

REVIEW OF WHEEL-RAIL ROLLING CONTACT THEORIES

J. J. Kalker

Department of Mathematics
Delft University of Technology
Delft, The Netherlands

SUMMARY

In the present paper, the rolling contact theories that emerged principally in the last two decades are described with special emphasis on their performance and on the ideas underlying them.

Although a number of these theories are obsolete, they all serve as a mine of ideas for those who seek to improve the present-day theory of rolling contact.

NOMENCLATURE

Symbol	Where defined	Symbol	Where defined
a, b, a', b'	(3), fig. 5	$\dot{s}$	(2)
a_{mn}, b_{mn}	(26)	t	(1)
B, C, D	(10)	u	(8)
$C_{ij}, i, j=1, 2, 3$	(18)	$\bar{v}$	(1)
$\underline{c}$	fig. 1, sec. 2	$\underline{v}$	fig. 1, sec. 2
d_{pq}, e_{pq}	(25)	$\bar{w}$	(4)
$\underline{F}(F_x, F_y)$	(6), (13), (17)	(X, Y)	(5), (6)
f	(5)	x, y, z	(1)
G	(8)	x', y', z'	fig. 1, sec. 2
g	(23)	x_y	above (16)
$\underline{g}(y)$	(15)	Z	(3)
K	(28)	ξ, η, τ, χ	(12), (21), (22)
L, L_x, L_y	(7), (34)	σ	(8)
N	(3)	τ	above (33)
$\underline{p}(X, Y)$	(5)	v_x, v_y, ϕ	(2)
r	(30)	Φ, ψ_1	(11)
$\underline{r}$	(27)		

1. INTRODUCTION

Centrally in the study and realization of rail vehicle simulation stands the problem of the frictional contact between wheel and rail, which may be stated as follows: What is the connection between the motion of the wheel over the rail, and the tangential force the rail exerts on the wheel?

Many theories have been proposed to deal with this problem.

It is the purpose of this paper to review them with emphasis on the ideas under-

lying them, and on the speed of their realization, which is very important as the solution of the stated problem is required very many times during a simulation. The paper is organized as follows:

In section 2 the problem is stated, and some concepts are introduced which are necessary to understand the theories.

In section 3 we treat theories whose domain of application is restricted, e.g. to small creepage and spin. These theories give the force-creep relation explicitly, so that their application is virtually instantaneous.

In section 4 we treat numerical theories whose domain of application is unrestricted and which are what is termed "exact" (see sec. 2), and which are slow in operation.

In section 5 we review realizations of the so-called simplified theory (see sec. 2), and finally,

in section 6 theories which fit the curves produced by experiments or other theories.

We conclude with a tabulated bird's eye view of all theories.

The paper is intended to give an overview of all significant theories of the past. A coherent account of the final product of these theories, viz. present day rolling contact theory, is provided by [7] which is recommended as a collateral reading. In effect, sec. 2 and the beginning of sec. 3 are taken directly from [7].

2. STATEMENT OF THE PROBLEM

Consider a rail, see fig. 1.

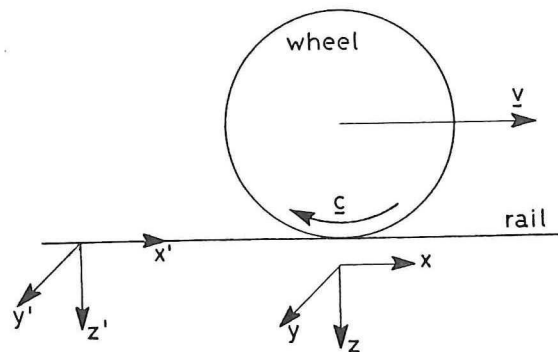


Fig. 1. A wheel rolling over a rail.

A Cartesian coordinate system is attached to it of which the x' -axis points along the rail, the z' -axis points vertically downwards, and the y' -axis points to the right when one looks along the direction of the x' -axis. A wheel rolls over the rail in the direction of positive x' with a vectorial rolling velocity $\underline{v}$. The (non-vectorial) rolling velocity is defined as $V = |\underline{v}|$. A new coordinate system is introduced which moves with the contact point over the rail.

$$\begin{aligned} x &= x' - Vt; \quad t: \text{time}; \quad V: \text{rolling velocity} \\ y &\text{ in plane of contact, pointing to the right, } z \text{ pointing into the} \\ &\text{ rail, see fig. 1, 2c.} \end{aligned} \quad (1)$$

In addition to a rolling velocity $\underline{v}$, the wheel has a circumferential velocity $\underline{c}$. In the point of contact, these two are almost opposite. The sum $\dot{\underline{s}} = \underline{v} + \underline{c}$ is called the rigid slip of the wheel over the rail; ordinarily, $|\dot{\underline{s}}| \ll V$, say $|\dot{\underline{s}}| \approx 0.001 V$. This rigid slip is a velocity of the wheel as a rigid body; in the (x,y) -plane of contact, it is composed of a translation velocity in the plane of x and y , and a rotation about the z -axis,

$$\dot{\underline{s}} = \underline{v} + \underline{c} = V(v_x - \phi y, v_y + \phi x) = \text{rigid slip}; |\dot{\underline{s}}| \approx 0.001 V$$

$$v_x: \text{longitudinal creepage}, v_y: \text{lateral creepage}, \phi: \text{spin} \quad (2)$$

The longitudinal creepage is connected with the difference of $\underline{c}$ and $\underline{v}$, see fig. 2a. The lateral creepage is the angle between the wheel and the plane $y = 0$, see fig. 2b. The spin is connected with the conicity of the wheel, see fig. 2c.

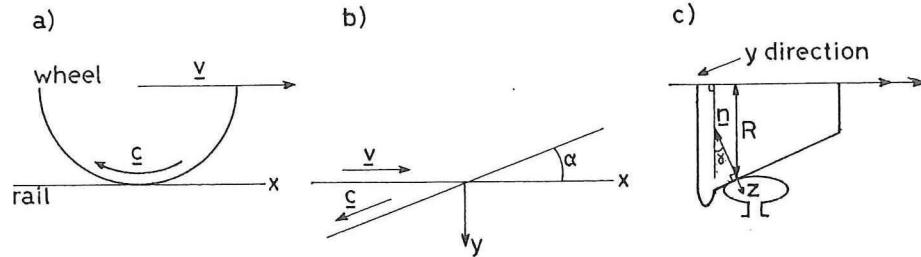


Fig. 2a. Longitudinal creepage. $v_x = (|\underline{v}| - |\underline{c}|)/V$.
 2b. Lateral creepage. $v_y = \alpha$.
 2c. Spin. Conicity: $\gamma \cdot \phi = (R \sin \gamma)/|\underline{c}| \approx (R \sin \gamma)/V$.

Owing to the elasticity of wheel and rail, the two bodies touch along a contact area, which, according to Hertz (1881) (see [1], p. 192) is elliptic in form. Indeed, the normal pressure Z exerted on wheel and rail is given by:

$$Z(x,y) = \frac{3N}{2\pi ab} \sqrt{1 - (x/a)^2 - (y/b)^2}$$

with:
 N : total normal force (i.e. force in z -direction), positive;
 a, b : semi-axes of the contact ellipse. (3)

The semi axes of the contact ellipse (a in the rolling direction, b in the other, so-called lateral direction) can in most cases*) be determined with the aid of the Hertz theory (see, e.g., [1]), once the local radii of curvature of wheel and rail are known, together with the normal force N . As to the latter, and ceteris paribus, a and b are proportional to $N^{1/3}$.

As a consequence of the compression and the friction, a deformation occurs in the wheel and the rail. We count it with respect to a reference state which is attached to the (x,y,z) coordinate system which moves with the contact. The material of wheel and rail flows through this coordinate system with a velocity $\underline{c}$ (wheel) and a velocity $(-\underline{v})$ (rail), which are both in first order equal to $(-V, 0)$ near the contact. We call $\underline{u}_w$ and $\underline{u}_r$ the surface displacement of wheel and rail in the (x,y) direction; $\underline{u} = \underline{u}_w - \underline{u}_r$ is their difference. Then the true slip $\dot{\underline{w}}$, i.e. the velocity of a particle of the wheel with respect to the contacting particle of the rail is given by the rigid slip $\dot{\underline{s}}$ plus the material time derivative $\dot{\underline{u}}$ of $\underline{u}$, where the material flows with a velocity $(-V, 0)$:

$$\dot{\underline{w}}(x,y) = V(v_x - \phi y, v_y + \phi x) + \dot{\underline{u}}(x,y,t), \text{ true slip};$$

$$= V(v_x - \phi y, v_y + \phi x) - V(\partial \underline{u}/\partial x) + (\partial \underline{u}/\partial t).$$

$$\underline{u} = \underline{u}_w - \underline{u}_r; \quad * = \text{material derivative with respect to time.}$$

In a steady state, $\partial \underline{u}/\partial t = 0$. (4)

The law of Coulomb-Amontons is commonly taken as the law governing the frictional behaviour. If $\underline{p} = (X, Y)$ are the (x,y) components of the traction at the contact exerted on the wheel, and if f is the coefficient of friction, then:

$$1. |\underline{p}| = |(X, Y)| \leq fZ.$$

$$2. \underline{\dot{w}} \neq 0 \Rightarrow \underline{p} = -fZ\dot{\underline{w}}/|\dot{\underline{w}}| \quad (\Rightarrow |\underline{p}| = fZ)$$
(5)

*) Sometimes it is not possible to apply Hertz's theory.

The contact problem may therefore be formulated as follows:
 Determine $\underline{p} = (X, Y)$, and in particular:
 $(F_x, F_y) = \iint_{\text{contact}} (X, Y) dx dy$
 so that (5) is satisfied with:

$$\dot{\underline{u}} = V(v_x - \phi y, v_y + \phi x) - V \frac{\partial \underline{u}}{\partial x} \left(+ \frac{\partial \underline{u}}{\partial t} \right)$$

where $v_x, v_y, \phi, V, N, a, b$ are given, and $\underline{u}$ and $\underline{p}$ are connected by certain constitutive relations. (6)

The connection between (v_x, v_y, ϕ) and (F_x, F_y) is called the creep-force law. When the constitutive relations in (6) are chosen as:

$$\underline{u} = \underline{u}_w - \underline{u}_r = (L_x X, L_y Y); L_x, L_y: \text{parameters} \quad (7)$$

we obtain the simplified theory, described in [2]; if the constitutive relations are derived from the theory of elasticity, which gives in the simplest case (see Love [1], p. 243, and Kalker [4], p. 21-22):

$$\underline{u}(x, y) = \frac{1}{\pi G} \iint_{\text{contact area}} \left\{ \left(\frac{1-\sigma}{R} + \frac{\sigma(x-x^*)^2}{R^3}, \frac{\sigma(x-x^*)(y-y^*)}{R^3} \right) X(x^*, y^*) + \right. \\ \left. + \left(\frac{\sigma(x-x^*)(y-y^*)}{R^3}, \frac{1-\sigma}{R} + \frac{\sigma(y-y^*)^2}{R^3} \right) Y(x^*, y^*) \right\} dx^* dy^*$$

with $R = \sqrt{(x-x^*)^2 + (y-y^*)^2}$; G : modulus of rigidity, $G = \frac{E}{2(1+\sigma)}$; σ : Poisson's ratio; E : Young's modulus. (8)

we speak of an exact theory.

A theory is called dynamic if inertial effects are taken into account. If such effects are neglected, we have a quasistatic theory. As inertial effects become only noticeable in contact theory for train speeds of over 500 km/h [3], there is at present no need for a dynamic contact theory in rail vehicle dynamics, and indeed there exists none. All theories mentioned in this paper, including simplified theory, are quasistatic in character.

All exact theories so far employ the half-space assumption, which means that as far as elastic effects near the contact are concerned, the wheel and the rail are regarded as the elastic half-spaces $z \geq 0$ (rail) and $z \leq 0$ (wheel).

The background is that specific contact effects due to the extensiveness of the contact have died out at approximately 3 contact widths away from the contact. Beyond that distance, the total force (F_x, F_y, N) may be regarded as a point load. With a contact area of 3 cm², the real dimensions of wheel and rail imply that the condition is approximately satisfied.

A theory is called three-dimensional if the $\underline{u}, \underline{p}$ dependence on all three coordinates (x, y, z) is taken into account. It is called two-dimensional if $\underline{u}$ and $\underline{p}$ are taken independent of the lateral coordinate y . Note that owing to the presence of the term ϕy in the expression (6) for the true slip $\dot{\underline{u}}$, there is no place for the spin ϕ in two-dimensional theory, so that such a theory has only limited application.

When in (6) $\partial \underline{u} / \partial t = 0$, we speak of steady state rolling contact. When $\partial \underline{u} / \partial t \neq 0$, we have transient rolling contact. As far as train dynamics is concerned, transient contact problems are of minor significance. This follows from Kalker's experience with two-dimensional [5], three-dimensional ([6], section 5), and simplified ([2], section 4) transient theories, that when the rigid slip $\dot{s}$, see (2), is constant in time, transient effects have died out after a distance of approximately 2 contact widths has been traversed. Transient theories are important however; the best three-dimensional steady state theory [6, 7, 8] is based on it.

3. THEORIES WITH RESTRICTED DOMAIN OF APPLICATION

The first theory of rolling contact with friction was developed by Carter [9] in 1926, in connection with vehicle dynamics: the paper is called "On the Action of

a Locomotive Driving Wheel". It is a two-dimensional theory; the wheel is globally regarded as a cylinder, and the rail as a very thick plate. The half-space assumption is employed, and only longitudinal creepage v_x is taken into account. Carter achieves an exact solution of this problem; a typical local loading is shown in fig. 3a, and the creep-force law is shown in fig. 3b.

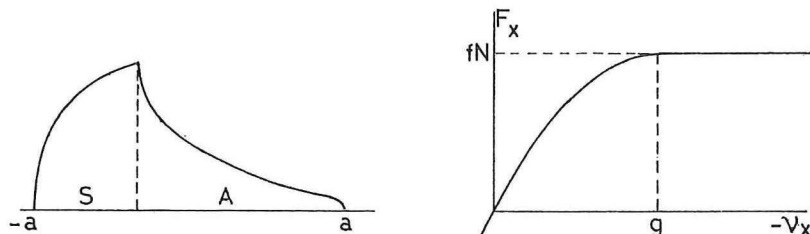


Fig. 3a. Local traction distribution according to Carter. S: area of slip, where $\dot{w} \neq 0$; A: area of adhesion, where $\dot{w} = 0$.

3b. F_x dependence on v_x (Carter). $g = 4fa/\rho$: at this point, complete sliding sets in. $a \approx 1$ cm, $f \approx 0.3$, $\rho = 2 \times \text{diameter wheel} = 200$ cm $\Rightarrow g = 0.006$.

Frommm, in 1927 [10] came up with the same solution. Lateral creepage v_y was taken into account by Kalker in 1967 [11], who gave a simple approximate solution very similar to Carter's, and, in 1967, by Heinrich and Desoyer [12], who gave an exact, but very complicated solution.

We will demonstrate the method of solution with the aid of simplified theory. To that end we will need the normal pressure between two smooth Winkler foundations, see fig. 4.

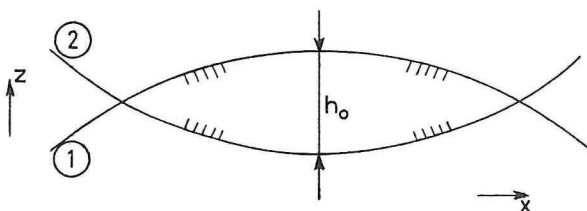


Fig. 4. Two Winkler foundations.

The foundations, numbered 1 and 2, have surfaces $z = -A_1x^2$, $z = A_2x^2$ respectively. We start from an unstressed state in which the undeformed foundations intersect to a depth h_0 . At the point x , the depth of penetration is $h(x) = h_0 + (A_1 + A_2)x^2$. In the deformed state, a vertical displacement w occurs, divided over the two foundations, such that the deformed distance $D = h(x) - w(x) = 0$. In a Winkler foundation, the pressure Z is proportional to w , hence, $Z = B\{h_0 - (A_1 + A_2)x^2\} = C(a^2 - x^2)$ inside contact, where B and C are some positive constants, and a is the half-width of contact.

We return to Carter's theory. In simplified theory, the slip is given by

$$\dot{w} = Vv_x - V \frac{\partial u}{\partial x} = Vv_x - VL \frac{\partial X}{\partial x}.$$

It is non-zero only if $|X| = fZ = fC(a^2 - x^2)$, by Coulomb's law. Let v_x be large and positive. Under these circumstances, $\dot{w}$ will be large and positive, and hence X will be at the negative bound:

$X = -fC(a^2 - x^2) \Rightarrow \dot{w} = Vv_x - 2VLfCx$
 This solution is valid as long as $\dot{w}$ is everywhere positive in the contact, i.e. when $0 \leq Vv_x - 2VLfCa \Rightarrow v_x \geq 2LfCa \stackrel{\text{def}}{=} v_{0x}$.
 Note that the minimum value of the creepage, v_{0x} , is independent of V . When v_x falls below v_{0x} , it was Carter's find that the solution traction X could be represented by the difference of two parabolic distributions, see fig. 5, viz.

$$\begin{aligned} X &= -fC(a^2 - x^2) + fC(a_0^2 - x_0^2) < 0, \quad x_0 \stackrel{\text{def}}{=} a_0 - a + x; \quad |x_0| \leq a_0 \quad \text{zone A} \\ &= -fC(a^2 - x^2) = -fZ; \quad |x_0| \geq a_0, \quad |x| \leq a \quad \text{zone B} \\ &= 0; \quad |x| \geq a \quad \text{zone C} \end{aligned} \quad (9)$$

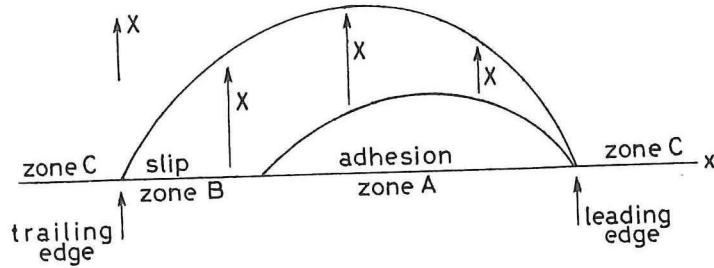


Fig. 5. Carter's solution (simplified vs.).

Indeed, in zone A, $\dot{w} = Vv_x + 2VLfC(a^2 - a) = 0$, when $a - a_0 = v_x / \{2fCL\} \Rightarrow a_0$, that is, in zone A the slip vanishes (area of adhesion, stick zone).
 In zone B, $\dot{w} = Vv_x - 2VLfCx > 0$, when $x < v_x / 2fCL = a - a_0$, so that the slip does not vanish in zone B (area of slip, slip zone).
 It is thus seen that (9) indeed constitutes the solution, as also in zone A, the stick zone, $|X| < fZ$, while in zone B, the slip zone, $|X| = fZ$.
 In Carter's version, Winkler foundations are replaced by half-spaces, and half-ellipses take the place of parabolas. The principle of the solution remains the same.

Carter's charming trick has inspired many authors to apply it to other situations. The first of these was K.L. Johnson who, in 1958, extended the method to the three-dimensional case of two spheres rolling with pure creepage (i.e. without spin) [13], and together with Vermeulen to arbitrary smooth half-spaces in 1964 [14]. The contact area of Johnson and Vermeulen, with proposed areas of slip and adhesion, is shown in fig. 6.

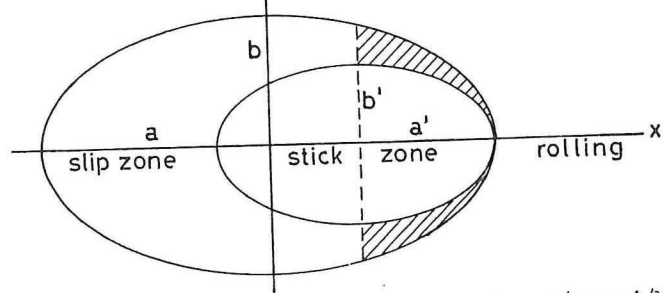


Fig. 6. The contact area according to Johnson & Vermeulen. $a/b = a'/b'$.

A semi-ellipsoidal tangential traction acts on each ellipse; the real tangential traction is found by taking the difference. This pretty picture is marred by the fact that the directions of slip and traction are incompatible in the shaded part of the area of slip. Two remarks are in order. Firstly, a simple generalization of the Carter results can in three-dimensional elasticity, be obtained only by an adhesion ellipse with the same axial ratio and orientation as the contact

ellipse, and secondly, the incompatibility of directions definitely indicates that the solution is wrong. However, Johnson performed a number of experiments of remarkably good quality from which it appears that error in the resulting force is not more than 25%. This total force has the following form:
Let a be the semi-axis of the contact area in rolling direction, and b the semi-axis of the contact ellipse in the lateral direction which lies in the plane of contact, and which is orthogonal to the rolling direction. Let further:

$$\left. \begin{aligned} B &= \int_0^{\pi/2} \cos^2 \theta \sqrt{1 - k^2 \sin^2 \theta}^{-1} d\theta \\ D &= \int_0^{\pi/2} \sin^2 \theta \sqrt{1 - k^2 \sin^2 \theta}^{-1} d\theta \\ C &= \int_0^{\pi/2} \cos^2 \theta \sin^2 \theta \sqrt{1 - k^2 \sin^2 \theta}^{-3} d\theta \end{aligned} \right\} \begin{array}{l} \text{complete elliptic integrals} \\ |k| \leq 1 \end{array} \quad (10)$$

and

$$\Phi = B - \sigma(D - C), \quad \sigma: \text{Poisson's ratio}; \quad \psi_1 = B - \sigma(a^2/b^2)C, \\ \text{where } a \leq b, \quad k = \sqrt{1 - a^2/b^2} \quad (11a)$$

$$\Phi = \{D - \sigma(D - C)\}(b/a); \quad \psi_1 = (D - \sigma C)(b/a), \\ \text{where } a \geq b, \quad k = \sqrt{1 - b^2/a^2} \quad (11b)$$

$$\xi = \pi ab G v_x / (fN\Phi), \quad \eta = \pi ab G v_y / (fN\psi_1) \\ \tau = \sqrt{\xi^2 + \eta^2}; \quad N: \text{total normal force}; \quad G: \text{modulus of rigidity} \quad (12)$$

Then, if $\underline{F}(F_x, F_y)$ is the total resulting tangential force:

$$\underline{F}/fN = \left\{ \left(1 - \frac{1}{3}\tau\right)^3 - 1 \right\} (\xi, \eta) / \tau \quad \text{if } |\tau| \leq 1 \\ = -(\xi, \eta) / \tau \quad \text{if } |\tau| \geq 1 \quad (13)$$

Owing to the fact that $\Phi \neq \psi_1$ unless $\sigma = 0$, the total force is never exactly in the direction of the creepage, however large that may be.

The theory of Carter as applied by Johnson and Vermeulen, confined as it is to the no-spin case, is of very limited use in vehicle dynamics.

We return to these formulae presently.

People continued to worry about the true shape of the area of adhesion, and in 1964 Haines and Ollerton [15], and independently Halling [31], published studies in which they considered a contact ellipse, which was relatively long in the lateral direction, so that they surmised a much slower change of the elastic field in lateral direction than in rolling direction, and on that basis they assumed that in each slice with constant y the two-dimensional Carter solution was valid, undisturbed by the behaviour at other values of y , see fig. 7.

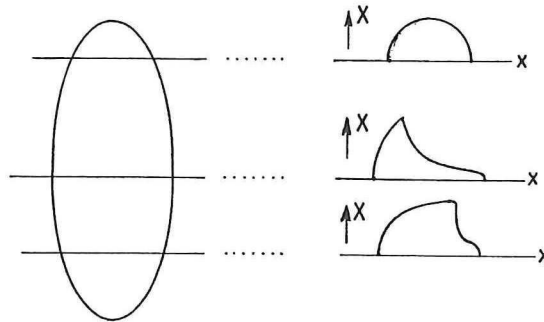


Fig. 7. The strip theory of Halling and Haines & Ollerton.

Such a theory is called a strip theory; it is mathematically justified in [30]. Note that the simplified theory is a strip theory in this sense, too. The theory of Haines and Ollerton is confined to longitudinal creepage v_x ; in 1967 Kalker extended the theory to lateral creepage v_y and (small) spin ϕ [11]. The adhesion and slip zones found with strip theory and later confirmed by other theories and by experiments, see [7], are shown in fig. 8.

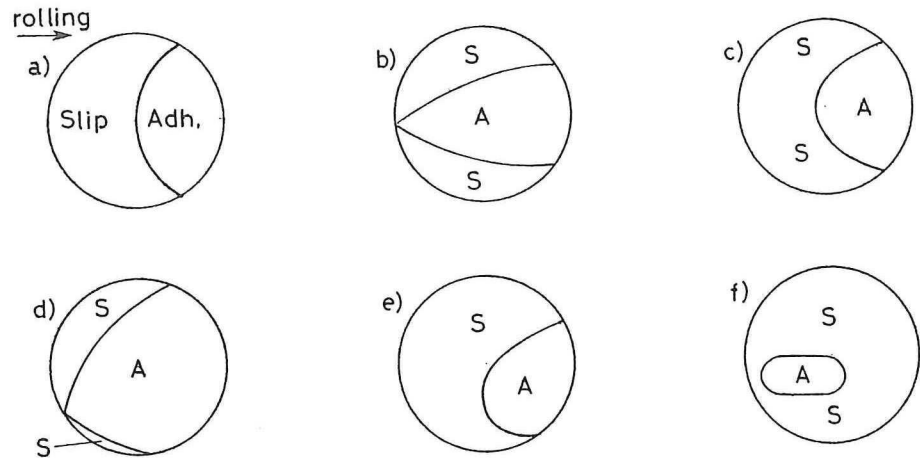


Fig. 8. Areas of slip and adhesion, as first discovered by strip theory ((f) excepted).

Strip theory is confined to contact ellipses long in the lateral direction and the spin may not exceed a fairly small number, so that it is virtually useless for vehicle dynamics. Its significance lies in that it revealed, in the mid-sixties, the true shape of the areas of slip and adhesion. Progress was made by Kalker in the early and middle sixties. His findings are laid down in his thesis [4] of 1967, in which, among other things, he published the almost complete version of the so-called linear theory of rolling contact. It is based upon an idea advocated by de Pater from 1956 onwards, that for very small creepage and spin the slip zone becomes very small indeed, so that the adhesion zone may be assumed to cover the entire contact area. The boundary conditions of steady state rolling simplify to:

$$0 = \dot{w} = V(v_x - \phi y, v_y + \phi x) - V \frac{\partial u}{\partial x} \quad \text{inside contact area} \quad (14a)$$

$$0 = p \quad \text{outside contact area} \quad (14b)$$

In his solution, Kalker integrated (14a) w.r.t. x :

$$-Vu + V(v_x x - \phi xy, v_y x + \frac{1}{2}\phi x^2) = g(y) \quad (15)$$

with $g(y)$ an arbitrary function (integration constant)

The arbitrary function $g(y)$ is determined by demanding that the traction should be continuous at the leading edge of the contact area, where the particles flow into the contact, and which is the edge with positive x . It then appears that at the trailing edge the condition $|p| < fZ$ is violated. That nevertheless this solution should be accepted, is a consequence of the following argument [7]: "A particle carrying no load enters the contact area at the leading edge. Traction builds up until the trailing edge is reached. Then the load is suddenly removed as the traction bound falls to zero very quickly, and, in the limiting case that the coefficient of friction becomes infinite, discontinuously." We demonstrate the determination of $g(y)$, see (15), with the aid of the simplified theory. To that end, we replace u by Lp , and $g = V(v_x x_y - \phi y x_y, v_y x_y + \frac{1}{2}\phi x_y^2)$, where x_y is the coordinate of the leading edge at lateral coordinate y , so that

$$p(x) = \frac{1}{2}[(v_x - \phi y)(x - x_y), [v_y + \frac{1}{2}\phi(x + x_y)](x - x_y)] \quad (16)$$

See fig. 9, where the case $v_y = 0$ is shown.

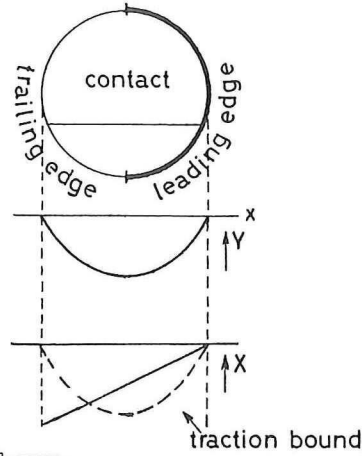


Fig. 9. Linear (simplified) theory.

Note that for all (v_x, v_y, ϕ) , $p(x_y) = 0$. In the exact linear theory, the picture is similar, except that the finite jump of X at the trailing edge is infinite, and that the finite slope of Y becomes likewise infinite. When $v_y \neq 0$, Y also has a discontinuity at the trailing edge. The resulting force $\underline{F}(F_x, F_y)$ has the following form:

$$\begin{aligned} F_x &= -abGC_{11}x & F_y &= -abG(C_{22}y + \sqrt{ab}C_{23}\phi) \\ F_x &= \iint_{\text{contact}} X dx dy & F_y &= \iint_{\text{contact}} Y dx dy \end{aligned} \quad (17)$$

where C_{11} , C_{22} , C_{23} are the so-called creepage and spin coefficients. They depend on the ratio of the axes of the contact ellipse (a/b) and on Poisson's ration σ . They are tabulated extensively in [4] p. 91 and in [7] p. 326.

It is of interest to compare the C_{11} , C_{22} of [4] and [7] with those computed from the theory of Johnson and Vermeulen, see eqns (10) - (13). As the linear theory is the asymptotic theory of small creepage and spin (de Pater [34]),

$$\begin{aligned} \{C_{11} &= -(\frac{\partial F_x}{\partial v_x}) / (abG), C_{22} = -(\frac{\partial F_y}{\partial v_y}) / (abG), \\ C_{23} &= -(\frac{\partial F_y}{\partial \phi}) / (\sqrt{ab}^3 G)\}_{v_x=v_y=\phi=0} \end{aligned} \quad (18)$$

and if C_{ij}^{JV} denotes C_{ij} according to Johnson & Vermeulen, then:

$$C_{11}^{JV}(a/b, \sigma) = \pi/\phi, C_{22}^{JV}(a/b, \sigma) = \pi/\psi_1; C_{23}^{JV} \text{ is absent} \quad (19)$$

In table 1, we compare three quantities, viz. C_{11}^{JV} , C_{11}^{Kalker} , and:

$$C_{11}^{\text{mean}}(a/b, \sigma) = C_{11}^{JV}(a/b, \sigma) C_{11}^{Kalker}(1,0) / C_{11}^{JV}(1,0) \quad (20)$$

It is seen that in the range $0.2 \leq a/b \leq 5$ C_{11}^{mean} , which is proportional to C_{11}^{JV} , has an error of no more than 7% w.r.t. Kalker's exact coefficient C^K . The error increases with the range.

So it is to be expected that the Johnson & Vermeulen formula (13) fairly accurately represents the creepage-force curve for vanishing spin ϕ :

$$\begin{aligned} F/fN &= \{(1 - \frac{1}{3}\tau)^3 - 1\}(\xi, \eta)/\tau & |\tau| \leq 3 \\ &= -(\xi, \eta)/\tau & |\tau| \geq 3 \end{aligned} \quad (13)$$

when, at any rate, we use the following definition of ξ , η , τ instead of (12):

$$\begin{aligned} \xi &= \frac{C_{11}(1,0)\phi(1,0)abGv_x}{fN\phi}, \quad \eta = \frac{C_{22}(1,0)\psi_1(1,0)abGv_y}{fN\psi_1} \\ \phi(1,0) &= \psi(1,0); \quad \tau = \sqrt{\xi^2 + \eta^2} \end{aligned} \quad (21)$$

Table 1. Comparison of the creepage coefficient according to Johnson & Vermeulen, Kalker, and (20).

	C^{JV}	$C^{(20)}$	C^K		C^{JV}	$C^{(20)}$	C^K
$C_{11} (0.2, 1/4)$	4.26	3.62	3.37	$C_{22} (0.2, 1/4)$	3.28	2.79	2.63
$C_{11} (1, 1/4)$	4.92	4.18	4.12	$C_{22} (1, 1/4)$	4.26	3.63	3.67
$C_{11} (5, 1/4)$	8.65	7.35	7.78	$C_{22} (5, 1/4)$	8.89	7.56	8.14

An alternative for (13), (20) based upon strip theory may be found in [16]. An alternative for (13), (20) based upon simplified theory may be derived from [2] p. 5, 6.

(21) can be viewed as a normalization of the creepage parameters, so that the creepage coefficients become fN , to which end the Kalker creepage coefficients can also be used directly (Hobbs [17]). This idea was extended to the spin parameter (Kalker [7], fig. 8b). In the case of vanishing spin, the theoretical creepage-force curves may be considered as functions of (ξ, η) alone. In the case of pure spin, a very clear view of the spin-force law is obtained. The normalization in question reads:

$$\xi = \frac{abGC_{11}v_x}{fN}, \quad \eta = \frac{abGC_{22}v_y}{fN}, \quad \chi = \frac{\sqrt{ab}^3 GC_{23}\phi}{fN} \quad (22)$$

In fig. 10 is shown the correspondence of the empirical formula (13), (22), with the exact numerical theory, see sec. 4.

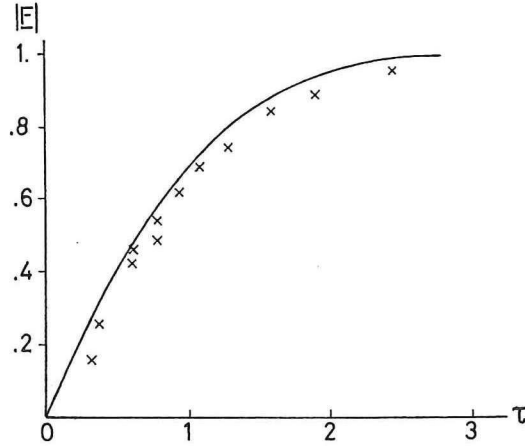


Fig. 10. Comparison of the adapted Johnson & Vermeulen theory (13), (22) with the exact numerical theory for various axial ratios.

4. EXACT NUMERICAL THEORIES

It has not proved possible to solve the rolling contact problem completely in any of the ways mentioned in the previous section. The method of Johnson & Vermeulen is confined to the no-spin case, and anyway the area of adhesion is wrong; strip theory is confined to slender contact areas and to small spin; the linear theory is confined to small creepage and spin. There exist three exact

theories which do not have this drawback, but they are numerical in nature, and slow in operation. We will treat them in turn.

4.1. The theory of Kalker's thesis (1967) [4]

The law of Coulomb may be formulated by

$$|\dot{\underline{w}}|\underline{p} + g\underline{\dot{w}} = 0 \text{ sub } |\underline{p}| \leq g \stackrel{\text{def}}{=} fZ; \underline{\dot{w}}: \text{ see (4)} \quad (23)$$

Proof:

a) Assume $|\dot{\underline{w}}| = 0$. Then $|\underline{p}| \leq g$.

b) Assume $|\dot{\underline{w}}| \neq 0$. Then $\underline{p} = -g\underline{\dot{w}}/|\dot{\underline{w}}| \Rightarrow |\underline{p}| = g$.

a) and b) together constitute Coulomb's law. Note that in (23) no distinction is made between stick and slip zones, so that these do not figure in the calculation. QED

We now try to represent $\underline{p}$ and $\underline{\dot{w}}$ by a linear combination of basis functions in such a way, that (8) is satisfied. In Kalker's thesis, the following basis functions are taken:

$$T_{pq} = \sqrt{1 - x^2/a^2 - y^2/b^2} x^p y^q, (x, y) \in K, = 0 \text{ outside } K$$

a, b: semi-axes of contact ellipse; x: rolling direction (24)

and

$$\{X(x, y), Y(x, y)\} = \sum_{p=0, q=0}^{p+q=M} T_{pq} \{d_{pq}, e_{pq}\} \quad (25)$$

It is shown in [4] that the tangential displacement is then given by a polynomial:

$$\underline{u} = \{u_x(x, y), u_y(x, y)\} = \sum_{m=0, n=0}^{m+n=M} x^m y^n \{a_{mn}, b_{mn}\}, (x, y) \in K \quad (26)$$

where $\{a_{mn}, b_{mn}\}$ are linear combinations of the $\{d_{pq}, e_{pq}\}$, the precise form of which is given in [4].

If we represent $\underline{p}$ in the manner of (25), we cannot satisfy the boundary condition (23) exactly:

$$|\dot{\underline{w}}|\underline{p} + g\underline{\dot{w}} = \underline{r} \text{ sub } |\underline{p}| \leq g; \text{ residual } \underline{r} \text{ generally } \neq 0 \quad (27)$$

We determine $\{d_{pq}, e_{pq}\}$ by means of a least squares fit of the residuals $\underline{r}$:

$$\min_{\substack{d_{pq}, e_{pq} \\ \text{sub } |\underline{p}| \leq g}} J_T = \iint_{\text{contact } K} |\underline{r}|^2 dx dy \quad (28)$$

(28a)

In Kalker's thesis (28a) is omitted because of the difficulty of its application to loads of the form (25). It turned out that reliable results could only be obtained when $b \geq a$.

It was conjectured in 1977 that this unreliability could be removed if (28a) was taken into account. To that end, and also to obtain more flexibility, the contact area was approximated by a net of rectangles, see fig. 11, and the tangential traction was taken constant over each mesh.

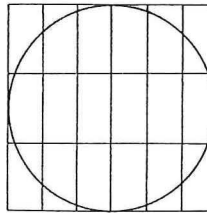


Fig. 11. Contact area discretized by rectangles.

A program called XCTROL was generated on these principles, but it turned out to be just as unreliable as the thesis program; XCTROL is briefly described in [6].

Another attempt was made in 1972 by Goedings & Kalker. They started from the following representation of Coulomb's law:

$$\underline{p} \cdot \dot{\underline{w}} + g|\dot{\underline{w}}| = 0 \text{ sub } |\underline{p}| \leq g; \dot{\underline{w}}: \text{ see (4)} \quad (29)$$

Proof:

a) If $|\dot{\underline{w}}| = 0 \Rightarrow |\underline{p}| \leq g$.

b) If $|\dot{\underline{w}}| \neq 0 \Rightarrow g = |(\underline{p} \cdot \dot{\underline{w}})|/|\dot{\underline{w}}| \leq |\underline{p}| \cdot |\dot{\underline{w}}|/|\dot{\underline{w}}| = |\underline{p}| \leq g$, from which it follows that $|\underline{p}| = g$ and $\underline{p} = -g \cdot \dot{\underline{w}}/|\dot{\underline{w}}|$;

a) and b) together constitute Coulomb's law. QED

Finally we observe that:

$$|\underline{p}| \leq g \Rightarrow \underline{p} \cdot \dot{\underline{w}} + g|\dot{\underline{w}}| = r \geq 0 \quad (30)$$

so that we can minimize

$$\min! J_{KG} \stackrel{\text{def}}{=} \iint_K r \, dx dy \text{ sub } |\underline{p}| \leq g \quad (31)$$

As in (27) - (28), the method does not distinguish between areas of slip and adhesion.

A program was generated [18, 19] to implement (31). The discretization of fig. 11 was used. The term $\dot{\underline{w}}$ occurring in (30) proved very difficult to handle; finally, it was replaced by:

$$|\dot{\underline{w}}| = \sqrt{\dot{w}_x^2 + \dot{w}_y^2} \approx \sqrt{\dot{w}_x^2 + \dot{w}_y^2} + \epsilon, \epsilon \rightarrow 0 \quad (32)$$

The limiting process indicates a sequential method. The inequalities $|\underline{p}| \leq g$ were taken into account by a penalty function method, which is also sequential. The performance of this program is disappointing, as it is very slow. By and large it is reliable, but in some cases the sequential technique fails. The unreliability of the thesis program, and the slowness and obvious inefficiency of the program based on (29) - (31), induced Kalker in 1977 to construct a new program without these defects. The first result was XCTRL, described above, which is based in (27) - (28). XCTRL proved just as unreliable as the program it intended to replace, and therefore an entirely new option was exercised, viz. the variational principle of Duvaut & Lions. This principle solves the following problem:

Consider two bodies in frictional contact over $K(t - \tau)$ at the time $(t - \tau)$, where t is an instant of time, and τ a small, positive time interval. The particles, which were at the surface point (x, y) in an unstressed state, have undergone a small elastic displacement $\underline{u}(x, y, t)$ and $\underline{u}(x, y, t - \tau)$ relative to each other at times t and $t - \tau$. Moreover, the bodies as a whole have a relative velocity $\dot{\underline{s}} = \dot{\underline{s}}(x, y, t)$, the rigid slip of eq. (2). Assuming that $\underline{u}(x, y, t - \tau)$ and $\dot{\underline{s}}$ are known, it is required to find the tangential traction $\underline{p}(x, y, t)$ which may not exceed the traction bound g , and the elastic displacement $\underline{u}(x, y, t)$ which is connected with $\underline{p}$.

When the bodies are half-spaces, the principle of Duvaut & Lions, which is actually the principle of virtual work (Kalker [21]), reads:

$$\left. \begin{array}{l} \min! \iint_{K(t-\tau)} \underline{p}(x', y', t) \{ \frac{1}{2} \underline{u}(x', y', t) - \underline{u}(x', y', t-\tau) - \dot{\underline{s}}(x', y', t)\tau \} dx' dy' \\ \text{sub } |\underline{p}| \leq g = fZ; \underline{u}: \text{ see (8)}; x', y': \text{ see sec. 2, fig. 1 (coordinates} \\ \quad \text{attached to the rail)} \end{array} \right\} \Rightarrow \underline{p} \quad (33)$$

The principle is used to simulate rolling in the following manner.

$\dot{\underline{s}}(x, y, t) = \dot{\underline{s}}(x, y)$ is given throughout, independent of t in a coordinate system which is attached to the contact area, see sec. 2, fig. 1. Set $t = 0$. Start with an arbitrary $\underline{p}(x, y, 0)$, $|\underline{p}| \leq g$. Find $\underline{u}(x, y, 0)$ from (8). Take a step ($t = t + \tau$). Determine $\underline{p}(x, y, t)$ from (33). This determination is eased by the uniqueness of the minimum of (31), and hampered by the non-linearity of the constraints $|\underline{p}| \leq g$. If $\underline{p}(x, y, t) \neq \underline{p}(x, y, t - \tau)$ the steady state has not yet been reached, and a new step is taken. If $\underline{p}(x, y, t) = \underline{p}(x, y, t - \tau)$, a steady state has been reached. In order to remove the effect of the finiteness of τ , the

process is repeated once with $\tau = \tau/2$, and the resulting total forces $\underline{F}$ are extrapolated to $\tau = 0$.

It is seen that this, again, is a sequential program. It has not proved possible to devise a virtual work based variational principle which yields the steady state directly.

The program DUVOROL is almost fully reliable. Only in very few cases it gives inaccurate results in a standard discretization. It is only slightly slower than XCTROL. For further details we refer to [6, 7, 8].

5. THREE VERSIONS OF THE SIMPLIFIED THEORY

The versions are: SIMROL, ROLCON, and FASTSIM.

SIMROL is a program written originally by Kalker in ALGOL-60, and subsequently translated into FORTRAN-4 by Goree [22]. The ALGOL version takes 2.5 sec/case on an IBM 370/158. A microprogrammed version by Demmelmeier [23] (100 ms/case) and an analog-hybrid version by Bansagi [24] (2 ms/case) also exist.

ROLCON was written by Knothe e.a. [25] in 1978 in FORTRAN-4; it is reported to be 5 times as fast as SIMROL.

Finally FASTSIM was written by Kalker in 1980. It is a very simple and extremely fast program: it is 25 times faster than SIMROL. It is currently being tested by Knothe of the TU Berlin.

Simplified theory is very popular because of the ease of understanding it and of the simplicity and speed of operation of the codes based on it. The principle of simplified theory was explained in sec. 2. The equations determining it are:

$$\text{Constitutive: } \underline{u} = L\underline{p}, \text{ with } L = \begin{pmatrix} L_x & 0 \\ 0 & L_y \end{pmatrix}; |\underline{p}| \leq g = fZ \quad (34)$$

$$\text{In stick zone: } \text{slip} = \underline{\dot{w}} \stackrel{\text{def}}{=} V(v_x - \phi y, v_y + \phi x) - V \frac{\partial \underline{u}}{\partial x} = 0 \quad (35)$$

$$\text{In slip zone: } \underline{\dot{w}} \neq 0, \underline{p} = -g\underline{\dot{w}}/|\underline{\dot{w}}| \quad (36)$$

L_x and L_y are determined so that the creepage/spin coefficients numerically coincide with the C_{ij} as calculated by Kalker [4]; see [2], [7].

In our solution, we simulate the process of rolling by following the particles along their paths in the direction of negative x . They enter the contact area at the leading edge, and we work backwards from that edge towards the interior along lines of constant lateral coordinate y . Eqns (34) - (35) may be solved explicitly.

Stick zone: $\underline{p} = (X, Y)$; $|\underline{p}|$ must be $\leq g$.

$$L_x \frac{\partial X}{\partial x} = v_x - \phi y \Rightarrow L_x X(x, y) = L_x X(x_0, y) + (v_x - \phi y)(x - x_0)$$

$$L_y \frac{\partial Y}{\partial x} = v_y + \phi x \Rightarrow L_y Y(x, y) = L_y Y(x_0, y) + [v_y + \frac{1}{2}\phi(x + x_0)](x - x_0)$$

$$(x_0, y), (x, y): \text{positions in the same stick zone, } x_0 \geq x \quad (37)$$

When the traction bound is reached at a certain x , eq. (36) is called into play. In SIMROL we set:

$$X = g \cos \theta, Y = g \sin \theta; \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = - \begin{pmatrix} \dot{w}_x / |\underline{\dot{w}}| \\ \dot{w}_y / |\underline{\dot{w}}| \end{pmatrix} \Leftrightarrow \dot{w}_y \cos \theta + \dot{w}_x \sin \theta = 0$$

($\underline{\dot{w}}$: see (8))

and we obtain an ordinary differential equation in θ . During the calculation, we repeatedly use the trigonometric functions $\cos \theta$, $\sin \theta$.

In ROLCON, $|\underline{p}| = g$ is taken into account by setting:

$$XX' + YY' = gg', \quad (' = \partial/\partial x),$$

and we use the ordinary differential equation (ODE):

$$X\dot{w}_y + Y\dot{w}_x = 0, \underline{\dot{w}}: \text{see (8)}.$$

So we have two ODE's instead of one as in SIMROL; but trigonometric functions need not be taken, which results in ROLCON being 5 times faster than SIMROL.

The fastest method is FASTSIM. In it, it is assumed that $L_x = L_y = L$. Call $\underline{\dot{s}} = V(v_x - \phi y, v_y + \phi x)$ (the rigid slip), then (35) becomes $\underline{\dot{w}} = \underline{\dot{s}} - V L \underline{p}$. Integrate this form $x + \tau$ to x , $\tau > 0$, whereby $\underline{\dot{w}}$ and $\underline{\dot{s}}$ are taken constant approximately:

$$\dot{w}\tau = \dot{s}\tau - VLp(x + \tau) + VLp(x).$$

We satisfy Coulomb's law in the following way ($p(x + \tau)$ assumed known):

$$\begin{aligned} &\text{Call } p_H = -\dot{s}\tau/(VL) + p(x + \tau) \\ &- \text{ If } |p_H| \leq g \Rightarrow p(x) = p_H \Rightarrow |p(x)| \leq g, \dot{w} = 0 \\ &- \text{ If } |p_H| > g \Rightarrow p(x) = p_H g/|p_H| \Rightarrow |p(x)| = g \end{aligned} \quad (38)$$

$$\dot{w}\tau = \dot{w}\tau - VLp(x + \tau) + VLp(x) = VL(1 - |p_H|/g)p(x) = -\lambda p(x), \lambda \geq 0.$$

from which we see that the construction (38) satisfies Coulomb's law.

A correction may be made for the non-constancy of $\dot{w}$ and $\dot{s}$ in the interval $(x, x+\tau)$ [29]. Apart from being the simplest code by far, FASTSIM is 25 times faster than SIMROL.

6. CURVE FITTING THEORIES

There have been a number of attempts to fit numerically the data on the creepage-force law which are due to experiments or to numerical theories, such as DUVOROL and SIMROL. The oldest of these attempts is due to Levi [26] (1935); it was

modified later (1950-1952) by Chartet [27]. An up-to-date version of this law reads ($v \stackrel{\text{def}}{=} |(v_x, v_y)|$):

$$\left| \frac{N}{F_x} \right|^n = \left| \frac{v}{v_x f} \right|^n + \left| \frac{N}{abGC_{11}v_x} \right|^n; \quad \left| \frac{N}{F_y} \right|^n = \left| \frac{v}{v_y f} \right|^n + \left| \frac{N}{abGC_{22}v_y} \right|^n \quad (39)$$

Levi takes the value 1 for n , and Chartet indicates the value 2.

The method is confined to the case of pure creepage; the role of spin in rail vehicle simulations had not been explicated in 1952 and note that it is only the occurrence of spin which upsets the neat picture of eq. (39) and of sec. 3!

Without spin, the law (39) is accurate enough, and the only thing needed is a theory of the C_{ij} , which was provided around 1962.

A recent attempt to use SIMROL and DUVOROL for curve fitting purposes is due to Jaschinski [28], who will talk of it himself in the present Meeting. Table books of exact rolling contact theory were constructed by Kalker [33] ("XCTROL") and by Rose of British Rail, Derby [32] by means of DUVOROL.

7. CONCLUSIONS

We tabulate the reviewed theories.

Table 2. Theories described in sec. 3

	Author(s) and Ref.	Characterization	Restrictions	Notes
2.1	Carter-type theories			
2.11	Carter [9] - Fromm [10]	Two-dimensional	Long. creepage	1, 2
2.12	Johnson & Vermeulen [13, 14]	Three-dimensional, approximate	No spin	1, 2
2.13	Haines & Ollerton [15] Halling [31], Kalker [11]	Strip theory	Contact narrow in rolling direction	2
2.14	Kalker [16, 2]	Empirical	No spin	1, 3
2.2	De Pater - Kalker [4, 7]	Linear theory	Small creepage and spin	1, 2

Notes:

1. Closed form solutions, easy to code.
2. Exact theory.
3. Simplified theory.

Table 3. Exact, numerical theories (sec. 4)

All theories of this table are unrestricted.
1 unit = 3.5 sec. on an IBM 370/158.

	Author(s), reference; name of program	Reliability	Speed
3.1	Kalker [4, 6]; XCTROL	60%	1 unit
3.2	Kalker & Goedings [18], Goree [19], "New numerical theory"	85%	30 units
3.3	Kalker [6, 7, 8]; DUVOROL	95%	1.5 units

Table 4. Simplified theory based programs (sec. 5)

Speeds will be expressed in real time, or in $\mu = 0.001$ unit, see table 3.
All programs are unrestricted.

	Author(s), reference; name of program	Reliability	Computer	Speed
4.1	Based on Kalker's program SIMROL [2]	95%		
4.11	Kalker, Goree [2, 22]	95%	Digital	700 μ
4.12	Demmermeier [23]	95%	Microprog.	100 ms
4.13	Bansagi [24]	95%	Analog	2 ms
4.2	Knothe e.a. [25]; ROLCON	100%	Digital	140 μ
4.3	Kalker [29]; FASTSIM	100%	Digital	28 μ

4. Interpolation programs

Lévi [28], Chartet [27]: Simple formulae, but no spin.

Jaschinski: [28].

Kalker [33] and Rose [32] constructed Table books of exact rolling contact theory.

REFERENCES

- [1] A.E.H. Love, A treatise on the mathematical theory of elasticity, 2nd ed., Cambridge (1926).
- [2] J.J. Kalker, Simplified theory of rolling contact. Delft Progr. Rept. 1 (1978) p. 1-10. Also as Appendix in Ref. [22].
- [3] A.L. Galin, Contact problems in the theory of elasticity. North Carolina State College, 1961.
- [4] J.J. Kalker, On the rolling contact of two elastic bodies in the presence of dry friction, Thesis Delft (1967).
- [5] J.J. Kalker, A minimum principle for the law of dry friction with application to elastic cylinders in rolling contact, J. Appl. Mech. 38 (1971) p. 875-887.
- [6] J.J. Kalker, The computation of three-dimensional rolling contact with dry friction, Int. J. Num. Meth. Engng. 14 (1979) p. 1293-1307.
- [7] J.J. Kalker, Survey of wheel-rail rolling contact theory, Veh. Syst. Dyn. 5 (1979) p. 317-358.
- [8] A.S.K.S. Tjoeng, J.J. Kalker, User's manual for the program "DUVOROL" in ALGOL-60 & FORTRAN for the computation of three-dimensional contact with dry friction (1980). Available from J.J. Kalker.
- [9] F.C. Carter, On the action of a locomotive driving wheel, Proc. Roy. Soc. A112 (1926) p. 151-157.
- [10] H. Fromm, Calculation of the slipping in the case of rolling deformable bars (German), ZAMM VTN1 (1927).
- [11] J.J. Kalker, A strip theory for rolling with slip and spin, Proc. KNAW B70 (1967) p. 10-62.
- [12] G. Heinrich, K. Desoyer, Rollreibung mit axialem Schub (German), Ing. Arch. 36 (1967) p. 48-72.
- [13] K.L. Johnson, The effect of a tangential force upon the rolling motion of an elastic sphere upon a plane, J. Appl. Mech. 15 (1958) p. 339-346.
- [14] K.L. Johnson, P.J. Vermeulen, Contact of non-spherical bodies transmitting tangential forces, J. Appl. Mech. (1964) p. 338-340.
- [15] D.J. Haines, E. Ollerton, Contact stress distributions on elliptical contact surfaces subjected to radial and tangential forces, Proc. Inst. Mech. Engrs. 179 (1964-5) pt. 3.
- [16] J.J. Kalker, The tangential force transmitted by two elastic bodies rolling over each other with pure creepage, Wear 11 (1968) p. 421-430.