for many railways is rolling contact fatigue. In Europe, there are more than one hundred broken rails each year due to rolling contact fatigue. In 1995, rail maintenance costs within the European Union were estimated to total 300 million Euro annually.

A. WEAR

Wear is the loss or displacement of material from a contacting surface. Material loss may be in the form of debris. Material displacement may occur by transfer of material from one surface to another by adhesion or by local plastic deformation. There are many different wear mechanisms that can occur between contacting bodies, each of them producing different wear rates. The simplest classification of the different types of wear that produce different wear rates is “mild wear” and “severe wear”. Mild wear results in a smooth surface that often is smoother than the original surface. On the other hand, severe wear results in a rough surface that often is rougher than the original surface. Mild wear is a form of wear characterised by the removal of materials in very small fragments. Mild wear is favourable in many cases for the wear life of the contact as it causes a smooth run-in of the contacting surfaces. However, in some cases it has been observed that it worsens the contact condition and the mild wear can change the form of the contacting surfaces in an unfavourable way. Another wear process that results in a smooth surface is the oxidative wear process characterised by the removal of the oxide layer on the contacting surfaces. In this case the contact temperature and asperity level influence the wear rate. Abrasive wear caused by hard particles between the contacting surfaces can also cause significant wear and reduce the life of the contacting bodies.
In wheel–rail contact, both rolling and sliding occur in the contacting zone. Especially in curves, there can be a large sliding component on the contact patch at the track side of the rail head (gauge corner). Due to this sliding, wear occurs in the contact under the poorly lubricated condition that is typical of wheel–rail contact, as shown in Figure 5.5. An observation that can be made on sliding wear is that an increase of the severity of loading (normal load, sliding velocity, or bulk temperature) leads, at some stage, to a sudden change in the wear rate (volume loss per sliding distance). The severe wear form is often associated with seizure. The transfer from mild acceptable wear to severe/catastrophic wear depends strongly on the surface topography. The loading capability of a sliding contact may be increased considerably by smoothing the surface. \(^{28}\) Chemically reacted boundary layers imposed by additives in the lubricant can improve the properties of lubricated contacting surfaces and reduce the risk of seizure. \(^{29}\) Also, as shown by Lewis and Dwyer-Joyce, \(^{30}\) the surface temperature influences the transition from mild to severe wear.

In addition to the contact pressure and the size of the sliding component, natural and applied lubrication strongly influenced the wear rate \(^{13–15}\) for the full-scale test results from the Stockholm test case. Both lubricated and nonlubricated, as well as seasonal variations, were studied. In addition, two different rail hardneses were studied in the same test curves. Track side lubrication reduced the wear significantly, and a lubrication benefit factor 9 for small radius curves (300 m) was reported. For 600–800 m radius curves the lubrication benefit factor was approximately 4. The variation seen in wear rates over the year was probably due to natural lubrication caused by changing weather conditions. An analysis of the relationship between weather conditions and measured rail wear shows that the precipitation has a significant effect on rail wear as shown in Figure 5.6. Waara \(^{31}\) reports that gauge face wear in a northern Sweden heavy haul application can be reduced 3–6 times with correct full year lubrication. Engel \(^{32}\) also reports significant reduction of wear by lubrication, here the lubricant benefit factor was 4 in a twin-disc test. An on-board
The on-board lubrication system was evaluated by Cantara\textsuperscript{33} in a Spanish study. The results were that the flange wear was reduced by a factor of 4.5 for wheels equipped with the on-board lubrication device.

The curve radius of the track has a strong influence on rail wear. The influence also strongly depends on the vehicles and their behaviour. In the Stockholm test case all vehicles were of the same type and passed over all the test sites with the same frequency. In this case the influence of curve radius can be clearly seen when comparing rail wear rate as a function of curve radius. The rail wear rate seems to increase exponentially for decreasing curve radius, as shown in Figure 5.7.

For a given situation a higher steel grade usually reduces rail wear. This effect is shown in Figure 5.8 for two different high rails with steel grade UIC 900A and UIC 1100, respectively, within the same lubricated, as well as a parallel nonlubricated, 300 m radius curve. For the nonlubricated curve the ratio between rail wear rate for the 900A grade rail compared to that of the 1100 grade rail is approximately 2. This can be compared with the lubricant benefit factor that was approximately 9 in this curve, as can be seen in Figure 5.8a–d, when comparing the nonlubricated and lubricated cases. The difference between rail head wear (low sliding velocities and contact pressure) and rail gauge wear (high contact pressure and sliding velocities) was seen to be a factor 10. This is also comparably higher than the rail grade benefit for modern rail steels as UIC 900A and UIC 1100. The observation that the contact conditions in terms of contact pressure and sliding velocity are more important than the grade of steel (900A and 1100) has also been verified in two-roller tests.\textsuperscript{13} However, when Lewis and Olofsson\textsuperscript{34} compared rail steel wear coefficients taken from laboratory tests run on twin disc and pin-on-disc machines, as well as those derived from measurements taken in the field, they found that the introduction of more modern rail materials had reduced wear rates by up to an order of magnitude in the last 20 years.

Fully pearlitic rail steels are still the most common and are used by most railways. Pearlite is a lamellar product of eutectoid composition that is formed in steel during transformation under isothermal continuous cooling. It consists of ferrite and cementite. Perez-Uzeta and Beynon\textsuperscript{35} have shown that the wear rate of pearlitic rail steel decreases with lower interlamellar spacing between the cementite lamella giving a corresponding increase in hardness. Steels with a bainitic microstructure are the other main rail steels. They have shown better rolling contact fatigue resistance than pearlitic rail steels. However, the wear resistance of bainitic rail steels is inferior to that of pearlitic rail steels at a fixed tensile strength, as shown by Graham and Beynon\textsuperscript{36} and Mitao et al.\textsuperscript{37}

**B. Plastic Deformation**

On a straight track, the wheel is in contact with the top of the rail, but in curves, the wheel flange may be in contact with the gauge corner of the rail. The wheel load is transmitted to the rail through

---

**FIGURE 5.7** Wear rate for high rail as function of curve radius in the Stockholm test case (from Nilsson\textsuperscript{15}).

MGT = mega gross tonne traffic.
a tiny contact area under high contact stresses. This results in repeated loading above the elastic limit, which leads to plastic deformation. The depth of plastic flow depends on the hardness of the rail and the severity of the curves; it can be as much as 15 mm.\textsuperscript{38,39} When a material is subjected to repeat loading, its response depends on the ratio of the amplitude of the maximum stress to the yield stress of the material. When the load increases above the elastic limit, the contact stresses exceed yield and the material flow plastically. After the wheel has passed, residual stresses will develop. These residual stresses are protective in nature in that they reduce the tendency of plastic flow in the subsequent passes of the wheel. This, together with any effect of strain hardening, makes it possible
for the rail material to support stresses that are much higher than its elastic limit. This process is called elastic shakedown and the contact pressure limit below which this process is possible is known as the elastic shakedown limit. There is also a plastic shakedown limit. Loads between the elastic and plastic shakedown limit will lead to cyclic plasticity of the rail. If repeated, cyclic plastic deformation takes place and the rail material can cyclically harden, which leads to an increase in the yield stress and reduces the tendency of plastic flow. For loads above the plastic shakedown limit, plastic ratchetting will occur, i.e., small increments of plastic deformation accumulate with each pass of the wheel. Plastic ratchetting can be found in a curved track as a lip down of the rail gauge corner, as shown in Figure 5.9. Plastic ratchetting is the main cause of headcheck surface cracks.

The consequences of ratchetting are wear and the initiating of fatigue cracks as the material accumulates strain up to its limiting ductility. Beyond this limit, failed materials can separate from the surface as wear debris or forms crack like flaws, as shown in Figure 5.10.

C. ROLLING CONTACT FATIGUE

Rolling contact fatigue cracks on the rail can be classified into those that are subsurface-initiated and surface-initiated. Subsurface-initiated cracks are often caused by metallurgical defects. On the other hand, surface initiated cracks seem to be the result of traffic intensity and axle load. A more specific division can be made into shelling, head checks, tache ovale, and squats. Shelling
(see Grassie and Kalousec\textsuperscript{42}) is a subsurface defect that occurs at the gauge corner of the high rail in curves on railways with a high axle load. An elliptical shell-like crack propagates predominantly parallel to the surface. In many cases the shell causes metal to spall from the gauge corner. However, when the crack length reaches a critical value, the crack may turn down into the rail, giving rise to fracture of the rail. Head checks (Boulanger et al.\textsuperscript{43}) generally occur as a surface initiated crack on or near the gauge corner in curves, as shown in Figure 5.11. Head checks may branch up towards the surface of the rail, giving rise to spalls. However, for reasons still not clearly understood, cracks can turn down into the rail and, if not detected, cause the rail to break. These events are rare, but are dangerous since surface cracks tend to form continuously.\textsuperscript{14} Frederick\textsuperscript{44} discusses the effect of train speed and wheel–rail forces as a result of surface roughness. Furthermore, he discusses whether hard rail or soft rails should be used in curves and also the relationship between wear rate and surface crack propagation. The conclusion was that hard rails are more prone to surface cracking. This was also seen in the Stockholm test case,\textsuperscript{14} where UIC 900A rail material was compared against UIC 1100 rail material. Both materials seemed to be similarly sensitive to crack initiation, but the 1100 grade rail was more sensitive to crack propagation and also more sensitive to the formation of headcheck cracks. More information on the initiation mechanisms and growth of rolling contact fatigue cracks can be found in Beynon et al.\textsuperscript{45} Tache ovale, or shatter cracks from hydrogen,\textsuperscript{42} are defects that develop approximately 10–15 mm below the railhead from cavities caused by hydrogen. They can occur in the rail or in welds from poor welding practice. Development of tache ovale is influenced by thermal or residual stresses from roller straightening. Squats\textsuperscript{42,43} occur on tangent tracks and in curves of large radius on the railhead and are characterised by the darkened area on the rail. Squats are surface initiated defects that can initiate from a white etching martensitic layer on the surface of the rail. Other mechanisms of squat formation are linked to longitudinal traction by wheels, which cause the surface layer of material to plastic ratchetting until a crack develops at the rail head.

Rolling contact fatigue cracks on wheels can be classified as shelling and spalling. Shelling is a subsurface rolling contact fatigue defect that occurs on the wheel thread and the mechanism is similar to the formation of shelling in rails. Spalling (Bartley\textsuperscript{46}) can be initiated on the wheel thread surface when the wheel experiences gross sliding on the rail (braking). Large wheel surface temperatures above the austenization limit (720°C) can form martensite, a hard brittle steel phase. This brittle phase will easily fracture under following wheel passages and eventually result in spalling.

Surface coating of the track has been shown to reduce the advent of RCF cracking in the laboratory and full-scale tests are currently underway to establish if this behaviour is replicated in the field.\textsuperscript{47}
IV. FRICTION

The friction force can be defined as the resistance encountered by one body moving over another body. This definition covers both sliding and rolling bodies. Note that even pure rolling nearly always involves some sliding and that the two classes of motion are not mutually exclusive. The resistive force, which is parallel to the direction of motion, is called the friction force. If the solid bodies are loaded together, the static friction force is equal to the tangential force required to initiate sliding between the bodies. The kinetic friction force is then the tangential force required to maintain sliding. Kinetic friction is generally lower than static friction.

For sliding bodies, the friction force, and thereby the coefficient of friction (friction force divided by normal force), depends on three different mechanisms in dry and mixed lubricated conditions: deformation of asperities, adhesion of the sliding surfaces, and ploughing caused by deterioration particles and hard asperities.\textsuperscript{48} For most metal pairs, the maximum value of the coefficient of friction ranges from 0.3 to 1.0.\textsuperscript{49} The ploughing component of the coefficient varies from 0 to 1.0 and the adhesion component varies from 0 to 0.4.\textsuperscript{50} It is generally recognised that friction due to rolling of nonlubricated surfaces over each other is considerably less than dry sliding friction of the same surfaces.\textsuperscript{51} For the steel wheel–steel rail contact, the rolling coefficient of friction is of the order of $1 \times 10^{-4}$.

As shown in Figure 5.12, the contact area between a wheel and rail can be divided into stick (no slip) and slip regions. Longitudinal creep and tangential (tractive) forces arise due to the slip that occurs in the trailing region of the contact patch. With increasing tractive force, the slip region increases and the stick region decreases, resulting in a rolling and sliding contact. When the tractive force reaches its saturation value, the stick region disappears, and the entire contact area is in a state of pure sliding. The maximum level of tractive force depends on the capability of the contact patch to absorb traction. This is expressed in the form of the friction coefficient, $\mu$ (ratio of tractive force to normal load, $N$). Normally, wheel–rail traction reaches a maximum at creep levels of 0.01 to 0.02.

The traction/creep curve can be dramatically affected by the presence of a third body layer in the wheel–rail contact. This could be formed either by a substance applied to increase/decrease

![Figure 5.12](image-url)
friction (friction modifier or lubricant), or by a naturally occurring substance acting to decrease friction (water or leaves etc.). Hou et al.\textsuperscript{52} have proposed a frictional model for rolling-sling contacts separated by an interfacial layer, which is based on the three rheological parameters: the shear moduli of elasticity ($G$) and plasticity ($k$) and the critical shear stress ($\tau_c$). It shows that the friction is greatly affected by the rheology of the third body, slip distance and load with the shear stress vs. slip distance relationship exhibiting the dominant influence.

A. Wheel–Rail Friction Conditions

The friction between the wheels and rail is extremely important as it plays a major role in the wheel–rail interface process such as adhesion, wear, rolling contact fatigue, and noise generation. Effective control of friction through the application of friction modifiers to the wheel–rail contact is therefore clearly advantageous, although the process has to be carefully managed. The aim of friction management is to maintain friction levels in the wheel–rail contact to give\textsuperscript{53}:

- Low friction in the wheel flange–rail gauge corner contact.
- Intermediate friction wheel tread-rail top contact (especially for freight trucks).
- High friction at the wheel tread-rail top contact for locomotives (especially where adhesion loss problems occur).

Ideal friction conditions in these contact regions for high and low rails are shown in Figure 5.13.\textsuperscript{54} These are similar to values quoted for Canadian Pacific.\textsuperscript{55}

Olofsson and Telliskivi\textsuperscript{13} compared coefficients of friction measured on track and in the laboratory. For pure nonlubricated sliding tests the level is roughly the same, varying between 0.5 and 0.6. For a full-scale lubricated rail, the coefficient of friction was lower and varied between 0.2 and 0.4. Other results found in the literature support the measured coefficients of friction from the full-scale tests. In another project, the Swedish National Rail Administration studied how leaves on the track influenced the coefficient of friction using a special friction measurement train.\textsuperscript{56} The reported coefficient of friction varied between 0.1 and 0.4. Harrison et al.\textsuperscript{57} compared a hand-pushed rail tribometer and a TriboRailer that operated from a companion vehicle. For the hand-pushed tribometer, the coefficient of friction was typically 0.7 under dry conditions and varied between 0.25 and 0.45 under lubricated conditions. The Triborailer presented lower values of the coefficient of friction. Under dry conditions, the coefficient of friction was approximately 0.5, and varied between 0.05–0.3 under lubricated conditions.

To gain the greatest benefit from friction management and to ensure efficient train operation, the coefficient of friction needs to be integrated with the overall wheel–rail management system. It has been noted that, for example, it should be closely tied in with grinding schedules used in the maintenance of wheels and rails.\textsuperscript{53} A scheme for the systematic approach to wheel–rail interface

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure5.13.png}
\caption{Ideal friction coefficients in the wheel–rail contact.}
\end{figure}
research and development is shown in Figure 5.14, which emphasises the consideration of all aspects including materials, dynamics, etc., as well as friction. Nothing can really be treated in isolation.

B. FRIC TION MODIFICATION

Friction modifiers can be applied to the wheel–rail contact to generate the required coefficients of friction. These can be divided into three categories:\n
- Low coefficient friction modifiers (lubricants) are used to give friction coefficients less than 0.2 at the wheel flange–gauge corner interface.
- High friction modifiers with intermediate friction coefficients of 0.2–0.4 are used in wheel tread-rail top applications.
- Very high friction modifiers (friction enhancers) are used to increase adhesion for both traction and braking.

Low friction modifiers can be solid or liquid (greases), the main difference between the two being the thickness of the film they form in the wheel/rail contact (solid lubricants will give a film of 10–30 \( \mu m \) and grease lubricants less than 5 \( \mu m \)). The primary application of these modifiers is in reducing friction in the wheel flange–rail gauge corner contacts, particularly in curves, where the contact conditions can be quite severe. The main focus of the remainder of this section is on low friction conditions and how to deal with them. Further discussion relating to reduction of friction can be found in the subsequent section on lubrication.

Friction modifiers are classified according to their influence after full slip conditions have been reached in the wheel–rail contact, as shown in Figure 5.15. If friction increases after the saturation point, the modifiers have positive friction properties, if friction reduces, the modifier has negative friction properties. Positive friction modifiers can be described as high positive friction (HPF) or very high positive friction (VHPF), depending on the rate of increase in friction.
Loss of friction or adhesion between the wheel and rail is particularly important as this has implications for both braking and traction. Poor adhesion in braking is a safety issue as it leads to extended stopping distances, and also in traction as it may lead to reduced acceleration which will increase the risk of a rear collision from a following train. In traction, however, it is also a performance issue. If a train experiences poor adhesion when pulling away from a station and a delay is enforced, the train operator will incur costs. Similar delays will occur if a train passes over areas of poor adhesion while in service.

A great deal of research was carried out on adhesion loss in the U.K. during the 1970s using both laboratory and field tests. This identified the major causes of adhesion as being: water (from rainfall or dew), humidity, leaves, wear debris, and oil contamination.

Relative humidity has been shown to influence the frictional behaviour of a wide variety of materials. By increasing the relative humidity, an absorbed layer of water molecules can be produced that can modify frictional behaviour. Relative humidity effects may also produce new chemical reactions on the surface together with other added substances.

The problems caused by leaves on the line remain prevalent today and each autumn can cause considerable delays to trains on the U.K. rail network. They are also a problem in Sweden, where it has been estimated by the Swedish National Railroad Administration that the cost of leaves on the rails is 100 million SEK annually (9 SEK = 1 EUR).

Work carried out on Japanese, American, and Canadian railways has re-emphasised the effect of the problems outlined above and identified further causes of adhesion loss, such as frost and mud deposited on rails by car wheels passing over level-crossings. This work also showed the varying effects on adhesion of different types of leaves. Oily leaves, such as pine and cedar, caused a larger decrease in adhesion. Tunnels were also highlighted as being a problem, especially where water was leaking onto the track. Full-scale testing has also shown that weather conditions affect both the coefficient of friction and wear rates.

Most of the work carried out in the U.K. was at relatively low speeds. Work on adhesion issues related to high speed lines, using both full-scale roller rigs and field measurements, has shown that adhesion decreases with train velocity and wheel–rail contact force.

A number of experimental and theoretical investigations have revealed other significant parameters affecting adhesion. Chen et al. carried out a detailed theoretical investigation of a water lubricated contact, studying the effect of rolling speed, slip, load, surface roughness, and
water temperature. The results indicated that the biggest influence on adhesion was the roughness of the wheel and rail surfaces (with adhesion rising with increased roughness). Third body effects due to material generated within the wheel–rail contact have been characterised by Niccolini and Bertier and due to externally applied materials by Hou et al. These can have a large influence on the adhesion, which is heavily dependent on the rheological properties of the layer formed in the contact. There is only limited data to validate these studies, but they give an important insight to aspects of the problem that are harder to evaluate in the field.

D. INCREASING ADHESION

While conditions leading to poor adhesion have been well investigated, methods for addressing the problems have not. The main adhesion enhancer used on railway networks worldwide is sand. Sanding is used in train operations to improve adhesion in both braking and traction. In braking it is used to ensure that the train stops in as short a distance as possible. It usually occurs automatically when the train driver selects emergency braking. Sanding in traction, however, is a manual process. The train driver must determine when to apply the sand and how long the application should last.

The sand is supplied from a hopper mounted under the train. Compressed air is used to blow the sand out of a nozzle attached to the bogie and directed at the wheel–rail contact region (see Figure 5.16). In most systems the sand is blown at a constant flow rate, but some can provide a variable flow rate.

While sanding is effective and easy to use, it can potentially cause complex and costly problems relating to both rolling stock and track infrastructure. Sand application has been shown to increase wear rates of both wheel and rail materials by up to an order of magnitude. Maintenance of sanders and control of sand build-up around track adhesion trouble spots are also issues that require particular attention.

Very high positive friction modifiers to enhance the coefficient of friction to 0.4–0.6 are available, but are really only in the development stage. There are a number of different products available, but most involve a solid stick of material that is applied directly to the wheel tread.

During autumn, when leaf fall occurs, leaf mulch is compressed in the wheel–rail contact and forms an extremely hard layer on the rail surface. This layer can cause adhesion loss problems, as already mentioned, but is also extremely hard to remove. A number of methods are used including using high pressure water-jets and blasting with Sandite (a mixture of sand and aluminum oxide particles), and a new system has now been developed that involves using a high power laser to burn away the layer. All of these, however, in the U.K., are applied by maintenance trains, of which there

![FIGURE 5.16 Sanding apparatus.](image-url)
are very few, and gaining track access is extremely difficult. Water-jets and Sandite also have
knock-on effects, which may be detrimental to the track infrastructure.

V. LUBRICATION AND SURFACE COATINGS

This section focuses on the problems of high friction coefficients and how to reduce them using
lubrication. High friction coefficients are most prevalent at the wheel flange–rail gauge corner
contact, particularly in curves. Load and slip conditions are also high, which means that wear and
rolling contact fatigue are more likely to occur at these sites. In order to reduce the wear problems,
lubrication can be applied to reduce friction and alter the load bearing capacity. Lubrication,
however, is also applied to alleviate other problems as will be shown. Surface coatings have also
been applied to the track to address the problem of high friction.

A. BENEFITS OF LUBRICATION

The benefits of lubrication have been well documented and are concerned with the reduction of:

- wheel flange and rail gauge corner wear
- energy consumption
- noise generation.

Laboratory and field tests have all shown the wear reducing benefits of lubrication in the wheel–rail contact.

Fuel savings of approximately 30% (compared to dry conditions) have been reported for
measurements taken on test tracks. Other studies carried out in the field have shown
improvements of a similar order of magnitude.

B. METHODS OF LUBRICATION APPLICATION

There are a number of different ways to apply lubricant:

- **Mobile lubricators:** these are basically railway vehicles designed to apply lubricant to the
gauge corner of the track.
- **Wayside lubricators:** these are mounted next to the track and apply lubricant to the rail
gauge corner. There are three types: mechanical, hydraulic, and electronic.
- **On-board lubricators:** these apply grease or solid lubricant or spray oil on to the wheel
flange, which is then transferred to the gauge corner of the rail. Complex control systems
are used in the application process to avoid the application of lubricant at inappropriate
locations.

Mechanical wayside lubricators rely on the wheel making contact with a plunger, which
operates a pump. The pump supplies lubricant from a reservoir to a distribution unit. The lubricant
is then picked up by the wheel flange and distributed along the rail. Problems exist because there is
only a single circuit so if a failure occurs the lubricant supply is ineffective. Mechanical lubricators
have a low initial cost because of their simple design, but require good maintenance to remain
effective. Hydraulic lubricators have been found to more reliable, but have some of the same
problems as their mechanical counterparts.

Electronic lubricators use sensors to detect the approach of a train and activate electric pumps to
deliver the lubricant. They are inherently more reliable than mechanical or hydraulic lubricators
and can also be adjusted away from the track.

On-board lubricators supply lubricant to the wheel flange–rail gauge corner. In most designs
the lubricant is deposited on the wheel flange and spread along the rail, although in some the
lubricant is directly applied to the rail. Grease or oil spray systems are used that employ complex control strategies using sensors measuring vehicle speed and track curvature to govern lubricant application. Solid stick lubricators are also available, in which a stick of lubricant is spring loaded against the wheel flange.

On board systems have a number of advantages over wayside lubricators\(^54\):

- reduced safety risk exposure to staff during installation, inspection and maintenance.
- easier inspection and maintenance (carried out in more controlled conditions).
- the rail will continue to receive some friction control protection in the event of the failure of an individual on board lubricator.

Despite these advantages, at problem tracks, site wayside lubricators will still be a necessity.

C. Problems with Lubrication

Problems with lubrication systems have been found to be related to both technical and human issues.\(^84\) The main technical problems with wayside lubricators have been highlighted as: blocked applicator openings; leaking holes; ineffective pumps and trigger mechanisms; and poor choice of lubricant. Human related problems can result from the technical issues. If over lubrication occurs and lubricant migrates onto the rail top, adhesion loss can occur. Train drivers may then be tempted to apply sand to compensate and increase friction, however, this will lead to increased wear and could cause the applicators to become blocked. The thought that the application of lubricant will lead to wheel slip can also lead train drivers to switch off on board lubrication systems.

Some of the consequences of poor wayside lubrication have been listed as\(^55\):

- wheel slip and loss of braking (and potentially, wheel flats and rail burn)
- poor train handling
- prevention of ultrasonic flaw detection
- wastage of lubricant
- high lateral forces in curves and subsequent increase in wear.

Other than adhesion problems, over-lubrication can cause an increase in rolling contact fatigue crack growth on the rail gauge corner.\(^22,23\) This can be due to pressurisation of the crack leading to increased growth rates or because reduced wear means that cracks are truncated less. However, full-scale test results from narrow curves show that well-maintained lubrication could reduce both the wear rate and the propagation rate of surface cracks.\(^14\)

D. Lubricator System Selection and Positioning

The effectiveness of a lubrication system is affected by a number of parameters including the climate, the railway operating conditions, the dispensing mechanism, and the maintenance of the lubricating equipment. Clearly, selection of the most appropriate type of lubricator and lubricant is very important, but also the positioning of the lubricator is critical to its successful operation.

The key characteristics required of a lubricant are\(^55\):

- Lubricity or the ability of the lubricant to reduce friction (although of greater importance is the effect on wear).
- Retentivity or the measure of time over which the lubricant retains its lubricity. Flash temperatures in the wheel–rail contact can be as high as 600 to 800°C, these lead to the lubricant in the contact being burned up. The retentivity is therefore a function of the
loads and creepages seen at the lubrication site as these dictate the temperature in the contact.

- Pumpability or how easily the lubricant can be applied to the track. The temperature is an issue here as some track locations will experience a wide range across which some lubricants may not maintain their pumpability. Some networks use different lubricants in the winter and summer for this reason.

Laboratory tests have been developed to assess the wear reducing capacity of lubricants and energy saving potential. These are good for screening and ranking purposes and selecting those lubricants suitable to take forward for field trials.

Monitoring the effectiveness of lubricants in the field, either during trials or in actual practice, is clearly essential. This will provide information necessary to decide on a lubrication strategy during trials or in maintaining performance once implemented. Measurements of friction can be taken using tribometers, either hand-propelled along the track or train mounted. The hand-propelled equipment is useful for monitoring short stretches of track. Obviously, for long stretches, a train or vehicle mounted system is preferable. It should be noted here that it has been shown that the benefits of lubrication may take some time to become evident on installation of a lubrication system.

Correct positioning of a wayside lubricator is critical to providing effective lubrication. Each site will require something different, which makes this task quite complex. Controlled field testing has been used to assess the reliability and efficiency of wayside lubricators based on a number of factors related to the lubricant including: waste prevention; burn up; distance covered; washing off by rain or snow; and migration to the rail top. This data and factors related to the track, such as length of curve, gradient and applicator configuration, and traffic, including direction, types of bogie, axle loads and speeds, have been combined to develop criteria and a model for positioning wayside lubricators.

Ultimately, however, the most critical element in preserving effective lubrication is maintenance. Once in place, wayside lubricators need regular maintenance to prevent the problems outlined occurring.

E. SURFACE COATINGS

Coating the rail surface is now being investigated as a means to control friction and reduce wear, rolling contact fatigue and noise problems. This is quite new technology, coatings have previously been applied to wheels to address these problems. If successful, this could address some of the problems evident with lubrication supply and reduce maintenance requirements.

ACKNOWLEDGMENTS

This work formed part of the activities of the Railway group KTH at the Department of Machine Design KTH in Stockholm Sweden. The partners in the Railway group KTH are Banverket, Green Cargo, SL Infrateknik, Bombardier Transportation, SJ and Traintech Engineering.

REFERENCES


22. Cassidy, P. D., *Variation of Normal Contact Stresses for Different Wheel/Rail Profile Combinations*, European Rail Research Institute, ERRI D 173/DT 336, 1996.


