

Journal of
Mechanics of
Materials and Structures

Volume 4, Nº 2

February 2009



mathematical sciences publishers

JOURNAL OF MECHANICS OF MATERIALS AND STRUCTURES

<http://www.jomms.org>

EDITOR-IN-CHIEF Charles R. Steele
ASSOCIATE EDITOR Marie-Louise Steele
Division of Mechanics and Computation
Stanford University
Stanford, CA 94305
USA

BOARD OF EDITORS

D. BIGONI University of Trento, Italy
H. D. BUI École Polytechnique, France
J. P. CARTER University of Sydney, Australia
R. M. CHRISTENSEN Stanford University, U.S.A.
G. M. L. GLADWELL University of Waterloo, Canada
D. H. HODGES Georgia Institute of Technology, U.S.A.
J. HUTCHINSON Harvard University, U.S.A.
C. HWU National Cheng Kung University, R.O. China
IWONA JASIUK University of Illinois at Urbana-Champaign
B. L. KARIHALOO University of Wales, U.K.
Y. Y. KIM Seoul National University, Republic of Korea
Z. MROZ Academy of Science, Poland
D. PAMPLONA Universidade Católica do Rio de Janeiro, Brazil
M. B. RUBIN Technion, Haifa, Israel
Y. SHINDO Tohoku University, Japan
A. N. SHUPIKOV Ukrainian Academy of Sciences, Ukraine
T. TARNAI University Budapest, Hungary
F. Y. M. WAN University of California, Irvine, U.S.A.
P. WRIGGERS Universität Hannover, Germany
W. YANG Tsinghua University, P.R. China
F. ZIEGLER Technische Universität Wien, Austria

PRODUCTION


PAULO NEY DE SOUZA Production Manager
SHEILA NEWBERY Senior Production Editor
SILVIO LEVY Scientific Editor

See inside back cover or <http://www.jomms.org> for submission guidelines.

Regular subscription rate: \$600 a year (print and electronic); \$460 a year (electronic only).

Subscriptions, requests for back issues, and changes of address should be sent to contact@mathscipub.org or to Mathematical Sciences Publishers, 798 Evans Hall, Department of Mathematics, University of California, Berkeley, CA 94720-3840.

©Copyright 2009. Journal of Mechanics of Materials and Structures. All rights reserved.

 mathematical sciences publishers

PREFACE: NEW TRENDS IN THE THERMOMECHANICAL MODELING OF SOLIDS

ANDREI CONSTANTINESCU AND NGUYEN QUOC SON

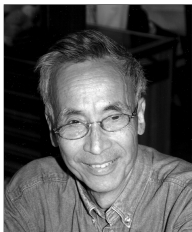
This special issue of JoMMS contains some contributions presented at the *International Conference on Thermo-Mechanical Modeling of Solids* organized by the Laboratoire de Mécanique des Solides (LMS) of the École Polytechnique, Paris, in July 2007.

The LMS was founded in 1961 jointly by Professors Jean Mandel and Pierre Habib as a mechanics research center bringing together researchers from the École Polytechnique, the École des Mines, the École Nationale des Ponts et Chaussées and the Centre National de la Recherche Scientifique (CNRS).

Professor Jean Mandel (1907–1982) graduated from the École Polytechnique in 1927 with highest honors, allowing him to join the prestigious Corps des Mines. His academic career started at the École des Mines of Saint-Étienne in 1932, and continued at the École des Mines of Paris in 1948. From 1951 to 1973 Mandel was a professor of mechanics at the École Polytechnique, where he held the prestigious chair occupied by Lagrange, Cauchy and Poisson. He is well known for his contributions to continuum mechanics, most notably in viscoelasticity, plasticity and geomechanics. His influence extended far beyond his research fields: at the École Polytechnique and in the LMS, several generations of researchers have been shaped by close interaction with Jean Mandel.

The conference was an occasion for the LMS to honor our colleagues Huy Duong Bui, Ky Dang Van, Minh Phuong Luong, Jean Salençon and Joseph Zarka, who started their scientific careers at the LMS in the sixties.

Huy Duong Bui graduated from the École Polytechnique in 1959 and obtained his Doctorat d'État in 1969. His contributions in plasticity and fracture mechanics were crucial for the safety assessment of structures in the nuclear industry. His favorite topic has always been duality. He extended the duality scheme from elasticity to plasticity and further to fracture mechanics [Bui 1978; 2006], and was the first to give the dual expression of the energy release rate in terms of the J-integral. In the field of inverse problems he proposed a series of methods and closed-form solutions based on the reciprocity gap [Bui 1973], which serves as a measure of the loss of duality. Bui is the author of several books and a member



From left to right: Huy Duong Bui, Ky Dang Van, Minh Phuong Luong, Jean Salençon and Joseph Zarka.

of the French Academy of Sciences, the European Academy of Sciences and the French Academy of Technology.

Ky Dang Van graduated from the Ecole Nationale des Ponts et Chaussées and obtained a Doctorat d'État in 1971. His contribution to fatigue theory which is based on a multiscale analysis of the materials at the grain size level and on the concept of shakedown. His what is broadly known as the Dang–Van criterion which is an efficient procedure to estimate the life-time of structures [Fayard et al. 1996; Ballard et al. 1995; Dang Van and Maitournam 1993]. The Dang Van criterion has since been adopted in a variety of companies in the automotive and the aerospace industries. His scientific impact has been recognized by many awards including the Wallenberg prize of the Swedish Academy of Engineering the Alexandre Darracq prize of the French Academy of Science and the Silver Medal of the CNRS.

Minh Phong Luong is a graduate of the École Nationale des Ponts et Chaussées and obtained his Doctorat d'État in 1964. His area of competence is geomechanics and geotechnical engineering, with a special emphasis in earthquake engineering. He made important contributions to granular materials [Evesque et al. 1993] and developed a novel technique for the prediction of lifetime using using infrared thermography [Luong 1995; 1998]. He has several notable results in nondestructive testing of engineering materials. His contributions were patented and used in various industrial applications. He is a recipient of the Henri Courbot prize of the French Academy of Science and several other awards from professional associations.

Jean Salençon graduated with high honors from École Polytechnique in 1959, which allowed him to join the prestigious Corps des Ponts et Chaussées. He obtained his Doctorat d'État in 1969. One of his major contributions is the the reformulation of yield design using the sophisticated mathematical convexity theory. His results were used in the formulation of standards and computational codes broadly used in civil engineering. His other research interests lie in earthquake engineering. Professor and Head of the Department of Mechanics at École Polytechnique and École Nationale des Ponts et Chaussées for years, Salençon is well known his extraordinary classes in different disciplines of mechanics [Salençon 1980; 2001; 2005; 2009]. He is currently the president of the French Academy of Science, and a member of the Istituto Lombardo (Milan).

Joseph Zarka graduated from École Polytechnique in 1962 and obtained his Doctorat d'État in 1968. His main contributions are in multiscale aspects of the polycrystalline plasticity and in the numerical analysis of inelastic structures, including the development of fast algorithms [Zarka et al. 1990]. His other contributions are in optimization techniques and in the domain of nondestructive testing. His current research interests are centered on the optimal design of materials and structures based on automatic learning including applications in areas such as control of fabrication and survey of structures. He was honored by the French Academy of Science with the Fourneyron prize and by the Swedish Academy of Engineering with the Wallenberg prize.

The present generation of LMS researchers is mostly composed of the students and collaborators of Huy Duong Bui, Ky Dang Van, Minh Phuong Luong, Jean Salençon, and Joseph Zarka. Their presence is constantly felt in the LMS because of the high standards of scholarship and intellectual honesty they transmitted to all of us.

References

- [Ballard et al. 1995] P. Ballard, K. D. Van, A. Deperrois, et al., “High cycle fatigue and the finite element analysis”, *Fat. Frac. Eng. Mat. Struct.* **18**:3 (1995), 397–411.
- [Bui 1973] H. D. Bui, *Problèmes inverses en mécanique des matériaux: une introduction*, Eyrolles, Paris, 1973. Translated as *Inverse problems in the mechanics of materials: an introduction*, CRC Press, Boca Raton, FL, 1994; also translated into Japanese (1994) and Chinese (1995).
- [Bui 1978] H. D. Bui, *Mécanique de la rupture fragile*, Masson, Paris, 1978.
- [Bui 2006] H. D. Bui, *Fracture mechanics: inverse problems and solutions*, Springer, Paris, 2006. Translated into Russian (Fizmalit, 2009).
- [Dang Van and Maitournam 1993] K. Dang Van and M. H. Maitournam, “Steady-state flow in classical elastoplasticity: applications to repeated rolling and sliding contact”, *J. Mech. Phys. Solids* **41**:11 (1993), 1691–1710.
- [Evesque et al. 1993] P. Evesque, D. Fargeix, P. Habib, M. P. Luong, and P. Porion, “Pile density is a control parameter of sand avalanches”, *Phys. Rev. E* **47**:4 (1993), 2326–2332.
- [Fayard et al. 1996] J. L. Fayard, K. D. Van, and A. Bignonnet, “Fatigue design criterion for welded structures”, *Fat. Frac. Eng. Mat. Struct.* **19**:6 (1996), 723–729.
- [Luong 1995] M. P. Luong, “Infrared thermographic scanning of fatigue in metals”, *Nucl. Eng. Des.* **158**:2-3 (1995), 363–376.
- [Luong 1998] M. P. Luong, “Fatigue limit evaluation of metals using an infrared thermographic technique”, *Mech. Mater.* **28** (1998), 1–4.
- [Salençon 1980] J. Salençon, *Calcul à la rupture et analyse limite*, Éditions de l’École Nationale des Ponts et Chaussées, Paris, 1980.
- [Salençon 2001] J. Salençon, *Handbook of continuum mechanics*, Springer, Berlin, 2001.
- [Salençon 2005] J. Salençon, *Mécanique des milieux continus* (2 vol. and CD-ROM), Éditions de l’École Polytechnique, Palaiseau, 2005.
- [Salençon 2009] J. Salençon, *Viscoélasticité pour le calcul des structures*, Éditions de l’École Polytechnique, Palaiseau, 2009.
- [Zarka et al. 1990] J. Zarka, J. Frelat, G. Inglebert, and P. Kasmai-Navidi, *A new approach to inelastic analysis of structures*, M. Nijhoff, Dordrecht, 1990.

ANDREI CONSTANTINESCU: constant@lms.polytechnique.fr

Laboratoire de Mécanique des Solides, CNRS - École Polytechnique, 91128 Palaiseau Cedex, France

NGUYEN QUOC SON: son@lms.polytechnique.fr

Laboratoire de Mécanique des Solides, CNRS - École Polytechnique, 91128 Palaiseau Cedex, France

CORRUGATION MODELS AND THE ROARING RAILS ENIGMA: A SIMPLE ANALYTICAL CONTACT MECHANICS MODEL BASED ON A PERTURBATION OF CARTER'S SOLUTION

LUCIANO AFFERRANTE AND MICHELE CIAVARELLA

Corrugation in railways, and especially short pitch corrugation (30–80 mm), is still considered something of an enigma, despite extensive research. Models based on repeated impacts or differential wear, such as Grassie and Johnson's (1985) and Bhaskar et al.'s (1997), seem not to be conclusive, or not to suggest the correct wavelength.

Further models have been suggested, either linear (Frederick, Valdivia, Hempelmann, Vassilly and Vincent) or nonlinear (Mueller), but most suggest a constant frequency mechanism invariably connected to vertical resonances of the system either in the low frequency range (50–100 Hz, the resonance of the vehicle's unsprung mass on the track stiffness referred to here as the "P2 resonance", close to the Hertz contact resonance), or at about 1000 Hz (pinned-pinned resonance, in which the rail vibrates almost as if it were a beam pinned at sleepers), or even higher frequencies still (1700–1800 Hz). The experimental data available, by contrast, do not fit these frequency ranges. The discrepancy is tentatively explained with "contact filtering" and varied traffic ideas, but do not convince completely.

In this paper, we stress the importance of wheel inertia in coupling the oscillations of normal load, with the variations of tangential load and longitudinal creepage. A simple zeroth order perturbation of the classical rolling contact solutions is suggested, which obtains good qualitative agreement with experimental evidence. The model also leads to the recognition that vertical resonances are not crucial in explaining corrugation, as believed in previous models, since we use an extremely simple model of an Euler beam with no elastic support, having no resonances. Important factors for the growth of corrugation are the friction coefficient and the tractive ratio. High longitudinal creepage is needed to promote rapid development, and this can arise from curving, hunting motion or misaligned axles, and is probably exacerbated by high contact conformity, since this increases the fluctuating component of longitudinal creepage due to the movement of the contact point. With discrete supports, we expect a modulation of corrugation wavelength and amplitude, but this requires a separate investigation, not just the inclusion of pinned-pinned resonance.

1. Introduction

Corrugations have been observed and studied for more than a century, and many tentative explanations have been put forward, but none seems convincing for short-pitch rail corrugation ("roaring rails") in the range of 20–80 mm wavelength [Grassie and Kalousek 1993]. This is because a nonproportional increase in corrugation wavelength with increasing trains speed is observed, as in Figure 1 of [Bhaskar et al. 1997a], which displays data from a 1911 BR report, David Harrison's thesis data (1979), and the Vancouver SkyTrain metro system data.

Keywords: short pitch corrugation, wear, rail-wheel contact, rolling contact, friction instabilities.

This makes simple vertical resonance models unsuccessful, and calls for nonlinear effects, or a threshold. For example, a simple mass vertically suspended on a spring-and-damper system — a model used to attempt to explain corrugation of roads, a process known in North America as washboarding (see [Both et al. 2001] and references therein, the more recent [Hoffmann and Misol 2007], and also <http://en.wikipedia.org/wiki/Washboarding> and its external links) predict that instability occurs for all wavelengths larger than a critical value linearly dependent on speed, which depends on the properties of the vehicles and the road surface. However, the wavelength observed in roads seems typically to correspond to frequencies one or two orders of magnitude higher than the lowest vertical resonance of the system.

Similar troubles emerge when attempting to use such simple models for railway corrugation, since the Hertz spring contact resonance model suggested by Carson and Johnson [1971], and observed in twin-disk machines in [Johnson and Gray 1975], suggests a frequency which is highly damped, and hence the normal impact mechanism inducing plastic deformations is observed only at higher wavelengths. Frederick [1987], from BR research, reports that plastic deformation occurs on the peaks where at the short-pitch corrugation frequency corresponds indeed the peak of the normal force, but not on the troughs. There is a need to an alternative explanation for short pitch corrugation, perhaps competing with the plastic deformation mechanism. Indeed, Frederick also reports that high plastic deformation resistant material show corrugation quickly, although the increase of wear resistance slows the rate of formation. Other general observations were given in the well known paper [Grassie and Kalousek 1993], namely that short pitch corrugation is (i) primarily observed on high-speed track, at 100–250 km/h; (ii) mainly on tangent track and on large radius curves with relatively low axle loads; and (iii) with wheel-rail excitation in the frequency range 350–2800 Hz. The first two observations clearly point at a phenomenon which requires sufficient energy to develop and probably sufficient creepage.

A model based on differential wear was then proposed by Grassie and Johnson [1985]. They calculated the frictional energy dissipation in the slip zone of the contact patch as a wheel rolls over a sinusoidally corrugated rail. However, since they assumed a *constant longitudinal creepage* (perhaps under the assumption that the large mass of the vehicle would stop the rotational speed of the wheel to change?), the phase of wear remained very close to the peak of normal force (close to the crests), promoting a mechanism for reduction of corrugation in the frequency range of interest. There was no maximum in the energy dissipation at a particular fixed wavelength independent of vehicle speed, and at typical speeds, the predicted phase did not correspond to a wear maximum in the troughs of an existing corrugation.

This attempt certainly became well known in railways industries and other academic contexts where corrugation was heavily studied, mainly BR, Berlin University, and later Chalmers in Sweden. Assuming that Grassie and Johnson showed longitudinal creepage removed corrugation, the possibility of traction or braking was not included in some later models, and a fortiori the inertia of the wheel and the rotational dynamics of the system were neglected. We can gather this impression from Frederick, head of research at BR, who in a discussion to the paper [Hempelmann and Knothe 1996] asks why not including longitudinal creepage, since it is known to be generally larger than lateral creepage, “although likely to suppress corrugation”.

Frederick [1987] first suggested a perturbation analysis of the nonlinear relationships between normal and tangential forces, creepage and wear, and using complex functions transfer functions, valid for both longitudinal as well as lateral creepage, for defining conditions for which the phase of dissipation would

be such that the component out-of-phase with the initial wave could progress. Unfortunately, many details of the models are very crudely represented. The corrugation is found to be enhanced if there is in the initial rail energy in the frequency range 800–900 Hz which correspond to surface profile positioned near or over a sleeper, attributed to a high vertical impedance and low lateral impedance of the rail. Frederick suggests that it may be possible that extending the analysis to lower frequencies, another high impedance may be found when the sleepers vibrate in antiphase to the rail. The results are however probably affected by the many assumptions in the model.

To explain the discrepancies, Frederick [1987] and Valdivia [1988b; 1988a] have introduced the idea of a “contact filter”, so that in fact the dynamic contact force is mostly amplified in a narrow band of frequency. In particular, for lower speed traffic, the filter makes the pin-pin frequency band “inactive” because it would produce corrugation at less than 1.5 times the contact patch length (i.e. about 20 mm). The 300–400 Hz band is then active (P2 resonance), but that, again, roughly produces corrugations of the observed wavelength. Mixed traffic can reinforce waves of roughly the same pitch. However, this idea of “contact filter” was never really validated.

Significant progress was made by the group of Prof. Knothe in Berlin, and at the Charnec research centre of Chalmers University in Sweden. Groß-Thebing [1993] looked at the transient dynamics effect devising a numerical method using the program CONTACT by Kalker defining generalized complex Carter coefficients to define the tangential load harmonic oscillations due to perturbation of the steady state for harmonic creepage. This code was used in [Hempelmann and Knothe 1996; Hempelmann 1994] but unfortunately, concentrating on lateral creepage alone. Hempelmann also attempts to take into account the discrete nature of support of the rail, although with a spurious Fourier analysis not allowing for parametric resonance; in [Hempelmann and Knothe 1996], the authors attributes the corrugation wavelength to pinned-pinned resonance at about 1000 Hz. This resonant frequency (or more precisely, the slightly higher antiresonance) may indeed have some effect in corrugation, but Hempelmann’s results do not seem to show this, whereas Müller [1998] in his more sophisticated nonlinear version, seems to conclude corrugation to correspond to the low vertical receptance at approximately 1000 Hz, although he also shows that other structural dynamics effects can also dominate the profile development, e.g. the high lateral rail receptance between 1600 and 1800 Hz and the low vertical rail receptance near 300 Hz.

Bhaskar et al. [1997a; 1997b] looked at the Vancouver SkyTrain tramway system where no traction or braking is done at the wheels, yet longitudinal creepage “can arise from curving, hunting motion or misaligned axles”, and indeed their reference case has three comparable components of steady creepage. However, the fluctuating parts of creepage are only caused by conformity (“principal cause of fluctuating longitudinal creepage was found to be the fluctuation in rolling radius due to the movement of the contact point [see equation (27) in [Bhaskar et al. 1997a], which is almost in phase with the angular ripple at most frequencies”), whereas rotational inertia of the wheelset is not included in the model, nor its complete receptance. Incidentally, these authors also attempted to simplify the results of the Groß-Thebing method fitting some results with a spring and dashpot in series at the contact: in particular, the stiffness of the spring is estimated by from a static Mindlin problem, and the dashpot from the zero-th order perturbation of the Carter-like steady state rolling contact solution. However, all this is based on a single fitting of Groß-Thebing’s results with oscillating longitudinal creepage and constant normal load, so while it is likely that the method has sufficient validation for pure oscillation of tangential load alone, the authors do not indicate how to estimate energy dissipation in the general case, and indeed only use the damping term

(no transient effects) in most later calculations — as it will be done in the present paper, which therefore is the equivalent of their paper with no account of conformity of the contact, but with the addition of inertia of the wheel and its dynamics.

It is surprising that the corrugation literature is vast and spans more than a century (Sato et al. [2002] estimate about 1500 papers), yet true comparisons between models and a careful consideration of the critical ingredients have not been attempted in the literature, leaving the impression that only extremely sophisticated models (and accordingly, accurate measurements for many parameters) could reproduce the phenomenon. We shall return to this later, when our simple model will indeed be developed in this paper.

Towards a simple model. A recent paper [Grassie and Edwards 2008] attracted our attention and motivated the present study. The first author is clearly one of the leading authorities in the field of corrugation, being one of the authors of the influential paper [Grassie and Johnson 1985] two decades ago, which however was unsuccessful in that the corrugation phase could not correspond to the trough of the profile. Hence, it was natural to compare the approaches of the two papers to see what was wrong in the early paper. [Grassie and Edwards 2008] distinguishes between corrugation initiated (i) as a result of a varying normal load with essentially constant tangential load — applied traction and braking, steering forces or a combination of the two — (ii) from a varying tangential traction with essentially constant normal load, and (iii) occasionally as a result of a combination of the two. Case (i) is said to be associated to either the resonance of the vehicle’s unsprung mass on the track stiffness referred to here as the “P2 resonance” (typically in the range 50–100 Hz) or the “pinned-pinned resonance” of the rail (typically at much higher frequencies, around 1000 Hz), in which this vibrates almost as if it were a beam pinned at sleepers. This second resonance is called the “dominant wavelength-fixing mechanism for main-line corrugation”. This association did not appear very rigorous, but one immediate reaction on this classification was that the normal load would vary as a result of corrugation or roughness on the railway, and hence its oscillation would indeed be in the range of high frequency. Therefore, it would be difficult to assume that tangential load could be constant also in this frequency unless the inertia of the wheel were very low, giving the first ingredient added in the present paper. However, the extremely simple analytical estimate of differential wear from the energy dissipation in the steady state fluctuating only as a result of the fluctuations of creepage from varying normal load, seem to qualitatively fit some experimental data. This results into a brutally simple equation, but no check is made on the *phase* of the differential energy dissipation, which had been the trouble in the earlier [Grassie and Johnson 1985]. Trusting Grassie’s intuition had pointed in the correct direction, we were therefore motivated to investigate more.

We shall try to consider if it is possible to include only just enough ingredient to explain corrugation, more precisely checking the issues raised by Grassie and Edward within a simple, analytical, perturbation of a Carter solution, in the form including 3D effects in Bhaskar et al. Removing the hopes to include transient effects with the spring+dashpot model, because we believe that the transient effects would anyway require separate and sophisticated treatments, we permit nevertheless both tangential load and creepage to oscillate. Then, we shall consider the phase of the differential energy dissipation, to see if the negative conclusions of Grassie and Johnson apply also to this more general case.

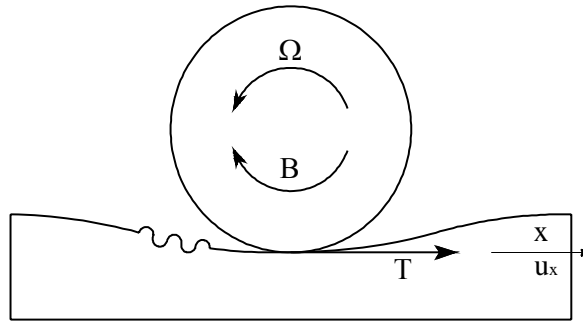


Figure 1. The model under investigation.

2. The model

We shall consider a 2D model with vertical and rotational degrees of freedom (Figure 1). The rail is corrugated in the simple form

$$\Delta \exp(i\omega t),$$

where $\omega = 2\pi V/\lambda$; here V is the velocity of the train and λ the corrugation wavelength. Finally, Δ is the amplitude of corrugation, which is up to about $100 \mu\text{m}$ when full unloading occurs and a linear model is no longer possible, although the noise would become so large that this is prevented in most maintenance strategies. We suppose the normal problem is independent of the tangential problem and results in a normal force

$$P = P_0 + P_1 \exp(i\omega t). \tag{1}$$

We linearize about a given steady state, given by the mean value normal force P_0 , tangential force Q_0 , and creep ratio ζ_0 . We will linearize the contact stiffness in the vertical and tangential direction, and make a perturbation of the steady state solution, to estimate the tangential velocities in the contact area.

The *steady state* creep ratio ζ_0 is defined as the relative velocity of the rail with respect to the wheel

$$\zeta_0 = 1 - \frac{\Omega_0 R}{V} = \frac{\partial u_0}{\partial x}, \tag{2}$$

where we are assuming braking conditions, so $\Omega_0 R < V$ and $\zeta_0 > 0$.

In the transient conditions, the local sliding velocity is a function of both x and t . We can define the transient creepage as the instantaneous rigid body velocity of the rail (V) and that of the wheel ($\Omega(t)R$) — i.e. the rigid body creepage

$$\zeta(t) = 1 - \frac{\Omega(t)R}{V}, \tag{3}$$

The perturbative approach will work well if things don't change too fast relative to the motion of the contact region. In a solution with only longitudinal creepage, perturb (3) about Ω_0, ζ_0 , to have

$$\zeta - \zeta_0 = - \frac{(\Omega - \Omega_0) R}{V} \tag{4}$$

We introduce a 3D solution equivalent to Carter's; it is also described in [Bhaskar et al. 1997a, (5a)], as the longitudinal creep ratio (we assume no lateral creep)

$$\zeta = \zeta_{\max} \left[1 - \left(1 - \frac{Q}{\mu P} \right)^{1/3} \right], \quad (5)$$

where Q is the resulting tangential force which is only in the x-direction, and ζ_{\max} is given by¹

$$\zeta_{\max} = \frac{3\mu}{C_{00}} \left[\frac{16P}{9(1-\nu)^2 R_e^2 G} \right]^{1/3}, \quad (6)$$

with the Kalker's creep coefficient expressed approximately as a function of a/b

$$C_{00} = 2.84 + 1.2 \frac{a}{b} \approx 2.84 + 1.2 \left(\frac{R_1}{R_2} \right)^{2/3}, \quad (7)$$

so that (5) can be rewritten as

$$Q = \mu P \left(1 - \frac{(\zeta_{\max} - \zeta)^3}{\zeta_{\max}^3} \right).$$

It then follows that the dissipation in the steady state is

$$W_0 = V \zeta_0 Q_0 = \mu P_0 V \left(1 - \frac{\Omega_0 R}{V} \right) \left(1 - \frac{(\zeta_{\max} - 1 + \Omega R/V)^3}{\zeta_{\max}^3} \right), \quad (8)$$

where we have substitute for ζ the expression $\zeta = 1 - \Omega R/V$ and ζ_{\max} is a function of P , according to (6). Hence, by differentiation, we obtain the zero-th order perturbation as

$$Q_P = \frac{\partial Q}{\partial P} \Big|_{P_0, \Omega_0} = \frac{2\mu \zeta_0}{\zeta_{\max}} \left(1 - \frac{\zeta_0}{2\zeta_{\max}} \right); \quad (9)$$

$$Q_\Omega = \frac{\partial Q}{\partial \Omega} \Big|_{P_0, \Omega_0} = -\frac{3\mu R P_0}{V \zeta_{\max}} \left(1 - \frac{\zeta_0}{\zeta_{\max}} \right)^2; \quad (10)$$

$$W_P = \frac{\partial W}{\partial P} \Big|_{P_0, \Omega_0} = \frac{2\mu V \zeta_0^2}{\zeta_{\max}} \left(1 - \frac{\zeta_0}{2\zeta_{\max}} \right) = V \zeta_0 Q_P; \quad (11)$$

$$W_\Omega = \frac{\partial W}{\partial \Omega} \Big|_{P_0, \Omega_0} = -\frac{6\mu P_0 R \zeta_0}{\zeta_{\max}} \left[1 - \frac{3\zeta_0}{2\zeta_{\max}} \left(1 - \frac{4\zeta_0}{9\zeta_{\max}} \right) \right]. \quad (12)$$

The fluctuating parts of tangential load and dissipation can therefore be written in the form

$$Q_1 = Q_P P_1 + Q_\Omega \Omega_1 \quad (13)$$

$$W_1 = W_P P_1 + W_\Omega \Omega_1. \quad (14)$$

¹We are defining the Kalker coefficients as positive, for simplicity, and hence change the sign of the creep-load relationships as more commonly found in the literature.

Coupling with the dynamics. The dynamic equilibrium of the wheelset, which we simplify now with no stiffness or damping, gives

$$I_w \frac{d\Omega}{dt} = (Q - Q_\Omega) R \quad (15)$$

where I_w is the inertia of the wheel. Moving to the oscillatory parts therefore, (15) reduces to

$$i\omega I_w \Omega_1 = Q_1 R \quad (16)$$

Substituting Ω_1 from (16) into (13), we have

$$Q_1 = Q_P P_1 + Q_\Omega \frac{Q_1 R}{i\omega I_w}, \quad (17)$$

and collecting Q_1 , we can write the tangential load oscillatory term in the perturbation as a function of the oscillatory term in normal load only,

$$Q_1 = \frac{Q_P}{1 - Q_\Omega R / (i\omega I_w)} P_1. \quad (18)$$

For dissipation, substituting Ω_1 from (16) into (14), we have

$$W_1 = W_1 = W_P P_1 + W_\Omega \frac{R}{i\omega I_w} Q_1 = Q_P \left(V \zeta_0 + \frac{W_\Omega}{i\omega I_w / R - Q_\Omega} \right) P_1 \quad (19)$$

In dimensionless form we can define the dissipated power $\hat{W}_1 = W_1 / (\mu P_0 V \zeta_0)$ as

$$\hat{W}_1 = \frac{2\zeta_0}{\zeta_{\max}} \left(1 - \frac{\zeta_0}{2\zeta_{\max}} \right) \left(1 - \frac{6}{\zeta_{\max}} \cdot \frac{1 - \frac{3\zeta_0}{2\zeta_{\max}} \left(1 - \frac{4\zeta_0}{9\zeta_{\max}} \right)}{i\hat{I}_w \zeta + \frac{3}{\zeta_{\max}} \left(1 - \frac{\zeta_0}{\zeta_{\max}} \right)^2} \right) \frac{P_1}{P_0} \quad (20)$$

where we have introduced the following dimensionless terms

$$\hat{I}_w = \frac{I_w V^2}{2\mu P_0 a_0 R^2}; \quad \zeta = \frac{2\omega a_0}{V} = \frac{4\pi a_0}{\lambda}. \quad (21)$$

where a_0 is the semiwidth of contact in longitudinal direction

$$a_0 = \left(\frac{3(1-\nu)RR_e P_0}{4GR_r} \right)^{1/3} \quad (22)$$

and $R_e = \sqrt{RR_r}$ with R the rolling radius of the wheel and R_r the other relative radius of curvature between the wheel and the rail.

Because microslip and dissipation for a small imposed creep occur towards the rear of the contact, whereas the present calculation associates dissipation with the position of the wheel, the lag of dissipation at a point on the rail is overestimated by about $2\pi a_0 / \lambda = \zeta / 2$. For this reason the phase of W_1 is corrected by introducing a phase lag of $-2\pi a_0 / \lambda$. However when the tractive ratio is large, near to full sliding conditions, dissipation occur nearer the centre of the contact area. In this case the lag of dissipation at a point on the rail is overestimated by about $\pi a_0 / \lambda = \zeta / 4$. In between full stick, and full slip, we assume a linear variation of this correction.

Vertical dynamic model. When the frequency of interest for corrugation is greater than about 500 Hz, the dynamics is dominated by that of the rail described with a simple Euler beam model, not even requiring the inclusion of the elastic supports usually described with Winkler foundations: an infinite beam subjected to a stationary force and whose magnitude oscillates in time at frequency $f = \omega/2\pi$. The mathematical description of the beam model is obtained from the case of load moving with speed V given in the [Appendix](#), although it has certainly been obtained elsewhere previously. For validation and comparison, we shall use the more sophisticated model of Bhaskar et al. [1997a; 1997b], where the rail is continuously supported by uniformly distributed rail pads, sleeper mass and ballast. The effect of discrete sleepers in that work was neglected, because the SkyTrain system has mostly a continuous support.

3. The model

[Figure 2](#) shows the variation with the frequency of the amplitude and phase of the vertical receptance. With solid line we plot the receptance of an infinite beam subjected to a force that moves at speed V , with dashed line the receptance of the model of rail presented in [Bhaskar et al. 1997a]. The comparison is done for $m_{\text{rail}} = 56 \text{ kg/m}$, $V = 27.8 \text{ m/s}$ and $I_{\text{rail}} = 2.35 \times 10^{-5} \text{ m}^4$ as typical values.

For their model, a sharp resonance peak is found at about 100 Hz, corresponding to a vibration mode in which the loaded track vibrates as a whole on the flexibility of the ballast. However, this mode of vibration has been associated with the long wavelength corrugations (greater than 200 mm), which display severe plastic deformation in their troughs. Short pitch corrugation is associated with much higher frequencies (corresponding to wavelength in the range of 20–80 mm), and for high frequencies we note the similarity between the Bhaskar model and the present beam model in terms of the asymptotic value of vertical receptance, for which we get (see [Appendix](#))

$$H_{\text{rail}} = \frac{\exp(-i3\pi/4)}{2\sqrt{2} (m_{\text{rail}}^3 E I_{\text{rail}})^{1/4} \omega^{3/2}}. \quad (23)$$

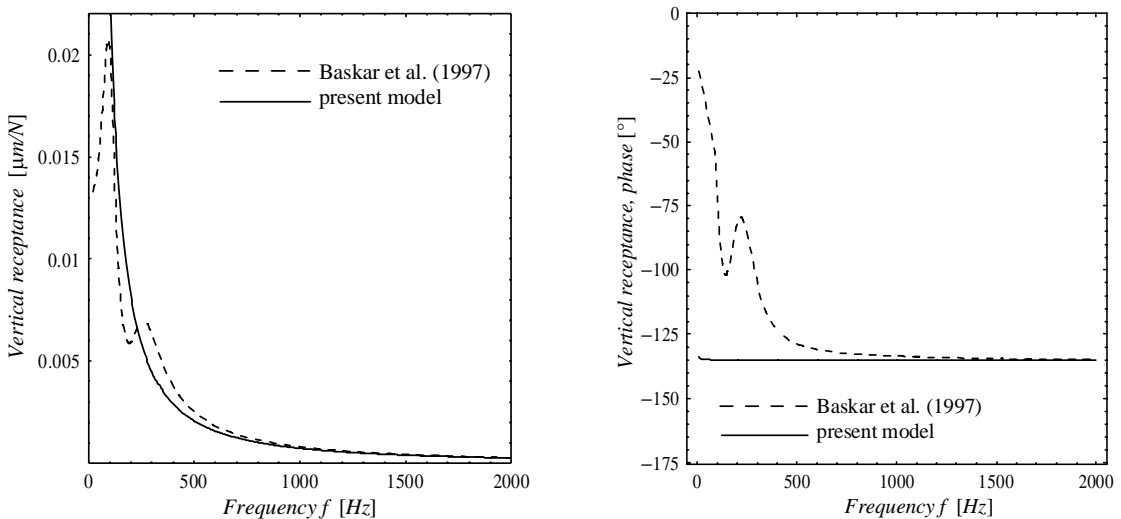


Figure 2. Vertical direct receptance ($E = 207 \text{ GPa}$; $m_{\text{rail}} = 56 \text{ kg/m}$; $m_w = 350 \text{ kg}$; $V = 27.8 \text{ m/s}$; $I_{\text{rail}} = 2.35 \times 10^{-5} \text{ m}^4$).

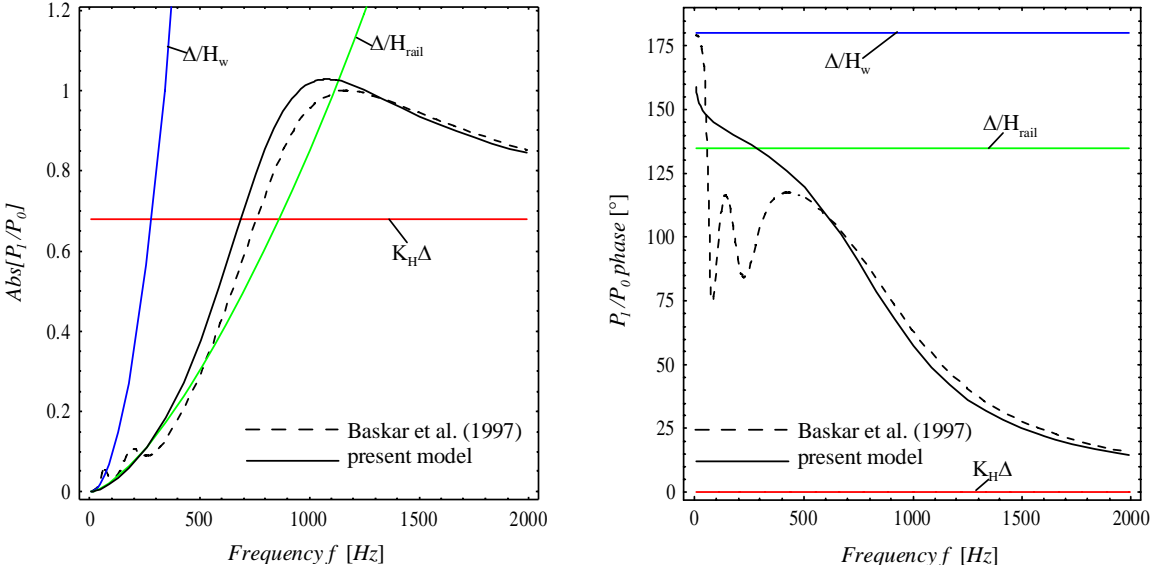


Figure 3. Normal load P_1/P_0 ($\Delta = 3.15 \times 10^{-5}$ m; $E = 207$ GPa; $m_{\text{rail}} = 56$ kg/m; $m_w = 350$ kg; $V = 27.8$ m/s; $I_{\text{rail}} = 2.35 \times 10^{-5}$ m⁴).

Figure 3 shows the amplitude and phase of normal load P_1 . The normal load is evaluated as

$$P_1 = \frac{\Delta}{H_{\text{rail}} + H_w + 1/k_H}, \quad (24)$$

where H_{rail} , H_w , $1/k_H$ are receptance of rail and wheel, respectively, and the inverse of Hertz stiffness. For H_w , we are simplifying the receptance with that of the concentrated mass,

$$H_w = -\frac{1}{m_w \omega^2} \quad (25)$$

and for the Hertz stiffness we use the expression

$$k_H = \left[\frac{6G^2 P_0 R_e}{(1 - \nu^2)} \right]^{1/3}. \quad (26)$$

To distinguish the various contributions, Figure 3 plots each term of the equation above Δ/H_{rail} , Δ/H_{wheel} , and $k_H \Delta$ separately, together with the resulting sum normalized with respect the mean normal load P_0 (the value of $\Delta = 3.15 \times 10^{-2}$ mm was taken such that the maximum value of P_1/P_0 is 1). The dominant effect at low f is the wheel, then the rail at intermediate frequencies. The phase moves from about 135° corresponding to the rail alone (somewhere near 300 Hz) down to 0° for the contact spring alone at very high frequencies.

4. Differential energy dissipation

We start with nominal conditions (normal load $P_0 = 50$ kN, creepage -0.4% , BR rail geometry and wheels, mass of the wheel or wheelset 350 Kg, but inertia reduced using the formula $0.75m_w R^2/2$).

Figure 4 shows contour plots of the real part of dissipation W_1 (ten equally spaced contours from 0 to the smallest value in the figure, which obviously is for the highest speed) with respect to the values of wavelength λ and speed V for tractive ratio $\tau = Q_0/\mu P_0 = 0.1$ and $\tau = 0.95$ (left and right panes, respectively) and for the simplified present model of the rail. Solid circles indicate the highest predicted growth for a given speed. In the plots the lines at constant frequency which gives the best fit of the minimum of real part of W_1 and experimental data points are also shown for reference. Notice for typical values of speed for which short pitch corrugation is observed ($\lambda = 20\text{--}80$ mm) the model gives the highest predicted growth at almost constant frequency, but a large area where growth is possible, and in good agreement with experimental data, except perhaps at the smallest speeds, where there seems to be some additional “filtering”.

Figure 5 shows the same contour plots with the model of the rail in [Bhaskar et al. 1997a]. Similar results to Figure 4 are observed, especially at low tractive ratio (Figure 5, left). However, at low speeds, we have a new possible one or two other lines of possible growth at very low frequencies (about 180 Hz). This regime becomes more important at high tractive ratio (Figure 5, right), showing that the additional “filtering” is due to the low frequency receptance, deviating from the simpler Euler beam behaviour, i.e., due to the supports.

In Figure 6, left, we plot curves of the minima of dissipation (now in the dimensionless form $\hat{W}_1 = W_1/W_0$ which removes the linear dependence on speed) for different tractive ratio τ . The minima are evaluated by considering a window of variation for λ equal to 20–80 mm. It is clear that a constant value

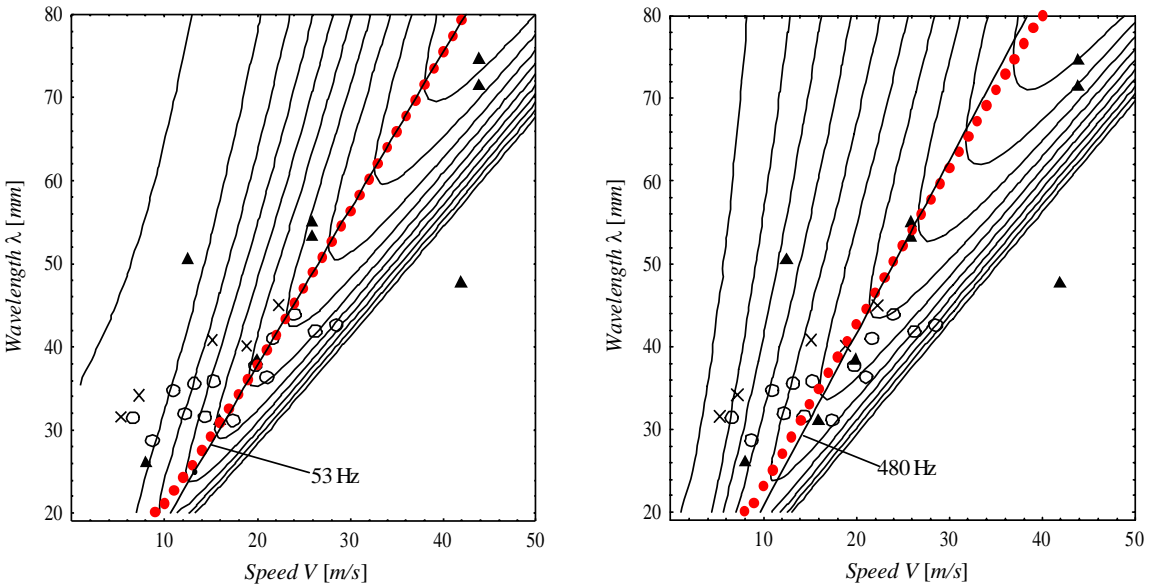


Figure 4. Contour plot of the real part of dissipation with respect to the initial undulation of the corrugation in term of λ and V , for calculations made using the present model (Euler beam for the rail): left, $\tau = 0.1$; right, $\tau = 0.95$. Solid circles indicate the highest predicted growth for a given speed. Data points from Figure 1 of [Bhaskar et al. 1997a] have been superimposed, as follows: open circles, BR survey (1911); triangles, Harrison (1979); crosses, Vancouver Skytrain (1992).

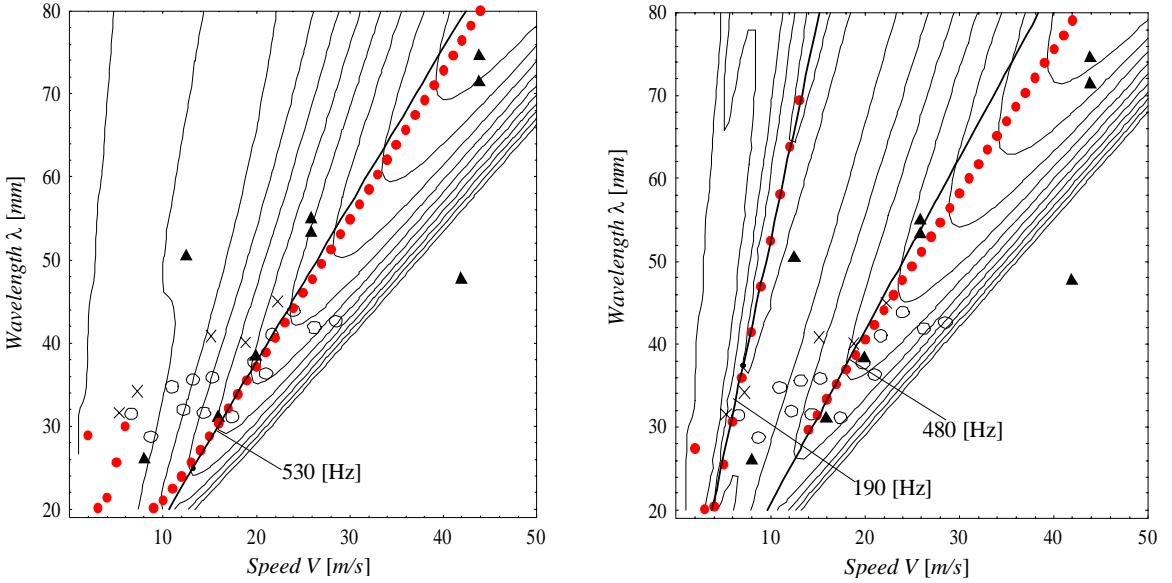


Figure 5. Counterpart of Figure 4 for calculations made using the model of [Bhaskar et al. 1997a]; see that figure for the legend.

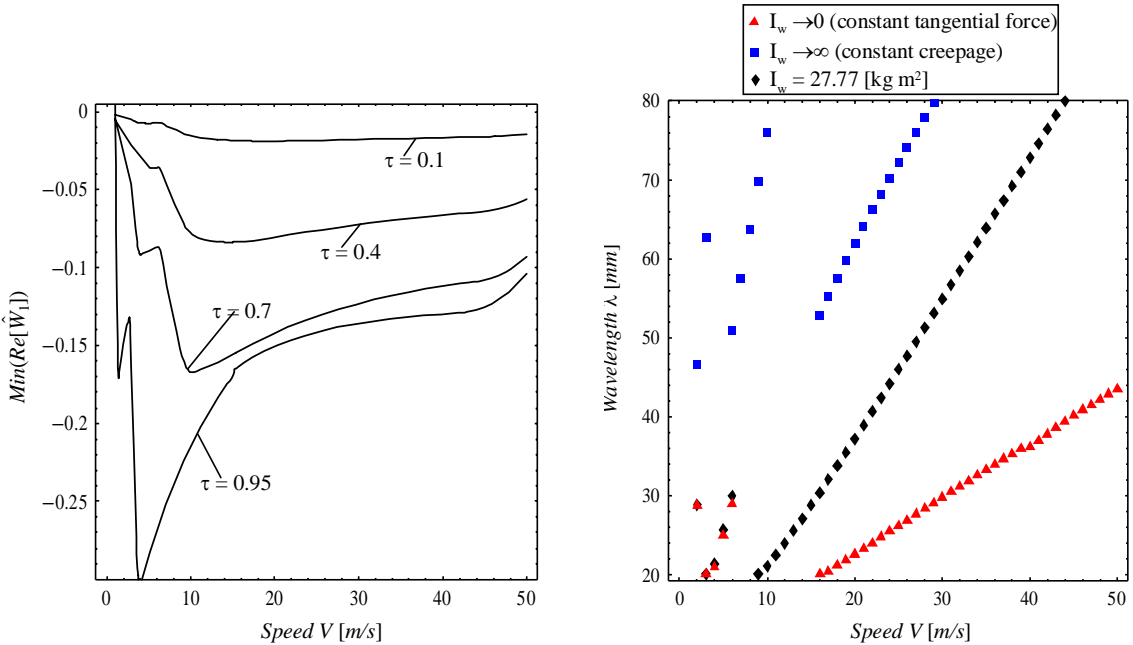


Figure 6. Left: Variation with the speed V of the maximum growth of dissipation, $\min(\hat{W}_1)$ for different tractive ratios. Right: Relation between the wavelength λ of corrugation and the speed V for which we have the maximum dissipation, for different values of inertia of the wheelset ($\tau = 0.1$). For both parts, $\Delta = 3.15 \times 10^{-5}$ m; $E = 207$ GPa; $\nu = 0.3$; $m_{\text{rail}} = 56$ kg/m; $m_w = 350$ kg; $I_{\text{rail}} = 2.35 \times 10^{-5}$ m⁴; $\mu = 0.4$; $P_0 = 50$ kN; $R = 0.46$ m; $R_r = 0.23$ m.

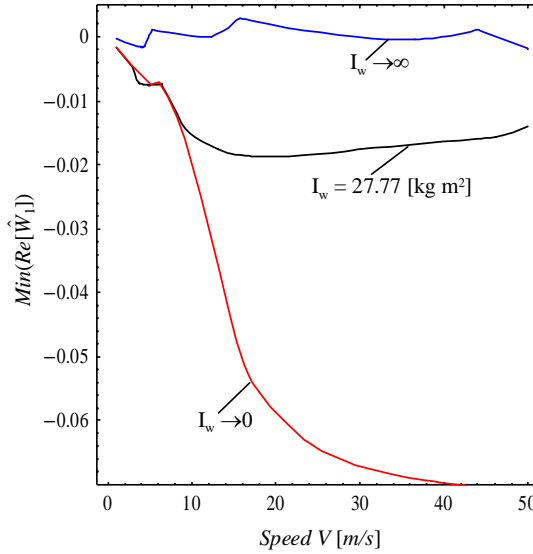


Figure 7. Variation with the speed V of the maximum growth of dissipation, $\min(\hat{W}_1)$, for different dissipation, for different values of inertia of the wheelset. ($\Delta = 3.15 \times 10^{-5}$ m; $E = 207$ GPa; $\nu = 0.3$; $m_{\text{rail}} = 56$ kg/m; $m_w = 350$ kg; $I_{\text{rail}} = 2.35 \times 10^{-5}$ m⁴; $\mu = 0.4$; $\tau = 0.1$; $P_0 = 50$ kN; $R = 0.46$ m; $R_r = 0.23$ m.)

indicates here a linear increase with speed, and the deviations occurring mostly at high tractive ratio, are due to the switching from one frequency line to the other.

Figure 6, right, shows the relation between the wavelength λ of corrugation and speed V for which we have the maximum growth of dissipation for different inertia of the wheelset. In particular, the limit cases of constant tangential force ($I_w \rightarrow 0$) and constant creepage ($I_w \rightarrow \infty$) are plotted and compared with a typical case of inertia ($I_w = 27.77$ kg m²). The limit cases can be considered as bounds for the values of (λ, V) which give growth of corrugation, and also explain why the assumptions of constant tangential load or constant creepage are significantly in error, particularly when considering also Figure 7 with the value of the dimensionless minima of \hat{W}_1 for the same cases of inertia considered in Figure 6. Again we can notice that constant tangential force and creepage bound the maximum growth of corrugation.

In Figure 8, for fixed speed ($V = 30$ m/s), we plot the variation of the minimum of \hat{W}_1 with the tractive ratio and inertia of wheelset, respectively. In the first case we have a monotonic decrease of the minimum of \hat{W}_1 (i.e. an increase of predicted exponential growth of corrugation) with τ . In the latter, an increase of the inertia I_w corresponds to a decrease of dissipation.

5. Discussion

Grassie and Johnson [1982a; 1982b; 1982c; 1982d] improved and extended beam models also to excitation on vertical, longitudinal and lateral excitation, using also careful experiments to compare the results. However, the wheelset dynamics has so many narrow resonances that its role is neglected. In the few cases where it is included, its role however seems crucial, perhaps even larger than what it really is, when considering other effects.

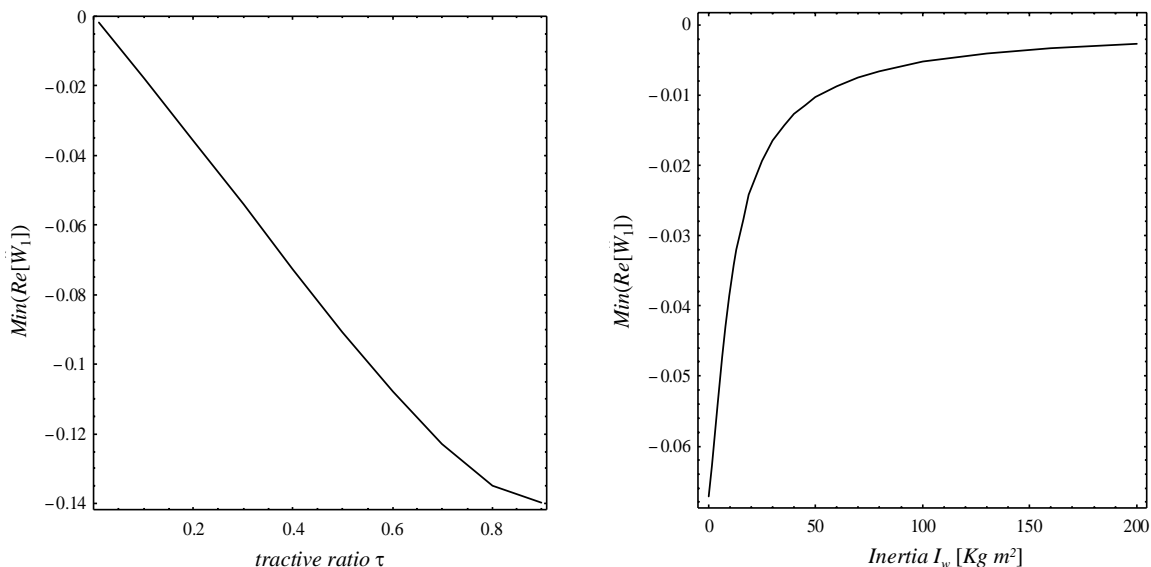


Figure 8. Variation with the tractive ratio τ (left) and with the inertia I_w (right) of the maximum growth of dissipation, $\text{min}(\dot{W}_1)$, for fixed speed $V = 30$ m/s ($\Delta = 3.15 \times 10^{-5}$ m; $E = 207$ GPa; $\nu = 0.3$; $m_{\text{rail}} = 56$ kg/m; $m_w = 350$ kg; $I_{\text{rail}} = 2.35 \times 10^{-5}$ m⁴; $\mu = 0.4$; $P_0 = 50$ kN; $R = 0.46$ m; $R_r = 0.23$ m; for the panel on the right, $\tau = 0.1$.)

For example:

- (a) Tassilly and Vincents [1991a; 1991b] introduce the wheelset behaviour in full curve via its frequency response function, showing predominantly transversal wear on the leading wheelset corresponding to its first bending mode, and longitudinal wear on the rear one, related to the first torsional mode of the wheelset.
- (b) Diana et al. [1998] suggest that the model of Tassilly and Vincents does not justify the case of changing corrugation with a simple change of the pad stiffness, as observed in the Milano subway they analyze. They notice corrugation wavelength is *not* driven by vertical resonances — neither the P2 frequency nor the pinned-pinned frequency (first bending mode) could correspond to the frequency of corrugation, and all the other higher resonant frequencies of the wheelset on the track do *not* change with the change from the stiff to the soft superstructure. Therefore they proposed a mechanism based on the discrete nature of the support and its periodic change of stiffness.
- (c) Elkins et al. [1998], in a study of North American transit railways suggest the second torsional resonance of wheelset also at about 300 Hz as the fixing mechanism.

The closest to our model is in [Diana et al. 1998], the first part of which proposes a simple wheel with concentrated inertia and mass, rolling over a corrugated rail, in turn supported by a structure with periodic change of stiffness. The contact mechanics is simplified with a relationship similar to a Carter solution, although linear in normal load. However, their model is purely integrated in time, and results only for one specific case are illustrated, whereas here we explore fully the behaviour of the system by a perturbation

analysis, permitting to explore the growth of corrugation as a function of all the crucial various parameters. The model shows significant enhancement of corrugation growth with the periodic change of stiffness, which we shall include in later studies. This could be interpreted as inducing parametric resonance on the system, as studied by some authors more in the context of noise than corrugation (see, for example, [Wu and Thompson 2006; Sheng et al. 2006; Wu and Thompson 2004]). However, from these same studies, we expect that corrugation is enhanced in its growth and also modulated in amplitude, but not necessarily will change significantly its wavelength.

Hence, it seems surprising that the corrugation wavelengths may have been frequently associated to the pinned-pinned resonance at about 1000 Hz [Hempelmann and Knothe 1996; Müller 1999; 2000; 1998; Grassie 2005].² Surprisingly, Hempelmann's thesis [1994], in Figure 7.15, has a full comparison of wavelength-speed like our Figure 1, two regimes are predicted around a 400 Hz line and around a 1450 Hz one, the first value not too far from one of our predictions. In [Hempelmann and Knothe 1996], reference to the 1450 Hz regime disappears, and instead there is a suggestion to the highest growth at the pinned-pinned resonance at 1060 Hz. Mueller, in his sophisticated nonlinear model [1999; 1998], seems to confirm Hempelmann's results, adding that other structural dynamics effects can also dominate the profile development, e.g. the high lateral rail receptance between 1600 and 1800 Hz and the low vertical rail receptance near 300 Hz. A detailed comparison is not possible but clearly would be interesting.

6. Conclusions

This note suggests a very simple model permitting a closed form treatment and full exploration of the phenomenon of corrugation with its possibly most crucial factors. The rotational dynamics of the wheel, which surprisingly is missing in many models, couples with the rotational dynamics with the vertical dynamics via the contact mechanics at the interface, leading to a strong effect of the train speed on the possible phase between normal load and local wear.

The results explain why Grassie and Johnson [1985] could not suggest an adequate mechanism, as the predicted phase corresponds to very large vehicle speeds, whereas our model shows that the phase between normal load and differential wear varies largely with speed, and hence should be included to justify the phenomenon.

The corrugation growth is found strongly increasing with the tractive ratio, and hence in stopping or departing conditions, the corrugation growth should be very high. This may explain the second observation of Grassie and Kalousek [1993], since a low normal load may increase the tractive ratio, and this may dominating over the decrease due to the fact that wear is also proportional to the dynamic normal load in turn proportional to the steady state mean value. A large inertia of the wheel may lower the speed for corrugation growth.

Appendix: Dynamic response of the rail

We consider the problem of an infinite beam subjected to a force that moves at speed V and whose magnitude oscillates in time at frequency $f = \omega/2\pi$ (where ω is the pulsation of the oscillation). In the

²Grassie reports that the pinned-pinned resonance is more precisely given as “about 800 Hz in the UK, for a sleeper spacing of 0.75 m and 56 kg/m rail, whereas in much of continental Europe it is more commonly about 1200 Hz because of the closer sleeper spacing (0.6 m) and heavier rail section (60 kg/m).”

steady state the vertical displacement of the beam will be of the form

$$u = U(x - Vt) \exp(i\omega t) \quad (27)$$

The equation on the motion for the beam, except under the load, is

$$\frac{dT}{dx} = -m\ddot{u},$$

where T is the shear force and m is the mass of the beam per unit length. From the linear elastic theory of the beam we also have

$$EI \frac{du^2}{dx^2} = M, \quad \frac{dM}{dx} = T,$$

where EI is the flexural rigidity of the beam and M is the bending moment. It follows that

$$EI \frac{d^4u}{dx^4} + m \frac{d^2u}{dt^2} = 0.$$

Now we introduce the damping in the above equation by adding the term $C\dot{u}$, giving

$$EI \frac{d^4u}{dx^4} + m \frac{d^2u}{dt^2} + C \frac{du}{dt} = 0, \quad (28)$$

where C is a damping coefficient.

By using (27) in (28) and cancelling the common factor $\exp(i\omega t)$, we obtain

$$EIU^{IV} + m(V^2U'' - 2iV\omega U' - \omega^2U) + C(-VU' + i\omega U) = 0.$$

If we introduce the following dimensionless variable ξ defined as

$$\frac{V\xi}{\omega} = x - Vt,$$

in which case

$$U'(x - Vt) = \frac{\omega}{V}U'(\xi)$$

etc., we obtain

$$KU^{IV}(\xi) + U''(\xi) - 2iU'(\xi) - U(\xi) + \hat{C}(-U'(\xi) + iU(\xi)) = 0,$$

where

$$K = \frac{EI\omega^2}{mV^4} = \frac{4\pi^2EI}{mV^2\lambda^2} \quad (29)$$

is a dimensionless flexural rigidity and

$$\hat{C} = \frac{C}{m\omega} \ll 1.$$

Noting that the parameter $\lambda = 2\pi V/\omega$ is the wavelength of the implied corrugation.

The problem is therefore governed by the dimensionless parameters K and \hat{C} and the phase shift between the excitation force and implied corrugation will be a function only of these parameters.

Solution. Equation (28) has the solution

$$U(\xi) = A \exp(s\xi),$$

where A is an arbitrary complex constant and s satisfies the fourth degree polynomial equation

$$Ks^4 + s^2 - 2\iota s - 1 + \hat{C}(-s + \iota) = 0$$

To avoid the complex coefficient in this equation, we can write

$$s = \iota\sigma,$$

giving

$$K\sigma^4 - \sigma^2 + 2\sigma - 1 + \hat{C}(-\iota\sigma + \iota) = 0. \tag{30}$$

For $\hat{C} = 0$ the solutions of (30) are

$$\sigma_{1,2} = \frac{1 \pm \sqrt{1-4k}}{2k}; \quad \sigma_{3,4} = \frac{-1 \pm \sqrt{1+4k}}{2k},$$

where $k^2 = K$.

We notice that all the roots are real if $k < 1/4$ and hence $K < 1/16$. For $K > 1/16$, $\sigma_{1,2}$ are complex and $\sigma_{3,4}$ are real. This in turn implies that roots for s are either all pure imaginary, or else two are complex and two are imaginary. The imaginary roots correspond to waves of constant amplitude, so it is difficult to know what to do about conditions at infinity. Physically, this corresponds to the fact that the undamped beam can support free vibrations and if the wavelength of these are suitably chosen, they can be combined to give a travelling wave of any given speed. We can get around it by imposing a small amount of damping.

Our concern is with the two real roots $\sigma_{3,4}$ and we anticipate that for small \hat{C} we can modify them in the form

$$\sigma_3 = \frac{-1 + \sqrt{1+4k}}{2k} + \hat{C}g_3(k); \quad \sigma_4 = \frac{-1 - \sqrt{1+4k}}{2k} + \hat{C}g_4(k).$$

Differentiating (30) with respect to \hat{C} , we have

$$(4K\sigma^3 - 2\sigma + 2 - \hat{C}\iota) \frac{d\sigma}{d\hat{C}} - \iota(\sigma - 1) = 0,$$

and hence

$$\frac{d\sigma}{d\hat{C}} = \frac{\iota(\sigma - 1)}{4K\sigma^3 - 2\sigma + 2 - \hat{C}\iota}.$$

Substituting $\sigma = \sigma_3$ and $\hat{C} = 0$, we obtain:

$$g_3(k) = \frac{d\sigma}{d\hat{C}} = -\frac{\iota}{2\sqrt{1+4k}}.$$

which shows that $\iota g_3(k)$ is always positive, hence the real part of the modified root s_3 is positive.

Next substitute $\sigma = \sigma_4$ so that

$$g_4(k) = \frac{d\sigma}{d\hat{C}} = \frac{\iota}{2\sqrt{1+4k}}.$$

Apart from the i factor, this is positive for all k and hence $ig_4(k) < 0$ and the real part of the modified root s_4 is negative.

Boundary conditions. We construct a solution for the region $\xi > 0$ using the roots with negative real part, so

$$U(\xi) = A_1 \exp(s_1 \xi) + A_4 \exp(s_4 \xi).$$

For $\xi < 0$, we use the roots with positive real part giving

$$U(\xi) = A_2 \exp(s_2 \xi) + A_3 \exp(s_3 \xi).$$

We require continuity in u and its first two derivatives (the second derivatives to achieve continuity of bending moment M), giving

$$A_1 + A_4 = A_2 + A_3, \quad A_1 s_1 + A_4 s_4 = A_2 s_2 + A_3 s_3, \quad A_1 s_1^2 + A_4 s_4^2 = A_2 s_2^2 + A_3 s_3^2. \quad (31)$$

The final equation comes from the requirement that the shear force has a discontinuity of magnitude $F_0 \exp(i\omega t)$ at the origin. In other words

$$T(0^+) - T(0^-) = F_0 \exp(i\omega t),$$

which implies

$$\frac{EI\omega^3}{V^3} (U'''(0^+) - U'''(0^-)) = F_0. \quad (32)$$

We therefore have

$$A_1 s_1^3 + A_4 s_4^3 - A_2 s_2^3 - A_3 s_3^3 = \frac{F_0 V^3}{EI\omega^3}.$$

The receptance will be determined by the complex ratio between F_0 and the displacement under the load which is

$$u_0 = A_1 + A_4. \quad (33)$$

It will be a function of K (and hence of the frequency f). Notice that this is physically meaningful only above the value of K that makes $g_3(K) > 0$. Further the result will be only slightly affected by g_3 and g_4 as long as these are small, and hence we can simplify the problem by setting these to zero.

Large K limit. Consider the case where $K \gg 1$, which is equivalent to $V \rightarrow 0$ in view of (29). In other words, in this limit, we should recover the solution for an oscillating load that is stationary at the origin. The roots then tend to the values

$$s_1 = -\frac{1}{\sqrt{k}}; \quad s_2 = \frac{1}{\sqrt{k}}; \quad s_3 = \frac{i}{\sqrt{k}}; \quad s_4 = -\frac{i}{\sqrt{k}},$$

and from the boundary conditions (31), (32) and (33), we then get

$$u_0 = -\frac{F_0(1+i)}{4(m^3 EI)^{1/4} \omega^{3/2}}.$$

Thus,

$$u(0) = u_0 \exp(i\omega t) = -\frac{F_0(1+i)}{4(m^3 EI)^{1/4} \omega^{3/2}} \exp(i\omega t),$$

which we can rewrite as

$$u(0) = \frac{F_0}{2\sqrt{2} (m^3 EI)^{1/4} \omega^{3/2}} \exp(i(\omega t - 3\pi/4)).$$

Thus, the displacement lags from the force by $3\pi/4 = 135^\circ$.

The vertical receptance can hence be written as

$$H_{\text{rail}} = \frac{u(0)}{F_0} = \frac{\exp(-i3\pi/4)}{2\sqrt{2} (m^3 EI)^{1/4} \omega^{3/2}}.$$

References

- [Bhaskar et al. 1997a] A. Bhaskar, K. L. Johnson, G. D. Wood, and J. Woodhouse, “Wheel-rail dynamics with closely conformal contact, 1: Dynamic modelling and stability analysis”, *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit* **211**:1 (1997), 11–26.
- [Bhaskar et al. 1997b] A. Bhaskar, K. L. Johnson, and J. Woodhouse, “Wheel-rail dynamics with closely conformal contact, 2: Forced response, results and conclusions”, *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit* **211**:1 (1997), 27–40.
- [Both et al. 2001] J. A. Both, D. C. Hong, and D. A. Kurtze, “Corrugation of roads”, *Physica A* **301**:1–4 (2001), 545–559.
- [Carson and Johnson 1971] R. M. Carson and K. L. Johnson, “Surface corrugations spontaneously generated in a rolling contact disc machine”, *Wear* **17**:1 (1971), 59–72.
- [Diana et al. 1998] G. Diana, F. Cheli, S. Bruni, and A. Collina, “Experimental and numerical investigation on subway short pitch corrugation”, *Veh. Syst. Dyn.* **29**:S1 (1998), 234–245.
- [Elkins et al. 1998] J. A. Elkins, S. Grassie, and S. Handal, “Rail corrugation mitigation in transit”, *TCRP Research Results Digest* **26** (1998), 1–33.
- [Frederick 1987] C. O. Frederick, “A rail corrugation theory”, pp. 181–211 in *Contact mechanics and wear of rail/wheel systems, II: Proceedings of the International Symposium* (Kingston, RI, 1986), edited by G. M. L. Gladwell et al., University of Waterloo Press, Waterloo, ON, 1987.
- [Grassie 2005] S. L. Grassie, “Rail corrugation: advances in measurement, understanding and treatment”, *Wear* **258**:7–8 (2005), 1224–1234.
- [Grassie and Edwards 2008] S. L. Grassie and J. W. Edwards, “Development of corrugation as a result of varying normal load”, *Wear* **265**:9–10 (2008), 1150–1155.
- [Grassie and Johnson 1985] S. L. Grassie and K. L. Johnson, “Periodic microslip between a rolling wheel and a corrugated rail”, *Wear* **101**:4 (1985), 291–309.
- [Grassie and Kalousek 1993] S. L. Grassie and J. Kalousek, “Rail corrugation: characteristics, causes and treatments”, *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit* **207**:1 (1993), 57–68.
- [Grassie et al. 1982a] S. L. Grassie, R. W. Gregory, D. Harrison, and K. L. Johnson, “The dynamic response of railway track to high frequency vertical excitation”, *J. Mech. Eng. Sci.* **24**:2 (1982), 77–90.
- [Grassie et al. 1982b] S. L. Grassie, R. W. Gregory, and K. L. Johnson, “The dynamic response of railway track to high frequency lateral excitation”, *J. Mech. Eng. Sci.* **24**:2 (1982), 91–96.
- [Grassie et al. 1982c] S. L. Grassie, R. W. Gregory, and K. L. Johnson, “The dynamic response of railway track to high frequency longitudinal excitation”, *J. Mech. Eng. Sci.* **24**:2 (1982), 97–102.
- [Grassie et al. 1982d] S. L. Grassie, R. W. Gregory, and K. L. Johnson, “The behaviour of railway wheelsets and track at high frequencies of excitation”, *J. Mech. Eng. Sci.* **24**:2 (1982), 103–111.
- [Groß-Thebing 1993] A. Groß-Thebing, *Lineare Modellierung des instationären Rollkontaktes von Rad und Schiene*, Fortschritt-Berichte VDI, 12: Verkehrstechnik/Fahrzeugtechnik **199**, VDI, Düsseldorf, 1993.
- [Hempelmann 1994] K. Hempelmann, *Short pitch corrugation on railway rails: a linear model for prediction*, Fortschritt-Berichte VDI, 12: Verkehrstechnik/Fahrzeugtechnik **231**, VDI, Düsseldorf, 1994.

- [Hempelmann and Knothe 1996] K. Hempelmann and K. Knothe, “An extended linear model for the prediction of short pitch corrugation”, *Wear* **191**:1–2 (1996), 161–169.
- [Hoffmann and Misol 2007] N. P. Hoffmann and M. Misol, “On the role of varying normal load and of randomly distributed relative velocities in the wavelength selection process of wear-pattern generation”, *Int. J. Solids Struct.* **44**:25–26 (2007), 8718–8734.
- [Johnson and Gray 1975] K. L. Johnson and G. G. Gray, “Development of corrugations on surfaces in rolling contact”, *Proc. Inst. Mech. Eng.* **189** (1975), 567–580.
- [Müller 1998] S. Müller, *Linearized wheel-rail dynamics: stability and corrugation*, Fortschritt-Berichte VDI, 12: Verkehrstechnik/Fahrzeugtechnik **369**, VDI, Düsseldorf, 1998.
- [Müller 1999] S. Müller, “A linear wheel-track model to predict instability and short pitch corrugation”, *J. Sound Vib.* **227**:5 (1999), 899–913.
- [Müller 2000] S. Müller, “A linear wheel-rail model to investigate stability and corrugation on straight track”, *Wear* **243**:1–2 (2000), 122–132.
- [Sato et al. 2002] Y. Sato, A. Matsumoto, and K. Knothe, “Review on rail corrugation studies”, *Wear* **253**:1–2 (2002), 130–139.
- [Sheng et al. 2006] X. Sheng, D. J. Thompson, C. J. C. Jones, G. Xie, S. D. Iwnicki, P. Allen, and S. S. Hsu, “Simulations of roughness initiation and growth on railway rails”, *J. Sound Vib.* **293**:3–5 (2006), 819–829.
- [Tassilly and Vincent 1991a] E. Tassilly and N. Vincent, “A linear model for the corrugation of rails”, *J. Sound Vib.* **150**:1 (1991), 25–45.
- [Tassilly and Vincent 1991b] E. Tassilly and N. Vincent, “Rail corrugations: analytical model and field tests”, *Wear* **144**:1–2 (1991), 163–178.
- [Valdivia 1988a] A. Valdivia, *Die Wechselwirkung zwischen hochfrequenter Rad-Schiene-Dynamik und ungleichförmigem Schienenverschleiss: ein lineares Modell*, Fortschritt-Berichte VDI, 12: Verkehrstechnik/Fahrzeugtechnik **93**, VDI, Düsseldorf, 1988.
- [Valdivia 1988b] A. Valdivia, “A linear dynamic wear model to explain the initiating mechanism of corrugation”, pp. 493–496 in *The dynamics of vehicles on roads and on tracks: Proceedings of 10th IAVSD Symposium* (Prague, 1987), edited by M. Apetaur, Swets & Zeitlinger, Amsterdam, 1988.
- [Wu and Thompson 2004] T. X. Wu and D. J. Thompson, “On the parametric excitation of the wheel/track system”, *J. Sound Vib.* **278**:4–5 (2004), 725–747.
- [Wu and Thompson 2006] T. X. Wu and D. J. Thompson, “On the rolling noise generation due to wheel/track parametric excitation”, *J. Sound Vib.* **293**:3–5 (2006), 566–574.

Received 14 Dec 2007. Revised 28 Aug 2008. Accepted 18 Sep 2008.

LUCIANO AFFERRANTE: luciano@poliba.it
CEMEC-PoliBA, Via le Japigia 182, Politecnico di Bari, 70125 Bari, Italy

MICHELE CIAVARELLA: mciava@poliba.it
CEMEC-PoliBA, Via le Japigia 182, Politecnico di Bari, 70125 Bari, Italy

STABILITY AND MEMORY EFFECTS IN A HOMOGENIZED MODEL GOVERNING THE ELECTRICAL CONDUCTION IN BIOLOGICAL TISSUES

MICOL AMAR, DANIELE ANDREUCCI, PAOLO BISEGNA AND ROBERTO GIANNI

We present a macroscopic model of electrical conduction in biological tissues. This model is derived via a homogenization limit by a microscopic formulation based on Maxwell's equations, taking into account the periodic geometry of the microstructure. We also study the asymptotic behavior of the model for large times. Our results imply that periodic boundary data lead to an asymptotically periodic solution. The model is relevant to applications like electric impedance tomography.

1. Introduction

In this paper we deal with a model of electrical conduction in composite media and, specifically, conduction in biological tissues. The classical governing equation is

$$-\operatorname{div}(\kappa \nabla u_t + \sigma \nabla u) = 0, \quad (1-1)$$

which is derived from the Maxwell equations in the quasistationary approximation (see for example, [Novožilov and Yappa 1978]). Here, u is the electrical potential and κ , σ are the permittivity and the conductivity of the material, respectively. The geometry of the composite media we have in mind is a periodic array of the unit cell depicted in Figure 1. More precisely, we look at a phase E_1^η which models the cell cytosol, coated by a shell Γ^η which models the cell membrane, included in a phase E_2^η which models the extracellular fluid [Foster and Schwan 1989]. In particular, the permittivity κ in E_1^η and E_2^η is lower, and the conductivity σ is higher, than in Γ^η . The diameter of the cell is of the order of tens of micrometers, while the width of the membrane is of the order of ten nanometers. This suggests that the thin shell Γ^η could be preferably modeled as a two dimensional interface Γ , in order to get a simpler model and, possibly, a better understanding of the effect of the geometric features of the microscopic structure. This simpler model can be obtained from Equation (1-1) via a concentration-of-capacity procedure [Amar et al. 2006], leading to Problem (2-1)–(2-6), below. In particular, Equation (2-3) takes into account the conductive/capacitive behavior of the concentrated membrane. As shown in (2-3), the electric potential jumps across the interface Γ , and its jump satisfies a dynamical condition (roughly speaking, in the form of a hyperbolic differential equation on the interface itself).

Our model is designed to investigate the response of biological tissues to the injection of electrical currents in the radio frequency range, that is, the Maxwell–Wagner interfacial polarization effect [Foster and Schwan 1989; Bisegna et al. 2001], at higher frequencies than those considered in [Amar et al. 2003; 2004b; 2005; 2006; 2008]. This effect is relevant to clinical applications like electric impedance tomography and body composition [De Lorenzo et al. 1997; Bronzino 1999].

Keywords: asymptotic decay, stability, homogenization, memory effects, electrical conduction, biological tissues.

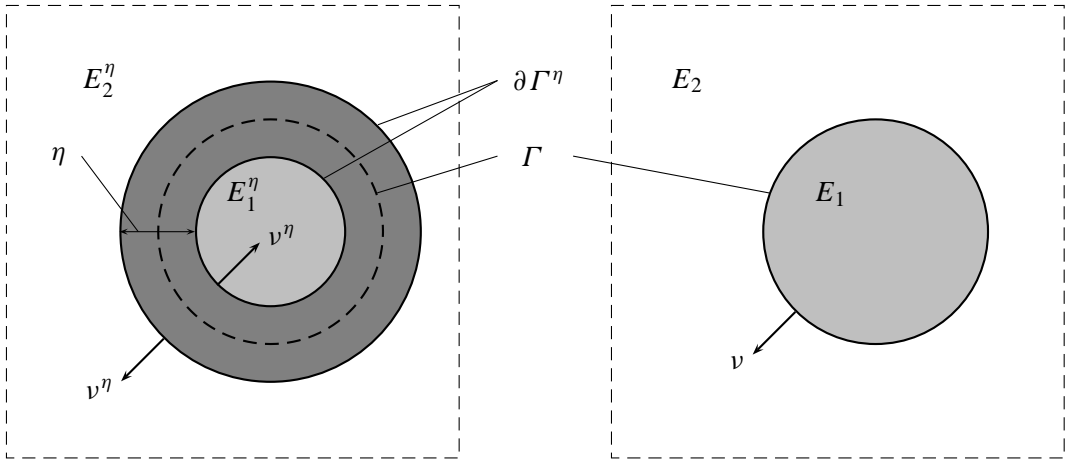


Figure 1. The periodic cell Y . Left: before concentration; Γ^η is the dark gray region, and $E^\eta = E_1^\eta \cup E_2^\eta$ is the union of the light gray and white regions. Right: after concentration; Γ^η shrinks to Γ as $\eta \rightarrow 0$.

Problem (2-1)–(2-6) contains a small parameter ε , coinciding with the period of the microstructure. The typical structure of the periodic array we have in mind is given in Figure 2. Some applications deal with measurements of the electric potential at the macroscopic (body) scale: this suggests that it would be advantageous to investigate the homogenization limit of Problem (2-1)–(2-6) when we let $\varepsilon \rightarrow 0$. Extensive surveys on this topic are, for example, in [Bensoussan et al. 1978; Sánchez-Palencia

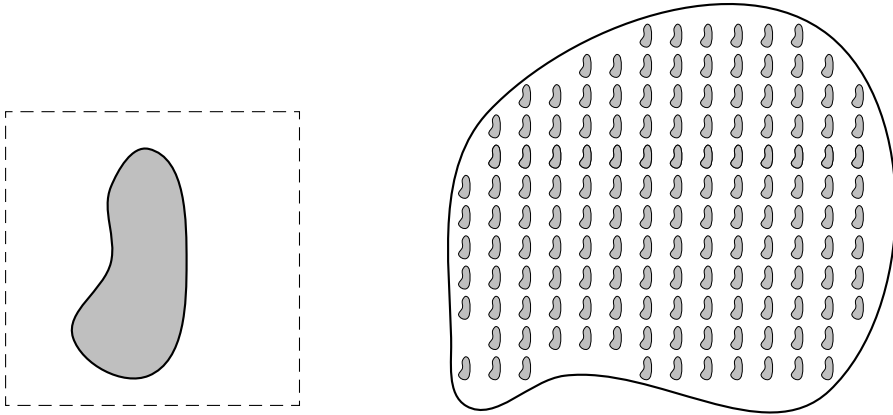


Figure 2. Left: an example of admissible periodic unit cell $Y = E_1 \cup E_2 \cup \Gamma$ in \mathbb{R}^2 . Here E_1 is the light gray region and Γ is its boundary. The remaining part of Y (the white region) is E_2 . Right: the corresponding domain $\Omega = \Omega_1^\varepsilon \cup \Omega_2^\varepsilon \cup \Gamma^\varepsilon$. Here Ω_1^ε is the light gray region and Γ^ε is its boundary. The remaining part of Ω (the white region) is Ω_2^ε .

1980; Lions 1981; Attouch 1984; Bakhvalov and Panasenko 1989; Oleinik et al. 1992; Sánchez-Hubert and Sánchez-Palencia 1992; Jikov et al. 1994; Braides and Defranceschi 1998; Cioranescu and Donato 1999]. It turns out that the partial differential equation obtained in the limit is nonstandard; see (3-39) below. Indeed, it is an equation exhibiting memory effects, that is, it contains explicitly the history of the unknown and hence is markedly different from the Laplace equation presently used as a standard in the bioelectrical impedance literature [Bronzino 1999].

Our model can be compared to some papers where homogenization theory is applied to linear stationary elliptic problems involving imperfect interfaces arising in fields like elasticity [Lene and Leguillon 1981] or heat conduction [Lipton 1998]. See also [Donato et al. 2007; Sánchez-Palencia 1980], where hyperbolic problems with interfaces are considered in the framework of elastodynamics and electro-dynamics.

In view of the applications, it is also of interest to study the time evolution of the homogenized potential (see Section 2). In particular, it is of interest to show that time-harmonic boundary data elicits a time-harmonic solution for large times. In this regard, following the same reasoning as that presented in [Amar et al. 2008], it is enough to prove that the solution u_0 of (3-39) exponentially decays to zero as time increases, provided that a zero Dirichlet boundary condition is assigned (see Theorem 2.1 and Corollary 2.2).

From a mathematical point of view, the asymptotic behavior of evolutive equations with memory is a classical problem [Fichera 1979; Slemrod 1981; Fabrizio and Morro 1988; Lazzari and Vuk 1992], currently drawing much interest in the literature [Lazzari and Nibbi 2002; Giorgi et al. 2001; 2005; Medjden and Tatar 2005; Appleby et al. 2006]. We note that the exponential decay of the memory kernel, in general, does not imply the existence of bounded solutions, as shown by a counterexample presented in Section 5 (see also, [Fichera 1979; Fabrizio and Morro 1988]).

We finally note that our methods could be easily applied to study the homogenization problem and the time-asymptotic behavior of Kelvin–Voigt viscoelastic composites with coated inclusions.

2. Position of the problem and main results

We look at the homogenization limit ($\varepsilon \rightarrow 0$) of the following problem for $u_\varepsilon(x, t)$:

$$-\operatorname{div}(\kappa \nabla u_{\varepsilon t} + \sigma \nabla u_\varepsilon) = 0 \quad \text{in } (\Omega_1^\varepsilon \cup \Omega_2^\varepsilon) \times (0, +\infty); \quad (2-1)$$

$$[(\kappa \nabla u_{\varepsilon t} + \sigma \nabla u_\varepsilon) \cdot \nu] = 0 \quad \text{on } \Gamma^\varepsilon \times (0, +\infty); \quad (2-2)$$

$$(\alpha/\varepsilon)\partial[u_\varepsilon]/\partial t + (\beta/\varepsilon)[u_\varepsilon] = ((\kappa \nabla u_{\varepsilon t} + \sigma \nabla u_\varepsilon) \cdot \nu)^{(2)} \quad \text{on } \Gamma^\varepsilon \times (0, +\infty); \quad (2-3)$$

$$u_\varepsilon(x, t) = 0 \quad \text{on } \partial\Omega \times (0, +\infty); \quad (2-4)$$

$$\nabla u_\varepsilon(x, 0) = \mathbf{G}_\varepsilon(x) \quad \text{in } \Omega_1^\varepsilon \cup \Omega_2^\varepsilon; \quad (2-5)$$

$$[u_\varepsilon](x, 0) = S_\varepsilon(x) \quad \text{on } \Gamma^\varepsilon. \quad (2-6)$$

The operators div and ∇ act with respect to the space variable x ; $\Omega = \Omega_1^\varepsilon \cup \Omega_2^\varepsilon \cup \Gamma^\varepsilon$, where Ω_1^ε and Ω_2^ε are two disjoint open subsets of Ω , and $\Gamma^\varepsilon = \partial\Omega_1^\varepsilon \cap \Omega = \partial\Omega_2^\varepsilon \cap \Omega$; ν is the normal unit vector pointing into Ω_2^ε ; the typical geometry we have in mind is depicted in Figure 2. We refer to Section 2 for a precise definition of the structure of Ω_1^ε , Ω_2^ε , Γ^ε .

Moreover, we assume that

$$\begin{aligned} \alpha > 0; \quad \beta \geq 0; \quad \kappa = \kappa_1 > 0, \quad \sigma = \sigma_1 > 0 \quad \text{in } \Omega_1^\varepsilon; \\ \kappa = \kappa_2 > 0, \quad \sigma = \sigma_2 > 0 \quad \text{in } \Omega_2^\varepsilon; \end{aligned} \quad (2-7)$$

where $\kappa_1, \kappa_2, \sigma_1, \sigma_2, \alpha$, and β are constants. From a physical point of view, Γ^ε represents the cell membranes having capacitance α/ε and conductance β/ε per unit area, whereas Ω_1^ε (respectively, Ω_2^ε) is the intracellular (respectively, extracellular) space, having permittivity κ_1 (respectively, κ_2) and conductivity σ_1 (respectively, σ_2).

Since u_ε is not, in general, continuous across Γ^ε , we have set

$$u_\varepsilon^{(2)} := \text{trace of } u_{\varepsilon|\Omega_2^\varepsilon} \text{ on } \Gamma^\varepsilon, \quad u_\varepsilon^{(1)} := \text{trace of } u_{\varepsilon|\Omega_1^\varepsilon} \text{ on } \Gamma^\varepsilon, \quad \text{and} \quad [u_\varepsilon] := u_\varepsilon^{(2)} - u_\varepsilon^{(1)}.$$

A similar convention is employed for the current flux density across the membrane $(\kappa \nabla u_{\varepsilon t} + \sigma \nabla u_\varepsilon) \cdot \nu$.

We assume that the restrictions of \mathbf{G}_ε to Ω_1^ε and Ω_2^ε are gradients of scalar fields, and that \mathbf{G}_ε strongly converges in L^2 . Moreover, we assume that $S_\varepsilon \in H^1(\Omega)$, and that $S_\varepsilon/\varepsilon$ strongly converges in L^2 . These assumptions are introduced in order to rule out the appearance of an initial layer (see [Amar et al. 2009]). Further assumptions on \mathbf{G}_ε and S_ε are introduced in the next paragraph.

Geometry. Following [Amar et al. 2004b], we introduce a periodic open subset E of \mathbb{R}^N , so that $E + z = E$ for all $z \in \mathbf{Z}^N$. For all $\varepsilon > 0$ we define $\Omega_1^\varepsilon = \Omega \cap \varepsilon E$, $\Omega_2^\varepsilon = \Omega \setminus \varepsilon \bar{E}$, $\Gamma^\varepsilon = \Omega \cap \partial(\varepsilon E)$. We assume that Ω , E have a regular boundary, say of class C^∞ for the sake of simplicity. We also employ the notation $Y = (0, 1)^N$, $E_1 = E \cap Y$, $E_2 = Y \setminus \bar{E}$, $\Gamma = \partial E \cap \bar{Y}$. We stipulate that E_1 is a connected smooth subset of Y such that $\text{dist}(\bar{E}_1, \partial Y) > 0$. Some generalizations may be possible, but we do not dwell on this point here. Finally, we assume that $\text{dist}(\Gamma^\varepsilon, \partial\Omega) > \gamma\varepsilon$ for some constant $\gamma > 0$ independent of ε , by dropping the inclusions contained in the cells $\varepsilon(Y + z)$, $z \in \mathbf{Z}^N$ which intersect $\partial\Omega$ (see Figure 2). For later usage, we introduce the set

$$\mathbf{Z}_\varepsilon^N := \{z \in \mathbf{Z}^N : \varepsilon(Y + z) \subseteq \Omega\}. \quad (2-8)$$

Energy estimate. Multiply (2-1) by u_ε and integrate by parts. Using (2-2)–(2-6), we arrive, for all $t > 0$, to the energy estimate

$$\begin{aligned} \int_\Omega \frac{\kappa}{2} |\nabla u_\varepsilon(x, t)|^2 dx + \int_0^t \int_\Omega \sigma |\nabla u_\varepsilon(x, \tau)|^2 dx d\tau + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [u_\varepsilon(x, t)]^2 d\sigma \\ + \frac{\beta}{\varepsilon} \int_0^t \int_{\Gamma^\varepsilon} [u_\varepsilon(x, \tau)]^2 d\sigma d\tau = \int_\Omega \frac{\kappa}{2} |\mathbf{G}_\varepsilon(x)|^2 dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} S_\varepsilon^2(x) d\sigma. \end{aligned} \quad (2-9)$$

We assume that

$$\int_\Omega \frac{\kappa}{2} |\mathbf{G}_\varepsilon(x)|^2 dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} S_\varepsilon^2(x) d\sigma < \gamma, \quad (2-10)$$

for a constant γ independent of ε . In fact (2-9), coupled with the Poincaré's inequality (Lemma 4.1), is a main tool in the rigorous proof of convergence of u_ε to its limit. In particular, up to a subsequence, u_ε converges weakly in $L^2(\Omega \times (0, \bar{T}))$ as $\varepsilon \rightarrow 0$ to a limit u_0 , for every $\bar{T} > 0$. The equation satisfied by u_0 will be formally derived via a homogenization procedure in Section 3.

Exponential decay.

Theorem 2.1. *Let $\Omega_1^\varepsilon, \Omega_2^\varepsilon, \Gamma^\varepsilon$ be as before. Assume that (2-7) holds, and that the initial data \mathbf{G}_ε are gradients of scalar fields and together with S_ε satisfy (2-10). Let u_ε be the solution of (2-1)–(2-6). Then*

$$\|u_\varepsilon(\cdot, t)\|_{L^2(\Omega)} \leq C(\varepsilon + e^{-\lambda t}) \quad \text{a.e. in } (0, +\infty), \tag{2-11}$$

where C and λ are positive constants independent of ε . Moreover, if $\beta > 0$, or else if S_ε has null mean average over each connected component of Γ^ε , it follows that

$$\|u_\varepsilon(\cdot, t)\|_{L^2(\Omega)} \leq C e^{-\lambda t} \quad \text{a.e. in } (0, +\infty). \tag{2-12}$$

This result easily yields the following exponential time-decay estimate for the limit u_0 under homogeneous Dirichlet boundary data:

Corollary 2.2. *Under the assumptions of Theorem 2.1, if $u_\varepsilon \rightarrow u_0$ weakly in $L^2(\Omega \times (0, \bar{T}))$ for every $\bar{T} > 0$, then*

$$\|u_0(\cdot, t)\|_{L^2(\Omega)} \leq C e^{-\lambda t} \quad \text{a.e. in } (0, +\infty). \tag{2-13}$$

3. Formal homogenization

To establish the notation, we summarize here some well known asymptotic expansions needed in the two-scale method (see, for example, [Bensoussan et al. 1978], [Sánchez-Palencia 1980]). Introduce the microscopic variables $y \in Y, y = x/\varepsilon$, assuming

$$u_\varepsilon = u_\varepsilon(x, y, t) = u_0(x, y, t) + \varepsilon u_1(x, y, t) + \varepsilon^2 u_2(x, y, t) + \dots \tag{3-1}$$

Note that u_0, u_1, u_2 are periodic in y , and u_1, u_2 are assumed to have zero integral average over Y . Recalling that

$$\operatorname{div} = \frac{1}{\varepsilon} \operatorname{div}_y + \operatorname{div}_x, \quad \nabla = \frac{1}{\varepsilon} \nabla_y + \nabla_x, \tag{3-2}$$

we compute, for example,

$$\nabla u_\varepsilon = \frac{1}{\varepsilon} \nabla_y u_0 + (\nabla_x u_0 + \nabla_y u_1) + \varepsilon(\nabla_y u_2 + \nabla_x u_1) + \dots \tag{3-3}$$

We also stipulate

$$\mathbf{G}_\varepsilon = \mathbf{G}_\varepsilon(x, y) = \mathbf{G}_0(x, y) + \varepsilon \mathbf{G}_1(x, y) + \varepsilon^2 \mathbf{G}_2(x, y) + \dots, \tag{3-4}$$

$$S_\varepsilon = S_\varepsilon(x, y) = S_0(x, y) + \varepsilon S_1(x, y) + \varepsilon^2 S_2(x, y) + \dots, \tag{3-5}$$

where the restrictions of $\mathbf{G}_0(x, \cdot), \mathbf{G}_1(x, \cdot), \dots$ to E_1 and E_2 are the gradients of scalar fields. According to Equation (2-10), recalling that $|\Gamma^\varepsilon|_{N-1} \sim 1/\varepsilon$, we assume $S_0 \equiv 0$ in (3-5). Moreover, according to the assumption on the strong convergence of \mathbf{G}_ε and $S_\varepsilon/\varepsilon$, the functions $\mathbf{G}_0(x, y)$ and $S_1(x, y)$ do not depend on y , that is $\mathbf{G}_0(x, y) = \mathbf{G}_0(x)$ and $S_1(x, y) = S_1(x)$.

For the sake of brevity, we introduce the operator

$$\mathfrak{D} := \kappa \frac{\partial}{\partial t} + \sigma. \tag{3-6}$$

Applying (3-2)–(3-3) to Problem (2-1)–(2-6), one readily sees, by matching corresponding powers of ε , that u_0 solves,

$$-\mathcal{D}\Delta_y u_0 = 0 \quad \text{in } (E_1 \cup E_2) \times (0, +\infty); \quad (3-7)$$

$$[\mathcal{D}\nabla_y u_0 \cdot \nu] = 0 \quad \text{on } \Gamma \times (0, +\infty); \quad (3-8)$$

$$\alpha \frac{\partial[u_0]}{\partial t} + \beta[u_0] = (\mathcal{D}\nabla_y u_0 \cdot \nu)^{(2)} \quad \text{on } \Gamma \times (0, +\infty); \quad (3-9)$$

$$\nabla_y u_0|_{t=0} = 0 \quad \text{on } E_1 \cup E_2; \quad (3-10)$$

$$[u_0]|_{t=0} = 0 \quad \text{on } \Gamma. \quad (3-11)$$

Reasoning as in Section 2, we obtain an energy estimate for (3-7)–(3-11), which implies that $[u_0] = 0$ for all times, and

$$u_0 = u_0(x, t).$$

Next, we find for u_1 that

$$-\mathcal{D}\Delta_y u_1 = 0 \quad \text{in } (E_1 \cup E_2) \times (0, +\infty); \quad (3-12)$$

$$[\mathcal{D}(\nabla_y u_1 + \nabla_x u_0) \cdot \nu] = 0 \quad \text{on } \Gamma \times (0, +\infty); \quad (3-13)$$

$$\alpha \frac{\partial[u_1]}{\partial t} + \beta[u_1] = (\mathcal{D}(\nabla_y u_1 + \nabla_x u_0) \cdot \nu)^{(2)} \quad \text{on } \Gamma \times (0, +\infty); \quad (3-14)$$

$$\nabla_y u_1|_{t=0} + \nabla_x u_0|_{t=0} = \mathbf{G}_0 \quad \text{on } E_1 \cup E_2; \quad (3-15)$$

$$[u_1]|_{t=0} = S_1 \quad \text{on } \Gamma. \quad (3-16)$$

Since both u_0 and \mathbf{G}_0 do not depend on y , Equation (3-15) implies $\nabla_y u_1|_{t=0} = 0$ on $E_1 \cup E_2$.

In order to represent u_1 in a suitable way, let $\mathbf{g} \in L^2(E_1 \cup E_2)$ and $s \in L^2(\Gamma)$ be assigned such that the restrictions of \mathbf{g} to E_1 and E_2 are gradients of scalar fields, and consider the problem

$$-\mathcal{D}\Delta_y v = 0, \quad \text{in } (E_1 \cup E_2) \times (0, +\infty); \quad (3-17)$$

$$[\mathcal{D}\nabla_y v \cdot \nu] = 0, \quad \text{on } \Gamma \times (0, +\infty); \quad (3-18)$$

$$\alpha \frac{\partial[v]}{\partial t} + \beta[v] = (\mathcal{D}\nabla_y v \cdot \nu)^{(2)} \quad \text{on } \Gamma \times (0, +\infty). \quad (3-19)$$

$$\nabla_y v|_{t=0} = \mathbf{g} \quad \text{on } E_1 \cup E_2; \quad (3-20)$$

$$[v]|_{t=0} = s \quad \text{on } \Gamma. \quad (3-21)$$

where v is a periodic function in Y , such that $\int_Y v(y, t) dy = 0$. Define the transform \mathcal{T} by

$$\mathcal{T}(\mathbf{g}, s)(y, t) = v(y, t), \quad y \in Y, t > 0.$$

Then, introduce the cell functions $\chi^0 : Y \rightarrow \mathbb{R}^N$ and $\chi^1 : Y \times (0, +\infty) \rightarrow \mathbb{R}^N$, whose components χ_h^0 and $\chi_h^1(\cdot, t)$, $h = 1, \dots, N$, are required to be periodic functions with vanishing integral averages over

Y for $t \geq 0$. The function χ_h^0 of the components of χ^0 satisfies

$$-\kappa \Delta_y \chi_h^0 = 0 \quad \text{in } E_1 \cup E_2; \quad (3-22)$$

$$[\kappa(\nabla_y \chi_h^0 - \mathbf{e}_h) \cdot \nu] = 0 \quad \text{on } \Gamma; \quad (3-23)$$

$$\alpha[\chi_h^0] = (\kappa(\nabla_y \chi_h^0 - \mathbf{e}_h) \cdot \nu)^{(2)} \quad \text{on } \Gamma. \quad (3-24)$$

The initial value $\chi_h^1(\cdot, 0)$ of the components of χ^1 satisfies

$$-\kappa \Delta_y \chi_h^1(\cdot, 0) - \sigma \Delta_y \chi_h^0 = 0, \quad \text{in } E_1 \cup E_2; \quad (3-25)$$

$$[(\kappa \nabla_y \chi_h^1(\cdot, 0) + \sigma(\nabla_y \chi_h^0 - \mathbf{e}_h)) \cdot \nu] = 0 \quad \text{on } \Gamma; \quad (3-26)$$

$$((\kappa \nabla_y \chi_h^1(\cdot, 0) + \sigma(\nabla_y \chi_h^0 - \mathbf{e}_h)) \cdot \nu)^{(2)} = \alpha[\chi_h^1(\cdot, 0)] + \beta[\chi_h^0] \quad \text{on } \Gamma. \quad (3-27)$$

Finally, χ_h^1 is defined for $t > 0$ by

$$\chi_h^1 = \mathcal{F} \left(\nabla_y \chi_h^1(\cdot, 0), [\chi_h^1(\cdot, 0)] \right). \quad (3-28)$$

Straightforward calculations show that u_1 may be written in the form

$$\begin{aligned} u_1(x, y, t) = & -\chi^0(y) \cdot \nabla_x u_0(x, t) - \int_0^t \chi^1(y, t - \tau) \cdot \nabla_x u_0(x, \tau) \, d\tau \\ & + \mathcal{F}(\nabla_y(\chi^0 \cdot \mathbf{G}_0(x)), S_1(x) + [\chi^0] \cdot \mathbf{G}_0(x))(y, t), \end{aligned} \quad (3-29)$$

so that

$$\begin{aligned} \mathcal{D}u_1(x, y, t) = & -\kappa \chi^0(y) \cdot \nabla_x u_{0t}(x, t) - (\kappa \chi^1(y, 0) + \sigma \chi^0(y)) \cdot \nabla_x u_0(x, t) \\ & - \int_0^t (\mathcal{D}\chi^1)(y, t - \tau) \cdot \nabla_x u_0(x, \tau) \, d\tau \\ & + \mathcal{D}\mathcal{F}(\nabla_y(\chi^0 \cdot \mathbf{G}_0(x)), S_1(x) + [\chi^0] \cdot \mathbf{G}_0(x))(y, t). \end{aligned} \quad (3-30)$$

Next we find for u_2 that

$$-\mathcal{D} \left(\Delta_y u_2 + 2 \frac{\partial^2 u_1}{\partial x_j \partial y_j} + \Delta_x u_0 \right) = 0, \quad \text{in } (E_1 \cup E_2) \times (0, +\infty); \quad (3-31)$$

$$[\mathcal{D}(\nabla_y u_2 + \nabla_x u_1) \cdot \nu] = 0 \quad \text{on } \Gamma \times (0, +\infty); \quad (3-32)$$

$$(\mathcal{D}(\nabla_y u_2 + \nabla_x u_1) \cdot \nu)^{(2)} = \alpha \frac{\partial[u_2]}{\partial t} + \beta[u_2] \quad \text{on } \Gamma \times (0, +\infty). \quad (3-33)$$

$$\nabla_y u_2|_{t=0} + \nabla_x u_1|_{t=0} = \mathbf{G}_1 \quad \text{on } E_1 \cup E_2; \quad (3-34)$$

$$[u_2]|_{t=0} = S_2 \quad \text{on } \Gamma. \quad (3-35)$$

Let us find the solvability conditions for this problem. Integrating by parts the partial differential equations (3-31) solved by u_2 , both in E_1 and in E_2 , adding the two contributions, and using (3-32), we get

$$\left[\int_{E_1} + \int_{E_2} \right] \mathcal{D} \left\{ \Delta_x u_0(x, t) + 2 \frac{\partial^2 u_1}{\partial x_j \partial y_j} \right\} dy = - \int_{\Gamma} [\mathcal{D} \nabla_x u_1 \cdot \nu] d\sigma. \quad (3-36)$$

Thus, we obtain

$$\left(\kappa_0 \frac{\partial}{\partial t} + \sigma_0 \right) \Delta_x u_0 = 2 \int_{\Gamma} [\mathcal{D} \nabla_x u_1 \cdot \nu] d\sigma - \int_{\Gamma} [\mathcal{D} \nabla_x u_1 \cdot \nu] d\sigma = \int_{\Gamma} [\mathcal{D} \nabla_x u_1 \cdot \nu] d\sigma, \quad (3-37)$$

where

$$\kappa_0 = \kappa_1 |E_1| + \kappa_2 |E_2|; \quad \sigma_0 = \sigma_1 |E_1| + \sigma_2 |E_2|. \quad (3-38)$$

Then, we substitute the representation (3-29) into Equation (3-37) and, after simple algebra, obtain the homogenized equation for u_0 in $\Omega \times (0, +\infty)$ as

$$- \operatorname{div} \left(K \nabla_x u_{0t} + A \nabla_x u_0 + \int_0^t B(t-\tau) \nabla_x u_0(\cdot, \tau) d\tau - \mathcal{F} \right) = 0, \quad (3-39)$$

where the matrices K , A , $B(t)$, and the vector $\mathcal{F}(x, t)$ are defined as follows:

$$K = \kappa_0 I + \int_{\Gamma} \nu \otimes [\kappa \chi^0(y)] d\sigma, \quad (3-40)$$

$$A = \sigma_0 I + \int_{\Gamma} \nu \otimes [\kappa \chi^1(y, 0) + \sigma \chi^0(y)] d\sigma, \quad (3-41)$$

$$B(t) = \int_{\Gamma} \nu \otimes [(\mathcal{D} \chi^1)(y, t)] d\sigma, \quad (3-42)$$

$$\mathcal{F}(x, t) = \int_{\Gamma} [\mathcal{D} \mathcal{T}(\nabla_y(\chi^0 \cdot \mathbf{G}_0(x)), S_1(x) + [\chi^0] \cdot \mathbf{G}_0(x))(y, t)] \nu d\sigma. \quad (3-43)$$

Equation (3-39) is complemented with the initial condition

$$\nabla_x u_0|_{t=0} = \mathbf{G}_0, \quad \text{on } \Omega. \quad (3-44)$$

Finally, integrating Equation (3-39) over time, changing the order in the double integral that results, and using (3-44), we obtain also the following formulation

$$- \operatorname{div} \left(K \nabla_x u_0 + \int_0^t \left(A + \int_0^{t-s} B(\tau) d\tau \right) \nabla_x u_0(\cdot, s) ds - K \mathbf{G}_0 - \int_0^t \mathcal{F}(\cdot, \tau) d\tau \right) = 0, \quad (3-45)$$

which shows that the homogenized equation has exactly the form of an equation with memory of the type derived in [Amar et al. 2003; 2004b] and studied in [Amar et al. 2004a].

4. Time-exponential asymptotic decay: proof of Theorem 2.1

The case $\beta > 0$ is quite simple. We introduce the space

$$H_\varepsilon^1(\Omega) := \{v \in L^2(\Omega) : v|_{\Omega_i^\varepsilon} \in H^1(\Omega_i^\varepsilon), i = 1, 2; v = 0 \text{ on } \partial\Omega\}. \quad (4-1)$$

It turns out that, for all $v \in H_\varepsilon^1(\Omega)$,

$$\int_\Omega \sigma |\nabla v|^2 \, dx + \frac{\beta}{\varepsilon} \int_{\Gamma^\varepsilon} [v]^2 \, d\sigma \geq \lambda \left(\int_\Omega \frac{\kappa}{2} |\nabla v|^2 \, dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [v]^2 \, d\sigma \right), \tag{4-2}$$

for $\lambda = \min\{2\sigma_1/\kappa_1, 2\sigma_2/\kappa_2, 2\beta/\alpha\}$. Taking $v = u_\varepsilon(\cdot, t)$ in the previous estimate and using equations (2-9), (2-10), and the differential version of Gronwall’s Lemma, we obtain

$$\int_\Omega \frac{\kappa}{2} |\nabla u_\varepsilon(\cdot, t)|^2 \, dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [u_\varepsilon(\cdot, t)]^2 \, d\sigma \leq \gamma e^{-\lambda t}, \quad \text{a.e. in } (0, +\infty), \tag{4-3}$$

and (2-12) follows from Poincaré’s inequality (Lemma 4.1).

Now we consider the case $\beta = 0$. We introduce the space $\tilde{H}^{1/2}(\Gamma^\varepsilon) \subset H^{1/2}(\Gamma^\varepsilon)$ of the functions which have a null average over each connected component of Γ^ε , that is, on $\varepsilon(\Gamma + z)$, for each z belonging to the set \mathbf{Z}_ε^N defined in (2-8). We decompose the initial datum $S_\varepsilon(x)$ in (2-6) as $S_\varepsilon(x) = \bar{S}_\varepsilon(x) + \tilde{S}_\varepsilon(x)$, where

$$\begin{aligned} \bar{S}_\varepsilon(x) &= \int_{\varepsilon(\Gamma+z)} S_\varepsilon \, d\sigma =: C_{\varepsilon z} \quad \text{on each } \varepsilon(\Gamma + z), \, z \in \mathbf{Z}_\varepsilon^N; \\ \tilde{S}_\varepsilon(x) &\in \tilde{H}^{1/2}(\Gamma^\varepsilon), \end{aligned} \tag{4-4}$$

and the initial datum $\mathbf{G}_\varepsilon(x)$ in (2-5) as $\mathbf{G}_\varepsilon(x) = \bar{\mathbf{G}}_\varepsilon(x) + \tilde{\mathbf{G}}_\varepsilon(x)$, where $\bar{\mathbf{G}}_\varepsilon(x) = 0$ and $\tilde{\mathbf{G}}_\varepsilon(x) = \mathbf{G}_\varepsilon(x)$. Accordingly, the solution u_ε to Problem (2-1)–(2-6) is decomposed as $\bar{u}_\varepsilon + \tilde{u}_\varepsilon$. Clearly,

$$\bar{u}_\varepsilon(x, t) = \begin{cases} 0 & \text{for } (x, t) \in \Omega_2^\varepsilon \times (0, +\infty), \\ -C_{\varepsilon z} & \text{for } (x, t) \in (\varepsilon(E_1 + z)) \times (0, +\infty), \, z \in \mathbf{Z}_\varepsilon^N. \end{cases} \tag{4-5}$$

Using the previous equation, we compute

$$\int_\Omega |\bar{u}_\varepsilon|^2 \, dx = \sum_{z \in \mathbf{Z}_\varepsilon^N} \int_{\varepsilon(E_1+z)} |\bar{u}_\varepsilon|^2 \, dx = \varepsilon^N |E_1| \sum_{z \in \mathbf{Z}_\varepsilon^N} \left| \int_{\varepsilon(\Gamma+z)} S_\varepsilon \, d\sigma \right|^2. \tag{4-6}$$

On the other hand, by Hölder’s inequality, we estimate

$$\sum_{z \in \mathbf{Z}_\varepsilon^N} \left| \int_{\varepsilon(\Gamma+z)} S_\varepsilon \, d\sigma \right|^2 \leq \frac{\gamma}{\varepsilon^{N-1}} \int_{\Gamma^\varepsilon} S_\varepsilon^2 \, d\sigma. \tag{4-7}$$

Hence, as a consequence of (2-10), it follows that

$$\|\bar{u}_\varepsilon(\cdot, t)\|_{L^2(\Omega)} \leq C \varepsilon, \tag{4-8}$$

where C is a constant independent of ε .

In order to obtain an estimate for \tilde{u}_ε , we introduce the space

$$\tilde{H}_\varepsilon^1(\Omega) := \{v \in H_\varepsilon^1(\Omega) : [v] \in \tilde{H}^{1/2}(\Gamma^\varepsilon)\}, \tag{4-9}$$

and, using [Lemma 4.2](#) and [Remark 4.3](#) below, we compute, for every $v \in \tilde{H}_\varepsilon^1(\Omega)$,

$$\int_{\Omega} \sigma |\nabla v|^2 dx \geq \sum_{z \in \mathbf{Z}_\varepsilon^N} \int_{\varepsilon(Y+z)} \sigma |\nabla v|^2 dx \geq \frac{\alpha \tilde{\lambda}}{\varepsilon} \sum_{z \in \mathbf{Z}_\varepsilon^N} \int_{\varepsilon(\Gamma+z)} [v]^2 d\sigma = \frac{\alpha \tilde{\lambda}}{\varepsilon} \int_{\Gamma^\varepsilon} [v]^2 d\sigma, \quad (4-10)$$

where $\tilde{\lambda}$ is defined in [\(4-15\)](#) and is independent of ε . Hence,

$$\int_{\Omega} \sigma |\nabla v|^2 dx \geq \lambda \left(\int_{\Omega} \frac{\kappa}{2} |\nabla v|^2 dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [v]^2 d\sigma \right), \quad (4-11)$$

for $\lambda = (\max\{\kappa_1/(2\sigma_1), \kappa_2/(2\sigma_2)\} + 1/(2\tilde{\lambda}))^{-1}$.

On the other hand, reasoning as in [Section 2](#) and using [\(4-4\)](#) and [\(2-10\)](#), we get that \tilde{u}_ε satisfies the energy estimate

$$\int_{\Omega} \frac{\kappa}{2} |\nabla \tilde{u}_\varepsilon(x, t)|^2 dx + \int_0^t \int_{\Omega} \sigma |\nabla \tilde{u}_\varepsilon(x, \tau)|^2 dx d\tau + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [\tilde{u}_\varepsilon(x, t)]^2 d\sigma < \gamma. \quad (4-12)$$

Hence, by using [\(4-11\)](#) written for $\tilde{u}_\varepsilon(\cdot, t)$ and the differential version of Gronwall's Lemma, we obtain

$$\int_{\Omega} \frac{\kappa}{2} |\nabla \tilde{u}_\varepsilon(\cdot, t)|^2 dx + \frac{\alpha}{2\varepsilon} \int_{\Gamma^\varepsilon} [\tilde{u}_\varepsilon(\cdot, t)]^2 d\sigma \leq \gamma e^{-\lambda t}, \quad \text{a.e. in } (0, +\infty), \quad (4-13)$$

and [\(2-11\)](#) follows from Poincaré's inequality ([Lemma 4.1](#)) and [\(4-8\)](#).

Lemma 4.1 (Poincaré's inequality [[Hummel 2000](#); [Amar et al. 2004b](#)]). *Let v belong to the space $H_\varepsilon^1(\Omega)$ introduced in [Equation \(4-1\)](#). Then,*

$$\int_{\Omega} v^2 dx \leq C \left\{ \int_{\Omega} |\nabla v|^2 dx + \varepsilon^{-1} \int_{\Gamma^\varepsilon} [v]^2 d\sigma \right\}. \quad (4-14)$$

Here C depends only on Ω and E .

Lemma 4.2 [[Amar et al. 2008](#)]. *Set $\tilde{H}^1(Y) := \{v \in L^2(Y) : v|_{E_i} \in H^1(E_i), i = 1, 2, [v] \in \tilde{H}^{1/2}(\Gamma)\}$, where $\tilde{H}^{1/2}(\Gamma)$ is comprised of the functions of $H^{1/2}(\Gamma)$ with null integral average. Then,*

$$\tilde{\lambda} := \min_{v \in \tilde{H}^1(Y), [v] \neq 0} \frac{\int_Y \sigma |\nabla v|^2 dy}{\alpha \int_{\Gamma} [v]^2 d\sigma} > 0. \quad (4-15)$$

Remark 4.3 [[Amar et al. 2008](#)]. The change of variables $y = x/\varepsilon$ applied to [Equation \(4-15\)](#) yields

$$\min_{\substack{v \in \tilde{H}^1(\varepsilon Y) \\ [v] \neq 0}} \frac{\int_{\varepsilon Y} \sigma |\nabla v|^2 dx}{\frac{\alpha}{\varepsilon} \int_{\varepsilon \Gamma} [v]^2 d\sigma} = \tilde{\lambda} > 0, \quad (4-16)$$

where $\tilde{H}^1(\varepsilon Y) := \{v \in L^2(\varepsilon Y) : v|_{\varepsilon E_i} \in H^1(\varepsilon E_i), i = 1, 2, [v] \in \tilde{H}^{1/2}(\varepsilon \Gamma)\}$, $\tilde{H}^{1/2}(\varepsilon \Gamma)$ is comprised of the functions of $H^{1/2}(\varepsilon \Gamma)$ with null integral average, and $\tilde{\lambda}$ is the positive constant introduced in [Lemma 4.2](#).

5. A counterexample

As pointed out in the Introduction, the structure of (3-39) is not enough to imply that the solution exponentially decays to zero, nor does it imply the solution’s boundedness, even if an exponentially decaying memory kernel and source are considered. Indeed, let $\Omega = (-1, 1)$, $\mu > 0$, $a > 0$, $b \in \mathbb{R}$, and $f(x), h(x)$ be smooth functions. Consider the problem

$$\begin{cases} -\left(u_{0xt} + au_{0x} + b \int_0^t e^{-\mu(t-\tau)} u_{0x}(x, \tau) d\tau + f(x)e^{-\mu t}\right)_x = 0, \\ u_0(\pm 1, 0) = 0, \\ u_{0x}(x, 0) = h(x). \end{cases} \tag{5-1}$$

Multiplying the previous equation by $e^{\mu t}$, we obtain

$$u_{0xxt}e^{\mu t} + au_{0xx}e^{\mu t} + b \int_0^t e^{\mu\tau} u_{0xx}(x, \tau) d\tau = f'(x). \tag{5-2}$$

Setting $v(x, t) = u_{0xx}e^{\mu t}$ and differentiating with respect to t , Equation (5-2) can be rewritten as

$$v_{tt} + (a - \mu)v_t + bv = 0,$$

which must be complemented with the initial conditions

$$\begin{cases} v(x, 0) = h'(x), \\ v_t(x, 0) = f'(x) + (\mu - a)h'(x). \end{cases}$$

This last equation has an explicit solution (if $(\mu - a)^2 - 4b > 0$) of the form,

$$v(x, t) = C_1(x) \exp\left(\frac{\mu - a + \sqrt{(\mu - a)^2 - 4b}}{2}t\right) + C_2(x) \exp\left(\frac{\mu - a - \sqrt{(\mu - a)^2 - 4b}}{2}t\right),$$

where $C_1(x)$ and $C_2(x)$ are easily determined by using the initial conditions, thus implying that

$$u_{0xx}(x, t) = C_1(x) \exp\left(\frac{-\mu - a + \sqrt{(\mu - a)^2 - 4b}}{2}t\right) + C_2(x) \exp\left(\frac{-\mu - a - \sqrt{(\mu - a)^2 - 4b}}{2}t\right).$$

Hence, u_0 can be obtained by integrating twice with respect to x and using the previous mentioned boundary conditions.

Note that in general, if b is negative and $-b > \mu a$, the first exponential tends to infinity as $t \rightarrow +\infty$. With the exception of particular choices of the initial data, C_1 is different from zero, and hence solutions to Problem (5-1) do not, in general, decay exponentially in time.

References

[Amar et al. 2003] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “Homogenization limit for electrical conduction in biological tissues in the radio-frequency range”, *Comptes Rendus Mécanique* **331**:7 (2003), 503–508.
 [Amar et al. 2004a] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “An elliptic equation with history”, *C. R. Acad. Sci. Paris Sér. I Math.* **338**:8 (2004), 595–598.

- [Amar et al. 2004b] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “[Evolution and memory effects in the homogenization limit for electrical conduction in biological tissues](#)”, *Math. Models Methods Appl. Sci.* **14**:9 (2004), 1261–1295.
- [Amar et al. 2005] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “[Existence and uniqueness for an elliptic problem with evolution arising in electrodynamics](#)”, *Nonlinear Anal. Real World Appl.* **6**:2 (2005), 367–380.
- [Amar et al. 2006] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “[On a hierarchy of models for electrical conduction in biological tissues](#)”, *Math. Methods Appl. Sci.* **29**:7 (2006), 767–787.
- [Amar et al. 2008] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “Exponential asymptotic stability for an elliptic equation with memory arising in electrical conduction”, Preprint, Università di Roma “La Sapienza”, 2008.
- [Amar et al. 2009] M. Amar, D. Andreucci, P. Bisegna, and R. Gianni, “Homogenization limit for electrical conduction in biological tissues in the high radiofrequency range”, Preprint, Università di Roma “La Sapienza”, 2009.
- [Appleby et al. 2006] J. A. D. Appleby, M. Fabrizio, B. Lazzari, and D. W. Reynolds, “[On exponential asymptotic stability in linear viscoelasticity](#)”, *Math. Models Methods Appl. Sci.* **16**:10 (2006), 1677–1694.
- [Attouch 1984] H. Attouch, *Variational convergence for functionals and operators*, Pitman, London, 1984.
- [Bakhvalov and Panasenko 1989] N. Bakhvalov and G. Panasenko, *Homogenization: averaging processes in periodic media: mathematical problems in the mechanics of composite materials*, Mathematics and its Applications **36**, Kluwer, Dordrecht, 1989.
- [Bensoussan et al. 1978] A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, North Holland, Amsterdam, 1978.
- [Bisegna et al. 2001] P. Bisegna, G. Caruso, and F. Lebon, “Bioelectrical impedance analysis: a matter of homogenization of composites with imperfect interfaces”, in *Proceedings of the 15th AIMETA Congress of Theoretical and Applied Mechanics* (Taormina), edited by G. Augusti et al., 2001.
- [Braides and Defranceschi 1998] A. Braides and A. Defranceschi, *Homogenization of multiple integrals*, Oxford University Press, New York, 1998.
- [Bronzino 1999] J. D. Bronzino, *The biomedical engineering handbook*, 2nd ed., CRC Press, Boca Raton, FL, 1999.
- [Cioranescu and Donato 1999] D. Cioranescu and P. Donato, *An introduction to homogenization*, Oxford University Press, New York, 1999.
- [De Lorenzo et al. 1997] A. De Lorenzo, A. Andreoli, J. Matthie, and P. Withers, “[Predicting body cell mass with bioimpedance by using theoretical methods: a technological review](#)”, *J. Appl. Physiol.* **82**:5 (1997), 1542–1558.
- [Donato et al. 2007] P. Donato, L. Faella, and S. Monsurro, “[Homogenization of the wave equation in composites with imperfect interface: a memory effect](#)”, *J. Math. Pures Appl.* **87**:2 (2007), 119–143.
- [Fabrizio and Morro 1988] M. Fabrizio and A. Morro, “[Viscoelastic relaxation functions compatible with thermodynamics](#)”, *J. Elasticity* **19**:1 (1988), 63–75.
- [Fichera 1979] G. Fichera, “[Avere una memoria tenace crea gravi problemi](#)”, *Arch. Ration. Mech. An.* **70**:2 (1979), 101–112.
- [Foster and Schwan 1989] K. R. Foster and H. P. Schwan, “Dielectric properties of tissues and biological materials: a critical review”, *Crit. Rev. Biomed. Eng.* **17** (1989), 25–104.
- [Giorgi et al. 2001] C. Giorgi, M. G. Naso, and V. Pata, “Exponential stability in linear heat conduction with memory: a semigroup approach”, *Commun. Appl. Anal.* **5** (2001), 121–133.
- [Giorgi et al. 2005] C. Giorgi, M. G. Naso, and V. Pata, “[Energy decay of electromagnetic systems with memory](#)”, *Math. Models Methods Appl. Sci.* **15**:10 (2005), 1489–1502.
- [Hummel 2000] H.-K. Hummel, “[Homogenization for heat transfer in polycrystals with interfacial resistances](#)”, *Appl. Anal.* **75**:3–4 (2000), 403–424.
- [Jikov et al. 1994] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik, *Homogenization of differential operators and integral functionals*, Springer, Berlin, 1994.
- [Lazzari and Nibbi 2002] B. Lazzari and R. Nibbi, “[Sufficient conditions for the exponential stability in linear conductors with memory](#)”, *Int. J. Eng. Sci.* **40**:10 (2002), 1131–1149.
- [Lazzari and Vuk 1992] B. Lazzari and E. Vuk, “[Constitutive equations and quasi-static problem in linear thermoviscoelasticity](#)”, *Int. J. Eng. Sci.* **30**:4 (1992), 533–544.

- [Lene and Leguillon 1981] F. Lene and D. Leguillon, “Étude de l’influence d’un glissement entre les constituants d’un matériau composite sur ses coefficients de comportement effectifs”, *J. Méc.* **20** (1981), 509–536.
- [Lions 1981] J. L. Lions, *Some methods in the mathematical analysis of systems and their controls*, Science Press, Beijing, 1981.
- [Lipton 1998] R. Lipton, “Heat conduction in fine scale mixtures with interfacial contact resistance”, *SIAM J. Appl. Math.* **58**:1 (1998), 55–72.
- [Medjden and Tatar 2005] M. Medjden and N.-E. Tatar, “Asymptotic behavior for a viscoelastic problem with not necessarily decreasing kernel”, *Appl. Math. Comput.* **167**:2 (2005), 1221–1235.
- [Novožilov and Yappa 1978] Y. J. Novožilov and Y. A. Yappa, Электродинамика, Nauka, Moscow, 1978. Translated as *Electrodynamics*, Mir, Moscow, 1981.
- [Oleinik et al. 1992] O. A. Oleinik, A. S. Shamaev, and G. A. Yosifian, *Mathematical problems in elasticity and homogenization*, North-Holland, Amsterdam, 1992.
- [Sánchez-Hubert and Sánchez-Palencia 1992] J. Sánchez-Hubert and E. Sánchez-Palencia, *Introduction aux méthodes asymptotiques et à l’homogénéisation*, Masson, Paris, 1992.
- [Sánchez-Palencia 1980] E. Sánchez-Palencia, *Non-homogeneous media and vibration theory*, Lecture Notes in Physics **127**, Springer, Berlin, 1980.
- [Slemrod 1981] M. Slemrod, “Global existence, uniqueness, and asymptotic stability of classical smooth solutions in one-dimensional non-linear thermoelasticity”, *Arch. Ration. Mech. An.* **76**:2 (1981), 97–133.

Received 18 Dec 2007. Revised 16 Jul 2008. Accepted 3 Nov 2008.

MICOL AMAR: amar@dmmm.uniroma1.it

University of Rome La Sapienza, Department of Mathematical Methods and Models, Via Antonio Scarpa 16, 00161 Rome, Italy

DANIELE ANDREUCCI: andreucci@dmmm.uniroma1.it

University of Rome La Sapienza, Department of Mathematical Methods and Models, Via Antonio Scarpa 16, 00161 Rome, Italy

PAOLO BISEGNA: bisegna@uniroma2.it

University of Rome Tor Vergata, Department of Civil Engineering, via del Politecnico 1, 00133 Rome, Italy

ROBERTO GIANNI: gianni@dmmm.uniroma1.it

University of Rome La Sapienza, Department of Mathematical Methods and Models, Via Antonio Scarpa 16, 00161 Rome, Italy

THERMAL EFFECTS OF COLLISIONS: DOES RAIN TURN INTO ICE WHEN IT FALLS ON FROZEN GROUND?

ANNA MARIA CAUCCI AND MICHEL FRÉMOND

Dedicated to H. D. Bui, V. K. Dang, G. Del Piero, J. Salençon, J. Zarka.

The phase changes produced by the thermal effects of collisions are investigated. The behaviour of warm rain falling on a deeply frozen ground is predicted. The ice-water phase change involves microscopic motions that are taken into account in the predictive theory.

1. Introduction

Consider warm rain falling on frozen ground, or hailstones falling on warm ground. We will investigate the thermal consequences of such collisions. Does the rain freeze, turning into the “black ice” that can make paved roads so hazardous in winter? Does the hailstone melt? We will investigate these very fast phase changes due to collisions, providing a scheme for the thermomechanical theory of collisions involving phase change.

The classical ice-water phase change involves microscopic motions, which, in our view, must be taken into account in the macroscopic predictive theory. The basic idea we have developed is to account for the power of the microscopic motions in the power of the interior forces [Frémond 2001; 2007]. We modify the expression of the power of the interior forces and assume it depends on the liquid water volume fraction velocity, $d\beta/dt$, which is clearly related to microscopic motion. The consequences of this assumption give the basic equations of motion, one for macroscopic motion and another for microscopic motion.

We assume the collisions are instantaneous, and thus it is wise to assume that the resulting phase changes are also instantaneous; the very fast evolution of the liquid water volume fraction β is described by a discontinuity with respect to time. The discontinuities of the liquid water volume fraction $[\beta]$ intervene in nonsmooth equations of motion, accounting for microscopic motion. Moreover, when phase change occurs, the temperatures are discontinuous with respect to time.

In Section 2, we derive the equations of motion for a collision of two balls of ice. The balance laws are given in Section 3 and the constitutive laws in Section 4. In Section 5, examples of phase changes produced by the thermal effects of collisions are described: the collision of two pieces of ice (Section 5.2), and the collision of a warm droplet of rain falling on frozen ground (Section 6).

2. The equations of motion

Let us consider the system made of two balls of ice, \mathcal{B}_1 and \mathcal{B}_2 , which move along an axis and collide. Note that such a system is deformable, because the relative position of the balls changes. For the sake

Keywords: predictive theory, collisions, phase change, black ice.

of simplicity, we assume the balls to be points. We restrict our investigation of the motion and of the thermal evolution to collisions of the two balls. The smooth evolution is straightforward.

In the following discussion, we use indices $_1$ and $_2$ to denote physical quantities relative to the balls \mathcal{B}_1 and \mathcal{B}_2 . The indices $-$ and $+$ are used to designate the quantities before and after the collision. The time discontinuity of a function $t \mapsto \mathcal{A}(t)$ is denoted by

$$[\mathcal{A}] = \mathcal{A}^+ - \mathcal{A}^-.$$

The equations of motion are derived from the principle of virtual work which introduces different contributions: the virtual works of the acceleration, interior, and exterior forces [Frémond 2001; 2007].

The virtual work of the acceleration forces is

$$\mathcal{F}^{\text{acc}}(\mathbf{U}, \mathbf{V}) = m_1[U_1] \frac{\mathbf{V}_1^+ + \mathbf{V}_1^-}{2} + m_2[U_2] \frac{\mathbf{V}_2^+ + \mathbf{V}_2^-}{2},$$

where m_1 and m_2 are the masses of the two balls, $\mathbf{U} = (U_1, U_2)$ their actual velocities, and $\mathbf{V} = (V_1, V_2)$ are the macroscopic virtual velocities.

The virtual work of the interior forces, \mathcal{F}^{int} , is a linear function of the virtual velocity fields; in particular, it is chosen to depend on the system velocity of deformation, $D(\mathbf{V}) = V_1 - V_2$, and also on $[\delta] = ([\delta_1], [\delta_2])$, where δ_1 and δ_2 are the virtual liquid water volume fractions of the two balls in the ice-water phase change, the actual volume fractions being β_1 and β_2 . It is worth noting that $[\delta_i] = (\delta_i^+ - \delta_i^-)$, $i = 1, 2$, is analogous to $d\delta_i/dt$, which represents, in a smooth evolution, a virtual velocity of phase change. This latter quantity is clearly related to microscopic motion.

We choose the virtual work of the interior forces as

$$\mathcal{F}^{\text{int}}(D(\mathbf{V}), [\delta]) = -P^{\text{int}} D\left(\frac{\mathbf{V}^+ + \mathbf{V}^-}{2}\right) - A_1[\delta_1] - A_2[\delta_2],$$

where P^{int} is the interior percussion that intervenes when collisions occur, and $A = (A_1, A_2)$ are the interior microscopic works.

Assuming no exterior percussion, for instance when there is no hammer stroke, and no exterior electrical, radiative, or chemical impulse, the virtual work of the exterior forces is $\mathcal{F}^{\text{ext}}(\mathbf{V}) = 0$.

The principle of virtual work,

$$\forall \mathbf{V}, [\delta]: \quad \mathcal{F}^{\text{acc}}(\mathbf{V}) = \mathcal{F}^{\text{int}}(D(\mathbf{V}), [\delta]),$$

gives easily the equations of motion

$$m_1[U_1] = -P^{\text{int}}, \quad m_2[U_2] = P^{\text{int}} \quad (2-1)$$

$$A_1 = 0, \quad A_2 = 0. \quad (2-2)$$

Remark 2.1. We have assumed that there is no exterior impulse inducing phase change. In case there is one, for instance an electrical impulse, or a chemical impulse able to produce a phase change, A_1^{ext} , the equation of motion (2-2) becomes $A_1 = A_1^{\text{ext}}$.

3. The laws of thermodynamics

The laws of thermodynamics are the same for the two balls.

3.1. The first law. We first consider each ball separately. For \mathcal{B}_1 , the first law of thermodynamics is

$$[\mathcal{E}_1] + [\mathcal{K}_1] = \mathcal{T}_1^{\text{ext}}(\mathbf{U}) + \mathcal{C}_1, \quad (3-1)$$

where \mathcal{E}_1 is the internal energy of the ball, \mathcal{K}_1 its kinetic energy, $\mathcal{T}_1^{\text{ext}}(\mathbf{U})$ the actual work of the percussions which are exterior to ball \mathcal{B}_1 , and $\mathcal{C}_1 = T_1^+(B_1^+ + B_{12}^+) + T_1^-(B_1^- + B_{12}^-)$ the heat impulse provided to the ball at the time of collision. This quantity includes the heats $T_1 B_1$ received from outside the system, and the heats $T_1 B_{12}$ received from inside, that is, from the other ball. We assume that these heats are received at a temperature of either T_1^+ or T_1^- .

The theorem of kinetic energy for ball \mathcal{B}_1 is

$$[\mathcal{K}_1] = \mathcal{T}_1^{\text{acc}}(\mathbf{U}) = \mathcal{T}_1^{\text{int}}([\beta_1]) + \mathcal{T}_1^{\text{ext}}(\mathbf{U}) = -A_1[\beta_1] + \mathcal{T}_1^{\text{ext}}(\mathbf{U}),$$

with $\mathcal{T}_1^{\text{int}}([\delta_1]) = -A_1[\delta_1]$. It gives, with the first law of thermodynamics (3-1),

$$\begin{aligned} [\mathcal{E}_1] &= \mathcal{C}_1 - \mathcal{T}_1^{\text{int}}([\beta_1]) = T_1^+(B_1^+ + B_{12}^+) + T_1^-(B_1^- + B_{12}^-) + A_1[\beta_1] \\ &= \bar{T}_1(\Sigma B_1 + \Sigma B_{12}) + [T_1](\Delta B_1 + \Delta B_{12}) + A_1[\beta_1], \end{aligned} \quad (3-2)$$

with the notations

$$\bar{T} = \frac{T^+ + T^-}{2}, \quad \Sigma B = B^+ + B^-, \quad \Delta B = \frac{B^+ - B^-}{2},$$

retained in the sequel.

Now we consider the system as a whole. The internal energy of the system is the sum of the internal energies of its components, \mathcal{E}_1 and \mathcal{E}_2 , to which an interaction internal energy, \mathcal{E}^{int} , may be added, giving $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}^{\text{int}}$.

The first law of thermodynamics for the system is $[\mathcal{E}] + [\mathcal{K}] = \mathcal{T}^{\text{ext}}(\mathbf{U}) + \mathcal{C}$, where \mathcal{K} is the kinetic energy of the whole system and \mathcal{C} is the exterior heat impulse received by the system in collision, given by $\mathcal{C} = T_1^+ B_1^+ + T_1^- B_1^- + T_2^+ B_2^+ + T_2^- B_2^-$.

The theorem of kinetic energy, that is, the principle of virtual power with the actual velocities

$$[\mathcal{K}] = \mathcal{T}^{\text{acc}}(\mathbf{U}) = \mathcal{T}^{\text{int}}(D(\mathbf{U}), [\beta_1], [\beta_2]) + \mathcal{T}^{\text{ext}}(\mathbf{U}),$$

and the first law of thermodynamics gives

$$[\mathcal{E}] = \mathcal{C} - \mathcal{T}^{\text{int}}(D(\mathbf{U}), [\beta_1], [\beta_2]) = \mathcal{C} + P^{\text{int}} D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right) + A_1[\beta_1] + A_2[\beta_2]. \quad (3-3)$$

Combining (3-2) and (3-3) we obtain

$$\begin{aligned} [\mathcal{E}] &= [\mathcal{E}_1] + [\mathcal{E}_2] + [\mathcal{E}^{\text{int}}] \\ &= \bar{T}_1(\Sigma B_1 + \Sigma B_{12}) + [T_1](\Delta B_1 + \Delta B_{12}) + A_1[\beta_1] \\ &\quad + \bar{T}_2(\Sigma B_2 + \Sigma B_{21}) + [T_2](\Delta B_2 + \Delta B_{21}) + A_2[\beta_2] + [\mathcal{E}^{\text{int}}] \\ &= T_1^+ B_1^+ + T_1^- B_1^- + T_2^+ B_2^+ + T_2^- B_2^- + P^{\text{int}} D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right) + A_1[\beta_1] + A_2[\beta_2] \\ &= \bar{T}_1 \Sigma B_1 + [T_1] \Delta B_1 + \bar{T}_2 \Sigma B_2 + [T_2] \Delta B_2 + P^{\text{int}} D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right) + A_1[\beta_1] + A_2[\beta_2]. \end{aligned}$$

Then

$$[\mathcal{E}^{\text{int}}] = P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - \bar{T}_1 \Sigma B_{12} - [T_1] \Delta B_{12} - \bar{T}_2 \Sigma B_{21} - [T_2] \Delta B_{21}. \quad (3-4)$$

This relationship is interesting because only interior quantities intervene.

3.2. The second law of thermodynamics. Again we start with each ball separately. Let \mathcal{S}_1 be the entropy of ball \mathcal{B}_1 . The second law of thermodynamics is

$$[\mathcal{S}_1] \geq B_1^+ + B_1^- + B_{12}^+ + B_{12}^- = \Sigma B_1 + \Sigma B_{12}, \quad (3-5)$$

where the quantity on the right-hand side is the sum of the exterior entropy impulses B_1 received from outside the system and the entropy impulses B_{12} received from ball \mathcal{B}_2 .

By means of relationship (3-2), the second law of thermodynamics gives

$$[\mathcal{E}_1] - \bar{T}_1 [\mathcal{S}_1] \leq [T_1] (\Delta B_1 + \Delta B_{12}) + A_1 [\beta_1],$$

or, by introducing the free energy $\Psi = \mathcal{E} - T\mathcal{S}$,

$$[\Psi_1] + \bar{\mathcal{F}}_1 [T_1] \leq [T_1] (\Delta B_1 + \Delta B_{12}) + A_1 [\beta_1]. \quad (3-6)$$

The free energy of ball \mathcal{B}_1 is

$$\Psi_1(T_1, \beta_1) = -C_1 T_1 \log T_1 - \beta_1 \frac{L}{T_0} (T_1 - T_0) + I(\beta_1),$$

where C_1 is the heat capacity, L is the latent heat at the phase change temperature T_0 , and I is the indicator function of the interval $[0, 1]$ (see [Moreau 1966–1967]), which takes into account the internal constraint on the volume fraction $0 \leq \beta_1 \leq 1$. We set

$$\widehat{\Psi}_1(T_1, \beta_1) = -\beta_1 \frac{L}{T_0} (T_1 - T_0) + I(\beta_1).$$

Then

$$[\Psi_1] = [-C_1 T_1 \log T_1] + [\widehat{\Psi}_1], \quad (3-7)$$

where

$$[\widehat{\Psi}_1] = \widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^-, \beta_1^-) = \widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) + \widehat{\Psi}_1(T_1^+, \beta_1^-) - \widehat{\Psi}_1(T_1^-, \beta_1^-) \quad (3-8)$$

and

$$\widehat{\Psi}_1(T_1^+, \beta_1^-) - \widehat{\Psi}_1(T_1^-, \beta_1^-) = -\beta_1^- \frac{L}{T_0} [T_1]. \quad (3-9)$$

In view of (3-7)–(3-9), inequality (3-6) transforms into

$$\begin{aligned} [\Psi_1] + \bar{\mathcal{F}}_1 [T_1] &= [-C_1 T_1 \log T_1] + \widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) - \beta_1^- \frac{L}{T_0} [T_1] + \bar{\mathcal{F}}_1 [T_1] \\ &\leq [T_1] (\Delta B_1 + \Delta B_{12}) + A_1 [\beta_1]. \end{aligned} \quad (3-10)$$

Since

$$\frac{[-C_1 T_1 \log T_1]}{[T_1]}$$

has a limit when $[T_1] \rightarrow 0$, we introduce the notation

$$[-C_1 T_1 \log T_1] + \left(\bar{\mathcal{F}}_1 - \beta_1^- \frac{L}{T_0} \right) [T_1] = -\mathcal{S}_1 [T_1]. \quad (3-11)$$

From (3-10), we have

$$\widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) \leq [T_1](\Delta B_1 + \Delta B_{12} + \mathcal{S}_1) + A_1[\beta_1]. \quad (3-12)$$

It is reasonable to assume that there is no dissipation with respect to $[T_1]$ as there is no dissipation with respect to dT_1/dt in a smooth evolution

$$\Delta B_1 + \Delta B_{12} + \mathcal{S}_1 = 0; \quad (3-13)$$

then, inequality (3-12) becomes

$$\widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) \leq A_1[\beta_1]. \quad (3-14)$$

Turning now to the system as a whole, its entropy is given by the sum

$$\mathcal{S} = \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}^{\text{int}}, \quad (3-15)$$

where \mathcal{S}^{int} is the interaction entropy. The second law of thermodynamics is

$$[\mathcal{S}] = [\mathcal{S}_1] + [\mathcal{S}_2] + [\mathcal{S}^{\text{int}}] \geq B_1^+ + B_1^- + B_2^+ + B_2^- = \Sigma B_1 + \Sigma B_2. \quad (3-16)$$

In the sequel, we choose constitutive laws such that laws of thermodynamics (3-5) and (3-16) are satisfied. They will be such that (3-5) and

$$[\mathcal{S}^{\text{int}}] \geq -\Sigma B_{12} - \Sigma B_{21} \quad (3-17)$$

are verified.

Let us get inequalities equivalent to (3-17). Let

$$\Theta = \frac{T_1 + T_2}{2}.$$

From (3-4) and (3-17), we have

$$\begin{aligned} [\mathcal{E}^{\text{int}}] - \bar{\Theta}[\mathcal{S}^{\text{int}}] \leq P^{\text{int}} D\left(\frac{U^+ + U^-}{2}\right) - \bar{T}_1 \Sigma B_{12} - [T_1] \Delta B_{12} \\ - \bar{T}_2 \Sigma B_{21} - [T_2] \Delta B_{21} + \bar{\Theta}(\Sigma B_{12} + \Sigma B_{21}). \end{aligned} \quad (3-18)$$

Introducing the free energy of interaction $\Psi^{\text{int}} = \mathcal{E}^{\text{int}} - \Theta \mathcal{S}^{\text{int}}$, inequality (3-18) can be written as

$$\begin{aligned} [\Psi^{\text{int}}] + \bar{\mathcal{F}}^{\text{int}}[\Theta] \leq P^{\text{int}} D\left(\frac{U^+ + U^-}{2}\right) - \bar{T}_1 \Sigma B_{12} - [T_1] \Delta B_{12} \\ - \bar{T}_2 \Sigma B_{21} - [T_2] \Delta B_{21} + \bar{\Theta}(\Sigma B_{12} + \Sigma B_{21}), \end{aligned} \quad (3-19)$$

where

$$\bar{\mathcal{F}}^{\text{int}} = \frac{(\mathcal{S}^{\text{int}})^+ + (\mathcal{S}^{\text{int}})^-}{2}, \quad [\Theta] = \frac{[T_1] + [T_2]}{2}.$$

We assume Ψ^{int} is a smooth function of Θ . Then when $[\Theta]$ tends to 0 and Θ^+ and Θ^- have limits

$$\lim_{[\Theta] \rightarrow 0} \Theta^+ = \lim_{[\Theta] \rightarrow 0} \Theta^- = \Theta,$$

we have

$$\lim_{[\Theta] \rightarrow 0} \left(\frac{[\Psi^{\text{int}}]}{[\Theta]} + \bar{\mathcal{F}}^{\text{int}} \right) = 0,$$

since

$$\lim_{[\Theta] \rightarrow 0} \frac{[\Psi^{\text{int}}]}{[\Theta]} = \frac{\partial \Psi^{\text{int}}}{\partial \Theta}(\Theta), \quad \lim_{[\Theta] \rightarrow 0} \bar{\mathcal{F}}^{\text{int}} = \mathcal{F}^{\text{int}}(\Theta),$$

and the Helmholtz relationship

$$\mathcal{F}^{\text{int}}(\Theta) = -\frac{\partial \Psi^{\text{int}}}{\partial \Theta}(\Theta).$$

Therefore, we can put

$$[\Psi^{\text{int}}] + \bar{\mathcal{F}}^{\text{int}}[\Theta] = -\mathcal{F}^{\text{int}}[\Theta], \quad (3-20)$$

and, from (3-19), we find

$$\begin{aligned} 0 \leq P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - \bar{T}_1 \Sigma B_{12} - [T_1] \left(\Delta B_{12} - \frac{\mathcal{G}^{\text{int}}}{2} \right) \\ - \bar{T}_2 \Sigma B_{21} - [T_2] \left(\Delta B_{21} - \frac{\mathcal{G}^{\text{int}}}{2} \right) + \bar{\Theta} (\Sigma B_{12} + \Sigma B_{21}). \end{aligned} \quad (3-21)$$

It is reasonable to assume that there is no dissipation with respect to $[T_1]$ and $[T_2]$ as there is no dissipation with respect to the dT/dt 's in a smooth evolution

$$\Delta B_{12} - \frac{\mathcal{G}^{\text{int}}}{2} = 0, \quad \Delta B_{21} - \frac{\mathcal{G}^{\text{int}}}{2} = 0; \quad (3-22)$$

then

$$0 \leq P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - \bar{T}_1 \Sigma B_{12} - \bar{T}_2 \Sigma B_{21} + \bar{\Theta} (\Sigma B_{12} + \Sigma B_{21}). \quad (3-23)$$

But

$$\bar{T}_1 \Sigma B_{12} + \bar{T}_2 \Sigma B_{21} = \bar{\Theta} (\Sigma B_{12} + \Sigma B_{21}) + \delta \bar{T} \frac{\Sigma B_{21} - \Sigma B_{12}}{2}, \quad (3-24)$$

where the difference in temperature, $\delta \bar{T}$, is defined by $\delta \bar{T} = \bar{T}_2 - \bar{T}_1$. Finally, inequality (3-23) yields

$$0 \leq P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - \delta \bar{T} \frac{\Sigma B_{21} - \Sigma B_{12}}{2}; \quad (3-25)$$

this last relationship links mechanical and thermal dissipations.

On the other hand, from (3-25), by means of (3-24) and assuming no dissipation with respect to $[T_1]$ and $[T_2]$, it follows that (3-21) is satisfied; thanks to (3-20), inequality (3-21) yields

$$\begin{aligned} [\Psi^{\text{int}}] \leq P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - \bar{T}_1 \Sigma B_{12} - [T_1] \Delta B_{12} - \bar{T}_2 \Sigma B_{21} \\ - [T_2] \Delta B_{21} + \bar{\Theta} (\Sigma B_{12} + \Sigma B_{21}) - \bar{\mathcal{F}}^{\text{int}}[\Theta]. \end{aligned} \quad (3-26)$$

Substituting the identity $[\Psi^{\text{int}}] = [\mathcal{E}^{\text{int}}] - [\Theta]\bar{\mathcal{G}}^{\text{int}} - \bar{\Theta}[\mathcal{G}^{\text{int}}]$ into (3-26), we have

$$-\bar{\Theta}[\mathcal{G}^{\text{int}}] \leq -[\mathcal{E}^{\text{int}}] + P^{\text{int}}D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right) - \bar{T}_1 \Sigma B_{12} - [T_1] \Delta B_{12} \\ - \bar{T}_2 \Sigma B_{21} - [T_2] \Delta B_{21} + \bar{\Theta}(\Sigma B_{12} + \Sigma B_{21}); \quad (3-27)$$

this in turn, by virtue of (3-4), entails $-\bar{\Theta}[\mathcal{G}^{\text{int}}] \leq \bar{\Theta}(\Sigma B_{12} + \Sigma B_{21})$, which is inequality (3-17).

As a consequence, the following theorem holds.

Theorem 3.1. *Let Ψ^{int} be a smooth function of Θ . We assume (3-22) holds, that is, there is no dissipation with respect to $[T_1]$ and $[T_2]$. If (3-4) is satisfied, then (3-17) and (3-25) are equivalent.*

4. The constitutive laws

For ball \mathcal{B}_1 , we split the interior force A_1 into a possibly nondissipative part, indexed by $^{\text{nd}}$, and a dissipative part, indexed by $^{\text{d}}$:

$$A_1 = A_1^{\text{nd}} + A_1^{\text{d}}. \quad (4-1)$$

The interior force A_1^{nd} is defined by the free energy $\Psi_1(T_1, \beta_1)$ as

$$A_1^{\text{nd}} \in \partial\Psi_1(T_1^+, \beta_1^+) = -\frac{L}{T_0}(T_1^+ - T_0) + A_1^{\text{ndr}}, \quad (4-2)$$

where the subdifferential $\partial\Psi_1$ [Moreau 1966–1967] is computed with respect to β_1^+ and the reaction A_1^{ndr} is such that

$$A_1^{\text{ndr}} \in \partial I(\beta_1^+). \quad (4-3)$$

The dissipative interior force A_1^{d} is defined by

$$A_1^{\text{d}} \in \partial\Phi_1([\beta_1]), \quad (4-4)$$

and satisfies

$$A_1^{\text{d}}[\beta_1] \geq 0, \quad (4-5)$$

where $\Phi_1([\beta_1])$ is a pseudopotential of dissipation [Moreau 1970] and the subdifferential is computed with respect to $[\beta_1]$.

For the system, we choose the pseudopotential of dissipation

$$\Phi\left(D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right), \delta\bar{T}, \chi\right),$$

which depends on the velocity $D((\mathbf{U}^+ + \mathbf{U}^-)/2)$, on the thermal heterogeneity $\delta\bar{T}$, and on the quantity $\chi = D(\mathbf{U}^-/2)$ depending on the past, to ensure the noninterpenetration of the two balls [Frémond 2001; 2007].

The constitutive laws are relationships (4-2)–(4-5) and

$$\left(P^{\text{int}}, \frac{\Sigma B_{21} - \Sigma B_{12}}{2}\right) \in \partial\Phi\left(D\left(\frac{\mathbf{U}^+ + \mathbf{U}^-}{2}\right), \delta\bar{T}, \chi\right), \quad (4-6)$$

where the subdifferential set of Φ is computed with respect to the first two variables.

Remark 4.1. The reaction to the internal constraint $0 \leq \beta^+ \leq 1$ is in general dissipative, because $\partial I(\beta^+)(\beta^+ - \beta^-)$ may be different from zero — in fact, positive. Thus, the reaction is dissipative, as we may expect. This property is not true in a smooth evolution where the internal constraint is workless or nondissipative, because

$$\partial I(\beta) \frac{d\beta}{dt} = 0, \quad \text{almost everywhere.}$$

Remark 4.2. From (4-6), applying the classical properties of pseudopotentials [Moreau 1970] we see that inequality (3-25) is satisfied.

Theorem 4.3. *If the constitutive laws (4-2) and (4-4) are satisfied, then inequality (3-14) holds and the internal constraint $0 \leq \beta_1^+ \leq 1$ is verified.*

Proof. If $A_1^{\text{nd}} \in \partial \Psi_1(T_1^+, \beta_1^+)$, then, using the definition of subdifferential, we have

$$\widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) \leq A_1^{\text{nd}}(\beta_1^+ - \beta_1^-) = A_1^{\text{nd}}[\beta_1];$$

with (4-5), resulting from (4-4), the previous inequality entails

$$\widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) \leq A_1^{\text{nd}}[\beta_1] + A_1^{\text{d}}[\beta_1] = A_1[\beta_1].$$

Since $\partial \Psi_1(T_1^+, \beta_1^+) \neq \emptyset$, the internal constraint $0 \leq \beta_1^+ \leq 1$ is verified. \square

Theorem 4.4. *Assume (3-13) and (3-22) hold, that is, there is no dissipation with respect to $[T_1]$ and $[T_2]$. If the first laws of thermodynamics (3-2) and (3-3) and the constitutive laws (4-2), (4-3), (4-4) and (4-6) are verified, then the second law of thermodynamics holds for each ball and for the system.*

Proof. Due to Theorem 4.3 and relationships (3-7)–(3-9), we find

$$\begin{aligned} [\Psi_1] &= [-C_1 T_1 \log T_1] + \widehat{\Psi}_1(T_1^+, \beta_1^+) - \widehat{\Psi}_1(T_1^+, \beta_1^-) - \beta_1^- \frac{L}{T_0} [T_1] \\ &\leq [-C_1 T_1 \log T_1] + A_1[\beta_1] - \beta_1^- \frac{L}{T_0} [T_1]. \end{aligned} \quad (4-7)$$

On the other hand, we have

$$[\Psi_1] = [\mathcal{E}_1] - [T_1 \mathcal{S}_1] = [\mathcal{E}_1] - [T_1] \bar{\mathcal{S}}_1 - \bar{T}_1 [\mathcal{S}_1]. \quad (4-8)$$

From (3-2), (4-7) and (4-8), it follows that

$$\begin{aligned} \bar{T}_1 [\mathcal{S}_1] &\geq [\mathcal{E}_1] - [T_1] \bar{\mathcal{S}}_1 + [C_1 T_1 \log T_1] - A_1[\beta_1] + \beta_1^- \frac{L}{T_0} [T_1] \\ &= \bar{T}_1 (\Sigma B_1 + \Sigma B_{12}) + [T_1] (\Delta B_1 + \Delta B_{12}) + A_1[\beta_1] - [T_1] \bar{\mathcal{S}}_1 \\ &\quad + [C_1 T_1 \log T_1] - A_1[\beta_1] + \beta_1^- \frac{L}{T_0} [T_1] \\ &= \bar{T}_1 (\Sigma B_1 + \Sigma B_{12}) + [T_1] (\Delta B_1 + \Delta B_{12} + \mathcal{S}_1), \end{aligned}$$

with \mathcal{S}_1 defined as in (3-11). From the previous inequality, assuming no dissipation with respect to $[T_1]$ (relationship (3-13)), we obtain

$$[\mathcal{S}_1] \geq \Sigma B_1 + \Sigma B_{12}, \quad (4-9)$$

that is, the second law of thermodynamics for ball \mathcal{B}_1 .

It is worth noting that, if (4-9) holds for each ball, then, in view of (3-15) and (3-17), we have

$$[\mathcal{S}] = [\mathcal{S}_1] + [\mathcal{S}_2] + [\mathcal{S}^{\text{int}}] \geq \Sigma B_1 + \Sigma B_{12} + \Sigma B_2 + \Sigma B_{21} - \Sigma B_{12} - \Sigma B_{21} = \Sigma B_1 + \Sigma B_2,$$

which is the second law of thermodynamics for the system. Therefore, all that is necessary is proving that our hypotheses imply (3-17). Due to the constitutive law (4-6), inequality (3-25) is satisfied, and, from Theorem 3.1, our conclusion follows. \square

5. Examples of thermal effects with phase changes

We choose a pseudopotential of dissipation, Φ , without nondiagonal terms, for example

$$\Phi\left(D\left(\frac{U^+ + U^-}{2}\right), \delta\bar{T}, D\left(\frac{U^-}{2}\right)\right) = \Phi^{\text{mech}}\left(D\left(\frac{U^+ + U^-}{2}\right), D\left(\frac{U^-}{2}\right)\right) + \frac{\lambda}{4}(\delta\bar{T})^2;$$

with this choice, the mechanical problem is split from the thermal one. The thermal constitutive law is

$$\Sigma B_{21} - \Sigma B_{12} = -\lambda\delta\bar{T}. \quad (5-1)$$

Let us recall that the free energies chosen for the balls are

$$\Psi_i(T_i, \beta_i) = -C_i T_i \log T_i - \beta_i \frac{L}{T_0} (T_i - T_0) + I(\beta_i), \quad i = 1, 2.$$

We choose the free energy of interaction $\Psi^{\text{int}} = 0$; we have $\mathcal{G}^{\text{int}} = 0$ and, from relationships (3-22), $\Delta B_{12} = 0$, $\Delta B_{21} = 0$.

Now, we suppose that the macroscopic mechanical problem is solved, that is, we know the quantity $P^{\text{int}} D(U^+ + U^-)/2$ and we assume no external percussion work $A_1 = A_2 = 0$, due to the equations of motion (2-2); the thermal equations are

$$\begin{aligned} [\mathcal{S}_1] &= [C_1 \log T_1] + \frac{L}{T_0} [\beta_1] = C_1 \log \frac{T_1^+}{T_1^-} + \frac{L}{T_0} [\beta_1] = \Sigma B_1 + \Sigma B_{12}, \\ [\mathcal{S}_2] &= [C_2 \log T_2] + \frac{L}{T_0} [\beta_2] = C_2 \log \frac{T_2^+}{T_2^-} + \frac{L}{T_0} [\beta_2] = \Sigma B_2 + \Sigma B_{21}, \\ [\mathcal{E}_1] &= C_1 [T_1] + L [\beta_1] = \bar{T}_1 (\Sigma B_1 + \Sigma B_{12}) + [T_1] \Delta B_1, \\ [\mathcal{E}_2] &= C_2 [T_2] + L [\beta_2] = \bar{T}_2 (\Sigma B_2 + \Sigma B_{21}) + [T_2] \Delta B_2, \\ [\mathcal{G}^{\text{int}}] &= P^{\text{int}} D\left(\frac{U^+ + U^-}{2}\right) - \bar{T}_1 \Sigma B_{12} - \bar{T}_2 \Sigma B_{21} = 0, \end{aligned}$$

where the mechanical dissipation $P^{\text{int}} D((U^+ + U^-)/2)$ is positive due to the properties of pseudopotentials of dissipation. These equations are completed by the description of thermal relationships between the system and the outside and by the equations of microscopic motion

$$0 = A_i = A_i^{\text{nd}} + A_i^{\text{d}} \in -\frac{L}{T_0} (T_i^+ - T_0) + \partial I(\beta_i^+) + \partial \Phi_i([\beta_i]), \quad i = 1, 2.$$

We assume the collision is adiabatic, that is, no heat is exchanged with the exterior:

$$\begin{aligned} T_1^+ B_1^+ + T_1^- B_1^- &= \bar{T}_1 \Sigma B_1 + [T_1] \Delta B_1 = 0, \\ T_2^+ B_2^+ + T_2^- B_2^- &= \bar{T}_2 \Sigma B_2 + [T_2] \Delta B_2 = 0, \end{aligned}$$

and we have the equations

$$[\mathcal{G}_1] = C_1 \log \frac{T_1^+}{T_1^-} + \frac{L}{T_0} [\beta_1] = \Sigma B_1 + \Sigma B_{12},$$

$$[\mathcal{G}_2] = C_2 \log \frac{T_2^+}{T_2^-} + \frac{L}{T_0} [\beta_2] = \Sigma B_2 + \Sigma B_{21},$$

$$C_1 [T_1] + L [\beta_1] = \bar{T}_1 \Sigma B_{12}, \quad C_2 [T_2] + L [\beta_2] = \bar{T}_2 \Sigma B_{21},$$

$$P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) = C_1 [T_1] + L [\beta_1] + C_2 [T_2] + L [\beta_2], \quad (5-2)$$

$$C_2 \frac{[T_2]}{T_2} + \frac{L}{T_2} [\beta_2] - C_1 \frac{[T_1]}{T_1} - \frac{L}{T_1} [\beta_1] = -\lambda \delta \bar{T}. \quad (5-3)$$

The last two equations give the temperatures T_1^+ and T_2^+ after the collision, and the first give the entropic heat exchanges, ΣB_1 and ΣB_2 , with the outside.

We assume also small perturbations, that is $T_i^\pm = T_0 + \theta_i^\pm$, $|\theta_i^\pm| \ll T_0$. We can write (5-2) and (5-3) as

$$P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) = C_1 [\theta_1] + L [\beta_1] + C_2 [\theta_2] + L [\beta_2], \quad (5-4)$$

$$C_2 \frac{[\theta_2]}{T_0} + \frac{L}{T_0} [\beta_2] - C_1 \frac{[\theta_1]}{T_0} - \frac{L}{T_0} [\beta_1] = -\lambda \left(\theta_2^- - \theta_1^- + \frac{[\theta_2]}{2} - \frac{[\theta_1]}{2} \right), \quad (5-5)$$

and we get the system

$$C_1 [\theta_1] = \frac{1}{2} \left(P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - 2L [\beta_1] + \lambda T_0 \left(\theta_2^- - \theta_1^- + \frac{[\theta_2]}{2} - \frac{[\theta_1]}{2} \right) \right), \quad (5-6)$$

$$C_2 [\theta_2] = \frac{1}{2} \left(P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - 2L [\beta_2] - \lambda T_0 \left(\theta_2^- - \theta_1^- + \frac{[\theta_2]}{2} - \frac{[\theta_1]}{2} \right) \right), \quad (5-7)$$

where the volume fractions β_1 and β_2 satisfy the equations of microscopic motion

$$0 \in -\frac{L}{T_0} (T_1^+ - T_0) + \partial I(\beta_1^+) + \partial \Phi_1([\beta_1]), \quad (5-8)$$

$$0 \in -\frac{L}{T_0} (T_2^+ - T_0) + \partial I(\beta_2^+) + \partial \Phi_2([\beta_2]), \quad (5-9)$$

which are equivalent to

$$\theta_1^+ \in \partial I(\beta_1^+) + \frac{T_0}{L} \partial \Phi_1([\beta_1]), \quad (5-10)$$

$$\theta_2^+ \in \partial I(\beta_2^+) + \frac{T_0}{L} \partial \Phi_2([\beta_2]). \quad (5-11)$$

It is easy to prove that Equations (5-6), (5-10), and (5-7), (5-11), giving the temperatures and phase fractions θ_1^+, θ_2^+ and β_1^+, β_2^+ , have unique solutions.

5.1. Collision of two pieces of ice at the same temperature. The temperatures after the collision are equal. Equations (5-6) and (5-10) show that they do not depend on λ .

We consider two identical balls of ice, $C_1 = C_2 = C$, at the same temperature, θ^- , before the collision. The balls have the same temperature and volume fraction, denoted by θ^+ and β^+ , after the collision:

$$\begin{aligned} \theta_1^- = \theta_2^- = \theta^-, & \quad \theta_1^+ = \theta_2^+ = \theta^+, \\ \beta_1^- = \beta_2^- = 0, & \quad \beta_1^+ = \beta_2^+ = \beta^+. \end{aligned} \tag{5-12}$$

We know that the smooth ice-water phase change is not dissipative. Thus we assume we're not in a nonsmooth situation; there is no dissipation with respect to the volume fractions' discontinuities $[\beta_1]$ and $[\beta_2]$:

$$\Phi_1([\beta_1]) = \Phi_2([\beta_2]) = 0. \tag{5-13}$$

Equations (5-6) and (5-7) give

$$C[\theta] = \frac{1}{2}(\mathcal{T} - 2L\beta^+), \tag{5-14}$$

where

$$\mathcal{T} = P^{\text{int}} D\left(\frac{U^+ + U^-}{2}\right) \geq 0.$$

Because of (5-13), we find easily from the equations of microscopic motion (5-10) and (5-11)

$$\theta^+ \in \partial I(\beta^+). \tag{5-15}$$

Theorem 5.1. (1) *If $\mathcal{T} \leq -2C\theta^-$, then $\beta^+ = 0$: the ice does not melt and has temperature (5-14) after collision.*

(2) *If $\mathcal{T} \geq 2(L - C\theta^-)$, then $\beta^+ = 1$: the ice melts. The liquid water has temperature (5-14) after collision.*

(3) *If $-2C\theta^- < \mathcal{T} < 2(L - C\theta^-)$, then $0 < \beta^+ < 1$: after collision there is a mixture of ice and liquid water with temperature $\theta^+ = 0$.*

Proof. (1) If $\mathcal{T} \leq -2C\theta^-$, then, from (5-14), it follows that $C[\theta] + L\beta^+ \leq -C\theta^-$. The previous inequality entails

$$\theta^+ \leq -\frac{L}{C}\beta^+ \leq 0;$$

therefore, in view of (5-15), we obtain $\beta^+ = 0$.

Cases (2) and (3) are easily proved. □

The results of the theorem agree with what is expected: A violent collision produces a phase change whereas a nonviolent collision does not. Violent means dissipative, that is, \mathcal{T} large.

Remark 5.2. In the case where the phase change, for another material, is dissipative, we choose the pseudopotential of dissipation as

$$\Phi_i([\beta_i]) = \frac{c}{2}[\beta_i]^2, \quad i = 1, 2,$$

where c is a positive constant. Thus, the equations of motion (5-8) give

$$\frac{L}{T_0}\theta_1^+ \in \partial I(\beta_1^+) + c[\beta_1], \quad \frac{L}{T_0}\theta_2^+ \in \partial I(\beta_2^+) + c[\beta_2],$$

and we have the equations

$$C[\theta] = \frac{1}{2}(\mathcal{T} - 2L\beta^+), \quad \frac{L}{T_0}\theta^+ \in \partial I(\beta^+) + c\beta^+.$$

The following result is easily proved, assuming (5-12) before collision.

- Theorem 5.3.** (1) *If $\mathcal{T} \leq -2C\theta^-$, then $\beta^+ = 0$ and there is solid with $\theta^+ < cT_0/L$.*
 (2) *If $\mathcal{T} \geq 2(L - C\theta^- + cCT_0/L)$, then $\beta^+ = 1$ and there is liquid with $\theta^+ \geq cT_0/L$.*
 (3) *If $-2C\theta^- < \mathcal{T} < 2(L - C\theta^- + cCT_0/L)$, then $0 < \beta^+ < 1$ and there is a mixture of solid and liquid with $\theta^+ = cT_0\beta^+/L$.*

When there is dissipation the collision has to be more violent to melt the solid balls. The phase change occurs with temperature slightly above T_0 , as is the case for dissipative phase changes [Frémond and Visintin 1985; Frémond 2001; 2005].

5.2. Collision of two pieces of ice at different temperatures. When two pieces of ice at different temperatures collide, the dissipation due to the collision may be large enough to melt the warmest of them. We look for conditions on the state quantities before the collision and on the dissipated work, such that this phenomenon occurs. We expect that the temperatures before collision cannot be very cold and that the dissipated work has to be large.

We assume that there is no dissipation with respect to $[\beta_1]$ and $[\beta_2]$, that is,

$$\Phi_1([\beta_1]) = \Phi_2([\beta_2]) = 0,$$

and thus the equations of microscopic motion are, from (5-10) and (5-11),

$$\theta_1^+ \in \partial I(\beta_1^+), \quad \theta_2^+ \in \partial I(\beta_1^+). \tag{5-16}$$

The two identical pieces of ice before collision satisfy $\beta_1^- = \beta_2^- = 0$, $\theta_1^- \leq 0$, $\theta_2^- \leq 0$. We look for conditions such that ball 1 melts and ball 2 remains frozen:

$$\beta_1^+ = 1, \quad \beta_2^+ = 0. \tag{5-17}$$

Thus from (5-16), we have

$$\theta_1^+ \geq 0, \quad \theta_2^+ \leq 0. \tag{5-18}$$

The values of λ and C depend on the relative importances of the volumes and surface areas of the pieces of ice. We study the two cases, when λ is either small or large with respect to C (see Remark 5.4).

Suppose λ is small with respect to C/T_0 . Equations (3-1) and (3-12) give

$$P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) = C_1[\theta_1] + L[\beta_1] + C_2[\theta_2] + L[\beta_2],$$

$$\left(\frac{C_2}{T_0} + \frac{\lambda}{2} \right) [\theta_2] + \frac{L}{T_0} [\beta_2] - \left(\frac{C_1}{T_0} + \frac{\lambda}{2} \right) [\theta_1] - \frac{L}{T_0} [\beta_1] = -\lambda(\theta_2^- - \theta_1^-).$$

We suppose λ to be small with respect to C_i/T_0 , $i = 1, 2$ and we get

$$P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) = C_1[\theta_1] + L[\beta_1] + C_2[\theta_2] + L[\beta_2],$$

$$\frac{C_2}{T_0} [\theta_2] + \frac{L}{T_0} [\beta_2] - \frac{C_1}{T_0} [\theta_1] - \frac{L}{T_0} [\beta_1] = -\lambda(\theta_2^- - \theta_1^-),$$

which gives the system

$$C_1[\theta_1] = \frac{1}{2} \left(P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - 2L[\beta_1] + \lambda T_0(\theta_2^- - \theta_1^-) \right),$$

$$C_2[\theta_2] = \frac{1}{2} \left(P^{\text{int}} D \left(\frac{U^+ + U^-}{2} \right) - 2L[\beta_2] - \lambda T_0(\theta_2^- - \theta_1^-) \right),$$
(5-19)

where the volume fractions β_1 and β_2 satisfy the equations of microscopic motion (5-16).

Equations (5-19) give with $C_1 = C_2 = C$, and (5-17)

$$C[\theta_1] = \frac{1}{2} (\mathcal{T} - 2L + \lambda T_0(\theta_2^- - \theta_1^-)), \quad C[\theta_2] = \frac{1}{2} (\mathcal{T} - \lambda T_0(\theta_2^- - \theta_1^-)).$$
(5-20)

By means of (5-20), conditions (5-18) are satisfied if and only if

$$(2C - \lambda T_0)\theta_1^- + \lambda T_0\theta_2^- + \mathcal{T} - 2L \geq 0,$$
(5-21)

$$\lambda T_0\theta_1^- + (2C - \lambda T_0)\theta_2^- + \mathcal{T} \leq 0,$$
(5-22)

with $\theta_1^- \leq 0$ and $\theta_2^- \leq 0$. Because of our hypothesis on λ , we have

$$2C - \lambda T_0 > 0.$$
(5-23)

- If $\mathcal{T} < 2L$, inequalities (5-21) and (5-23) show that it is impossible to satisfy system (5-21), (5-22) with $\theta_1^- \leq 0$ and $\theta_2^- \leq 0$. Thus, if the dissipation is small, it is impossible to melt one piece of ice. Both remain frozen.
- If $\mathcal{T} \geq 2L$, it is possible to find temperatures (θ_1^-, θ_2^-) satisfying the system (5-21), (5-22) with $\theta_1^- \leq 0$, $\theta_1^- \approx 0$ ($\theta_1^- = 0$ if $\mathcal{T} = 2L$), and $\theta_2^- < \theta_1^-$. Thus, if the dissipation is large, one piece of ice melts, and the other one remains frozen. The temperature of the coldest piece of ice has to be sufficiently cold.

Examples are given in Figures 1 and 2 for $\lambda = 0$ and $\lambda = 100$, $\mathcal{T} = 10L$, with $L = 3.33 \times 10^5$, $C = 10^6$, and $T_0 = 273$.

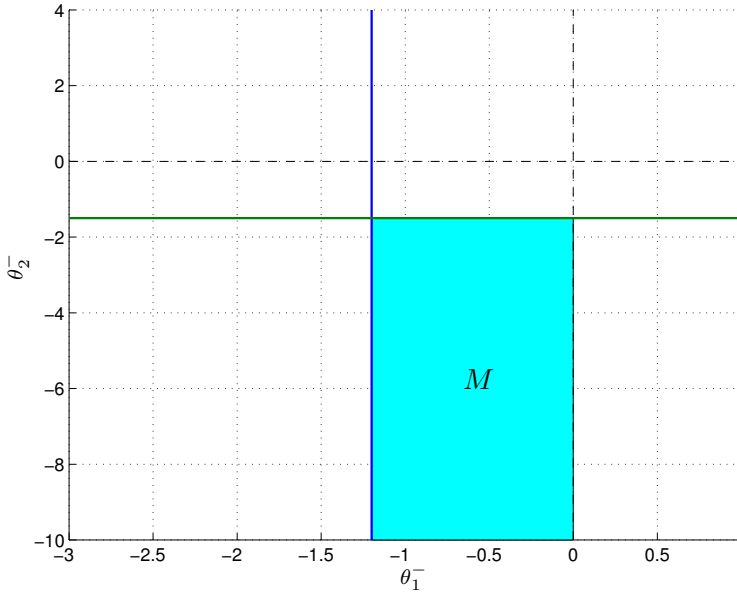


Figure 1. The case $\lambda = 0$ and $\mathcal{T} = 10L$ with $L = 3.33 \times 10^5$ and $C = 10^6$. The inequalities (5-21) and (5-22) have solutions if θ_1^- is negative and satisfies $\theta_1^- \geq -(\mathcal{T} - 2L)/2C$, equality on the blue line, and $\theta_2^- \leq -\mathcal{T}/2C$, equality on the green line. The point (θ_1^-, θ_2^-) has to belong to the set M .

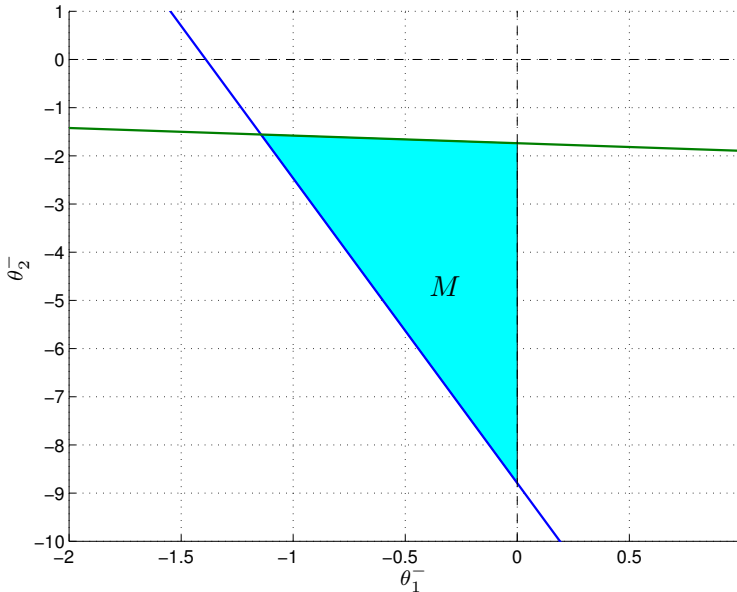


Figure 2. The case $\lambda = 1000$ and $\mathcal{T} = 10L$ with $L = 3.33 \times 10^5$ and $C = 10^6$. The inequalities (5-21) and (5-22) have solutions if point (θ_1^-, θ_2^-) belongs to the set M which is defined by the blue and green lines of (5-21), (5-22), and $\theta_1^- \leq 0, \theta_2^- \leq 0$.

Now suppose instead that λ is not small with respect to C/T_0 . We have $[\beta_1] = 1$ and $[\beta_2] = 0$, and (5-4) and (5-5) give

$$\mathcal{T} = C[\theta_1] + L + C[\theta_2], \tag{5-24}$$

$$C \frac{[\theta_2]}{T_0} - C \frac{[\theta_1]}{T_0} - \frac{L}{T_0} = -\lambda \left(\theta_2^- - \theta_1^- + \frac{[\theta_2]}{2} - \frac{[\theta_1]}{2} \right). \tag{5-25}$$

From (5-24), we get

$$[\theta_2] = \frac{1}{C} (\mathcal{T} - C[\theta_1] - L); \tag{5-26}$$

substituting (5-26) into (5-25), we obtain

$$\begin{aligned} \theta_1^+ &= \frac{2C}{2C + \lambda T_0} \theta_1^- + \frac{\lambda T_0}{2C + \lambda T_0} \theta_2^- + \frac{\mathcal{T} - L}{2C} - \frac{L}{2C + \lambda T_0}, \\ \theta_2^+ &= \frac{2C}{2C + \lambda T_0} \theta_2^- + \frac{\lambda T_0}{2C + \lambda T_0} \theta_1^- + \frac{\mathcal{T} - L}{2C} + \frac{L}{2C + \lambda T_0}. \end{aligned} \tag{5-27}$$

Conditions (5-18) are satisfied if and only if

$$\frac{2C}{2C + \lambda T_0} \theta_1^- + \frac{\lambda T_0}{2C + \lambda T_0} \theta_2^- + \frac{\mathcal{T} - L}{2C} - \frac{L}{2C + \lambda T_0} \geq 0, \tag{5-28}$$

$$\frac{2C}{2C + \lambda T_0} \theta_2^- + \frac{\lambda T_0}{2C + \lambda T_0} \theta_1^- + \frac{\mathcal{T} - L}{2C} + \frac{L}{2C + \lambda T_0} \leq 0. \tag{5-29}$$

Since $\theta_1^- \leq 0$ and $\theta_2^- \leq 0$, to satisfy (5-28) a necessary condition is

$$\frac{\mathcal{T} - L}{2C} \geq \frac{L}{2C + \lambda T_0}.$$

Thus the dissipation has to be large in order to melt one of the pieces of ice.

- If $\lambda < 2C/T_0$, the system (5-28), (5-29) has solutions (θ_1^-, θ_2^-) , θ_1^- , and $\theta_2^- \leq 0$, if

$$\frac{\mathcal{T} - L}{2C} \geq \frac{L}{2C - \lambda T_0} > \frac{L}{2C + \lambda T_0};$$

in agreement with the case λ negligible; see the left halves of Figures 3 and 4. If λ is small, only the mechanical effect warms the balls whereas the conduction has a negligible effect. We have $\theta_2^- \leq \theta_1^- \leq 0$ and $\theta_2^+ \leq 0 \leq \theta_1^+$. The warmest piece of ice melts in the collision and the coldest remains frozen.

- If $\lambda > 2C/T_0$, the system (5-28), (5-29) has solutions θ_1^- and $\theta_2^- \leq 0$ if

$$\frac{\mathcal{T} - L}{2C} \geq -\frac{L}{2C - \lambda T_0} > \frac{L}{2C + \lambda T_0},$$

see the right halves of Figures 3 and 4. If C is small, that is, if the heat capacity is negligible, it is difficult for the system to store energy (the only possibility for storing energy is with a phase change). Since we have assumed the system to be adiabatic, the heat has to remain in the system and very large temperature variations occur: the effect of conduction is added to the mechanical effect and it

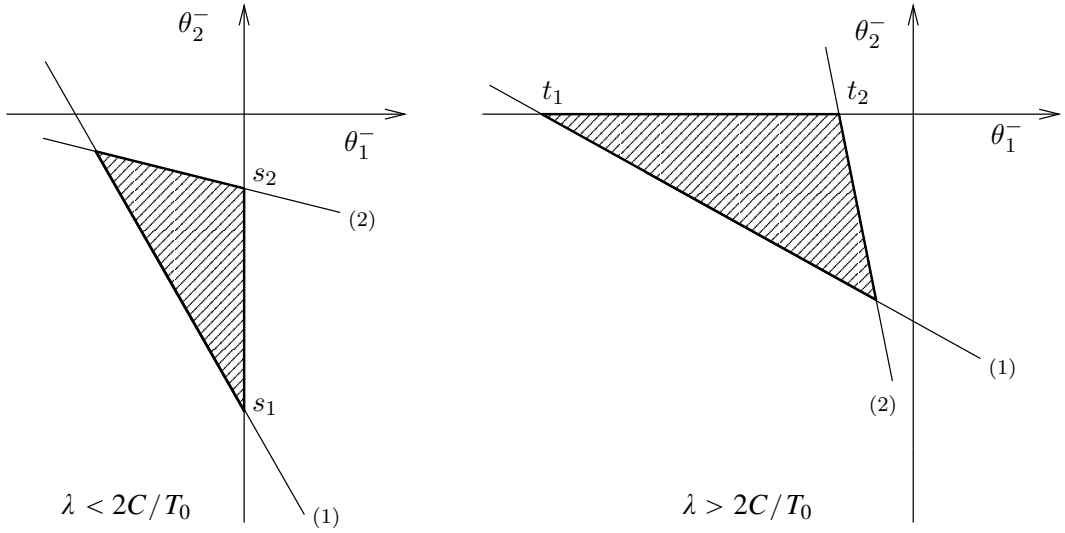


Figure 3. Condition (5-28), curve (1), and condition (5-29), curve (2). For ball 1 to be unfrozen after collision while ball 2 remains frozen, the temperatures θ_1^- and θ_2^- have to be in the hatched triangles.

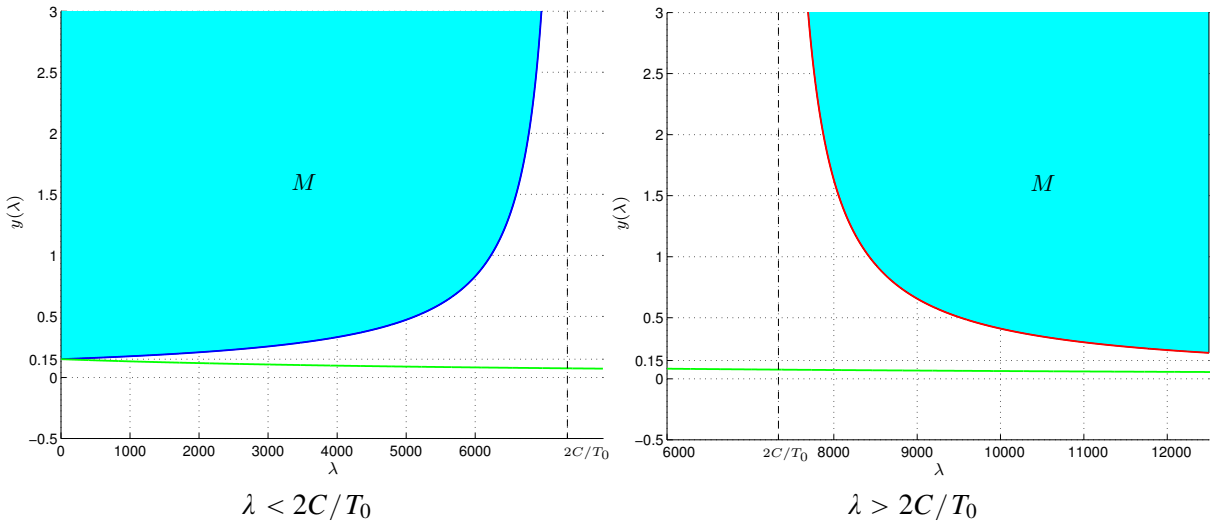


Figure 4. The curve $y_1(\lambda) = L/(2C - \lambda T_0)$ versus λ in blue, the curve $y_2(\lambda) = -y_1(\lambda)$ in red, and the curve $y_3(\lambda) = L/(2C + \lambda T_0)$ in green. The quantity $(\mathcal{F} - L)/2C$ has to be in the blue domain M for ball number 1 to melt. Thus the collision has to be violent enough, that is, \mathcal{F} large enough, for ball number 1 to melt.

increases the temperature of the coldest ball. Therefore, the coldest ball before collision becomes the warmest and vice versa. We have $\theta_1^- \leq \theta_2^- \leq 0$ and $\theta_2^+ \leq 0 \leq \theta_1^+$. In the extreme situation where $\lambda = \infty$, formulae (5-27) show that $\theta_1^+ = \theta_2^-$ and $\theta_2^+ = \theta_1^-$. When the thermal dissipation is very

large the temperatures exchange. Let us note that the same result holds for the velocities when the mechanical pseudopotential of dissipation,

$$\Phi^{\text{mech}}\left(D\left(\frac{U^+ + U^-}{2}\right), D\left(\frac{U^-}{2}\right)\right),$$

involves a quadratic potential

$$k\left(D\left(\frac{U^+ + U^-}{2}\right)\right)^2;$$

see for instance [Frémond 2001] or [Frémond 2007, page 51]. When the mechanical dissipation parameter $k = \infty$, the velocities exchange, $U_1^+ = U_2^-$, $U_2^+ = U_1^-$, if the two balls have same mass, see [Frémond 2001] or [Frémond 2007, page 53].

In the extreme situation where the heat capacity is zero, $C = 0$, the system cannot store energy except by changing phase. The energy L which is needed to melt the piece of ice has to be equal to the dissipated work \mathcal{T} and the discontinuities of temperature are opposite due to (5-24) extended by continuity.

- If $\lambda = 2C/T_0$, then, from (5-25), we have

$$\left(\frac{C}{T_0} + \frac{\lambda}{2}\right)(\theta_2^+ - \theta_1^+) = \frac{L}{T_0} + \left(\frac{C}{T_0} - \frac{\lambda}{2}\right)(\theta_2^- - \theta_1^-) = \frac{L}{T_0};$$

this entails $\theta_2^+ > \theta_1^+$, which forbids having $\theta_1^+ \geq 0 \geq \theta_2^+$. In this situation, conduction and heat storage have opposite effects and cancel each other.

Remark 5.4. The assumption $\lambda > 2C/T_0$ is often not realistic because λ is proportional to the contact surface of the two pieces of ice and C is proportional to the volume of the pieces of ice. To have $\lambda > 2C/T_0$, the contact surface has to be large compared to the volume — for instance, when the two pieces of ice are thin sheets. But in this case the contact surface with the atmosphere is also large in contradiction of the adiabatic assumption in the collision. Thus the assumption $\lambda > 2C/T_0$ is not consistent with the adiabatic collisions, that is, collisions without heat exchange with the exterior.

6. Does rain give black ice when falling on frozen ground?

6.1. An application of the previous results. We may consider as a simplifying approximation that the ground behaves like a ball of frozen water. Thus we use the previous results and assume that λ is small with respect to C/T_0 and that there is no dissipation with respect to $[\beta_1]$ and $[\beta_2]$:

$$\Phi_1([\beta_1]) = \Phi_2([\beta_2]) = 0.$$

A droplet of rain and the frozen ground before collision satisfy $\beta_1^- = 1$, $\beta_2^- = 0$, $\theta_1^- \geq 0$ and $\theta_2^- \leq 0$. We look for conditions such that the droplet of rain freezes, becomes black ice, and the ground remains frozen:

$$\beta_1^+ = 0, \quad \beta_2^+ = 0. \tag{6-1}$$

Thus from (5-16), we have

$$\theta_1^+ \leq 0, \quad \theta_2^+ \leq 0. \tag{6-2}$$

We assume that the collision is adiabatic and, for consistency, that λ is small with respect to C/T_0 . Equations (5-6) and (5-7), with $C_1 = C_2 = C$, and using (6-1), give

$$C[\theta_1] = \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)), \quad C[\theta_2] = \frac{1}{2}(\mathcal{T} - \lambda T_0(\theta_2^- - \theta_1^-)).$$

By means of (5-20), conditions (6-2) are satisfied if and only if

$$(2C - \lambda T_0)\theta_1^- + \lambda T_0\theta_2^- + \mathcal{T} + 2L \leq 0, \quad \lambda T_0\theta_1^- + (2C - \lambda T_0)\theta_2^- + \mathcal{T} \leq 0, \quad (6-3)$$

with $\theta_1^- \geq 0$ and $\theta_2^- \leq 0$. Because of the hypothesis $2C - \lambda T_0 > 0$, it is always possible to satisfy conditions (6-3) by having θ_2^- negative enough, that is, by having the ground very cold. The maximum value of θ_2^- is given by $\theta_1^+ = 0$. It is

$$\theta_2^- = -\frac{\mathcal{T} + 2L}{\lambda T_0} - \frac{2C - \lambda T_0}{\lambda T_0}\theta_1^-, \quad \text{or} \quad \theta_2^- = -\frac{\mathcal{T} + 2L}{\lambda T_0} - \frac{2C}{\lambda T_0}\theta_1^-,$$

because λ is small with respect to C/T_0 . As may be expected, the maximum value of θ_2^- is negative and decreasing when the dissipated work \mathcal{T} is increasing and when θ_1^- is increasing [Caucci and Frémond 2007; Caucci 2006].

6.2. Another assumption: the ground is massive and its temperature remains constant. We assume the ground temperature remains constant in the collision. We get this property if the heat impulse capacity C_2 is very large compared to C_1 . Thus we let $C_2 = \infty$ in formulas (5-6) and (5-7) and get the equations

$$C_1[\theta_1] = \frac{1}{2}\left(P^{\text{int}}D\left(\frac{U^+ + U^-}{2}\right) - 2L[\beta_1] + \lambda T_0\left(\theta_2^- - \theta_1^- - \frac{[\theta_1]}{2}\right)\right),$$

and $\theta_1^+ \in \partial I(\beta_1^+)$ — see (5-10) — assuming no dissipation with respect to $[\beta_1]$. They give

$$\left(C_1 + \frac{\lambda T_0}{4}\right)[\theta_1] + L[\beta_1] = \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)),$$

and

$$\left(C_1 + \frac{\lambda T_0}{4}\right)\theta_1^+ + L\partial I^*(\theta_1^+) \ni \left(C_1 + \frac{\lambda T_0}{4}\right)\theta_1^- + \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)),$$

where $I^*(\theta_1^+) = pp(\theta_1^+)$ is the dual function of I ($pp(x)$ is the nonnegative part of x). Assuming λ is small with respect to C/T_0 , we have

$$C_1\theta_1^+ + L\partial I^*(\theta_1^+) \ni C_1\theta_1^- + \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)),$$

with the following solutions:

(1) If

$$C_1\theta_1^- + \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)) \leq 0, \quad \theta_1^+ \leq 0, \quad \theta_2^- \leq \frac{(\lambda T_0 - 2C_1)}{\lambda T_0}\theta_1^- - \frac{(\mathcal{T} + 2L)}{\lambda T_0},$$

the droplet freezes. Its temperature is

$$\theta_1^+ = \theta_1^- + \frac{1}{2C_1}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)).$$

(2) If $0 \leq C_1\theta_1^- + \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-)) \leq L$, the droplet freezes partially at 0°C :

$$\theta_1^+ = 0.$$

(3) If $L \leq C_1\theta_1^- + \frac{1}{2}(\mathcal{T} + 2L + \lambda T_0(\theta_2^- - \theta_1^-))$, the droplet does not freeze. Its temperature becomes

$$\theta_1^+ = \theta_1^- + \frac{1}{2C_1}(\mathcal{T} + \lambda T_0(\theta_2^- - \theta_1^-)).$$

Note that the conditions for the droplet to freeze are similar under the two assumptions. The occurrence of black ice on roads due to rain falling on deeply frozen ground is predicted by this theory.

6.3. A more realistic assumption: the temperature of the ground is not uniform after collision. Under the two previous assumptions, the temperature of the ground is uniform after collision. It is clear that there is a local increase of the temperature where the droplet hits the ground but the temperature remains constant at some distance. Thus an interesting assumption is that the temperature of the ground is not uniform after collision: the temperature discontinuity $[\theta_2]$ becomes a function of x . This predictive theory is given in [Frémond 2001]. It is better but it has the disadvantage that closed form solutions are not available.

7. Conclusions

The thermomechanical theory of collisions involving phase change we have investigated is in agreement with everyday experiments.

We considered the collision of two pieces of ice. When they have different temperatures and collide, the dissipation due to the collision may be large enough to melt the warmest of them. We looked for conditions on the state quantities before the collision and on the dissipated work such that this phenomenon occurs. We showed that, if the dissipation is large, one piece of ice melts and the other remains frozen. For this to happen, the temperature of the coldest piece of ice has to be sufficiently cold.

We also studied the problem of warm rain falling on frozen ground, asking whether the rain freezes or the frozen ground thaws. We proved that the droplet of rain freezes if the ground is very cold. For this to occur, the temperature of the ground has to be lower than a maximum temperature, which decreases as the dissipated work and the temperature of the rain increase.

References

- [Caucci 2006] A. M. Caucci, *Modelli matematici per transizioni di fase*, Ph.D. Thesis, University of Bologna, 2006. Also to appear in *Boll. Unione Mat. Ital.*
- [Caucci and Frémond 2007] A. M. Caucci and M. Frémond, “Phase change and collisions”, in *Free boundary problems: theory and applications* (Coimbra, 2005), edited by I. N. Figueiredo et al., International Series of Numerical Mathematics **154**, Birkhäuser, Basel, 2007.
- [Frémond 2001] M. Frémond, *Non-smooth thermomechanics*, Springer, Berlin, 2001.
- [Frémond 2005] M. Frémond, “Phase change in mechanics”, Lecture notes, 30th Summer School in Mathematical Physics, Ravello, 2005. To be published by the Gruppo Nazionale per la Fisica Matematica of the INdAM, Rome.
- [Frémond 2007] M. Frémond, *Collisions*, Edizioni del Dipartimento di Ingegneria Civile dell’Università di Roma Tor Vergata, Rome, 2007.

[Frémond and Visintin 1985] M. Frémond and A. Visintin, “Dissipation dans le changement de phase. Surfusion. Changement de phase irréversible”, *C. R. Acad. Sci. II B Mec.* **301** (1985), 1265–1268.

[Moreau 1966–1967] J. J. Moreau, “[Fonctionnelles convexes](#)”, Séminaire Jean Leray aux équations différentielles partielles, 2, 1966–1967, Available at http://www.numdam.org/item?id=SJL_1966-1967__2_1_0. Also available from the Edizioni del Dipartimento di Ingegneria Civile dell’Università di Roma Tor Vergata, Rome, 2003.

[Moreau 1970] J. J. Moreau, “Sur les lois de frottement, de viscosité et de plasticité”, *C. R. Acad. Sci. Paris* **271** (1970), 608–611.

Received 20 Dec 2007. Revised 2 Nov 2008. Accepted 8 Nov 2008.

ANNA MARIA CAUCCI: caucci@dm.unibo.it

Università di Bologna, Dipartimento di Matematica, Piazza di Porta S. Donato, 5, 40126 Bologna, Italy

MICHEL FRÉMOND: michel.fremond@lagrange.it

Università di Roma Tor Vergata, Laboratorio Lagrange, Dipartimento di Ingegneria Civile, Via del Politecnico, 1, 00133 Roma, Italy

FIELDS OF STORED ENERGY ASSOCIATED WITH LOCALIZED NECKING OF STEEL

ANDRÉ CHRYSOCHOOS, BERTRAND WATTRISSE,
JEAN-MICHEL MURACCIOLE AND YVES EL KAÏM

This paper describes an experimental procedure for the simultaneous determination of heat sources and mechanical energy involved locally during a heterogeneous tensile test. This procedure involves two complementary imaging techniques: digital image correlation (DIC) and infrared thermography (IRT). The first technique gives displacement fields from which strains are derived while the second provides temperature fields with which the heat sources are estimated using a local form of the heat equation. Moreover, a method based on integration of equilibrium equations under the plane stress assumption is used to determine the stress distribution during the test. The distribution of the local deformation energy developed by the material is then assessed using stress and strain-rate fields.

Tensile tests were performed on thin flat steel samples. The results revealed early and gradual development of strain localization within the gauge part of the specimen. Energy balances were performed inside and outside the necking zone based on the assumption that the thermoelastic part of the behaviour remains linear and isotropic. Finally, indirect estimate of the stored energy led us to compute the time course of the local Taylor–Quinney coefficient.

Introduction

The conversion of mechanical energy into heat has been investigated in a wide range of materials by many researchers, including [Taylor and Quinney 1934; Schmidt et al. 1945; Chrysochoos and Martin 1989; Rittel 1999]. Using different experimental arrangements (calorimeter [Shenogin et al. 2002], thermocouples [Zehnder et al. 1998], IR sensors [Guduru et al. 2001; Chrysochoos and Louche 2000]), these studies gave similar results, showing that a variable amount of mechanical energy is converted into heat during inelastic transformation. Such techniques generally provide a macroscopic estimate of the Taylor–Quinney coefficient that links mechanical and dissipated energies. Here *macroscopic* means at the scale of the sample gauge part. This coefficient plays a key role in the modelling of plasticity each time the dissipated energy has to be introduced within a pure mechanical elastic-plastic framework. It has been used to compute the plastic strain-induced heat without having to invoke a particular thermodynamic framework [Batra and Chen 2001; Campagne et al. 2005; Rusinek et al. 2007]. Moreover, we stress that this coefficient is nearly always regarded as a constant material parameter, independent of the loading path and strain hardening history. In such particular cases, the stored energy ratio is then equal to the ratio of the stored energy rate, as we will see hereafter. This latter ratio often appears in the heat equation associated with anisothermal viscoplastic models developed to describe dynamic localization mechanisms [Mercier and Molinari 1998; Rosakis et al. 2000]. Nevertheless, in the case of heterogeneous

Keywords: digital image correlation, infrared thermography, dissipation, stored energy, plasticity, localization.

loading (necking, shear bands, etc.), energy fields are then both representative of the material behaviour combined with structure effects. Under such conditions, the overall Taylor–Quinney coefficient can no longer be only representative of the material behaviour.

In this paper, we present a specific setup for local estimate of different terms of the energy balance for quasistatic, heterogeneous tests. This setup combines two imaging techniques, that is, digital image correlation (DIC) and infrared thermography (IRT). The experimental design enables us to estimate the local distribution of deformation energy rate developed by the material and heat sources induced by the loading. DIC techniques are now commonly used to measure displacement and strain fields on the surface of thin specimens [Chu et al. 1985]. Besides, IRT camera performances have been considerably improved with the advent of infrared focal plane array sensors (IRFPA). These new cameras provide IR films with fine spatial resolution, low thermal noise, and reasonably high frame rates for quasistatic mechanical tests.

In what follows, we first review the theoretical background used to define the energy balance. We then describe the different devices involved in the setup and experimental procedure. In the third part, we detail the data processing procedures and their validation. Finally, we present and discuss some results obtained during straining of an interstitial-free steel (commercial grade: IF–Ti).

Energy balance and heat equation

To draw up the energy balance, we worked within the classical formalism of generalised standard materials [Halphen and Nguyen 1975] which are sometimes used to deal with material behaviours [Lemaitre and Chaboche 1985], particularly in plasticity [Lubliner 1991]. In this framework, the thermodynamic state of each volume element of the material is characterised by a finite set of variables. In the case of plasticity, this set includes the absolute temperature T , a strain tensor denoted by ϵ , and a vector α whose components characterize the microstructural state of the material. The chosen thermodynamic potential associated with such a state variable set is the Helmholtz free energy $\psi(T, \epsilon, \alpha)$.

The dissipation is a volume heat source associated with irreversible processes induced by deformation mechanisms and heat diffusion. Its definition is classically derived from the local expression of the 2nd principle of thermodynamics defining the irreversible entropy source. As usual, we suppose that the intrinsic (mechanical) dissipation d_1 and the thermal dissipation d_2 are separately positive. With the chosen set of state variables, d_1 is defined by

$$d_1 = \sigma : D - \rho \frac{\partial \psi}{\partial \epsilon} : \dot{\epsilon} - \rho \frac{\partial \psi}{\partial \alpha} \cdot \dot{\alpha} \geq 0, \quad (1)$$

where ρ is the mass density, σ is the Cauchy stress tensor, and D the Eulerian strain-rate tensor. The superimposed dot represents the time derivative. Naturally, the equality $d_1 = 0$ corresponds to mechanically reversible processes. Note that the intrinsic dissipation d_1 is the difference between the deformation energy rate w_{def}^\bullet , and the sum of the elastic w_e^\bullet and stored w_s^\bullet energy rates

$$w_{\text{def}}^\bullet = \sigma : D, \quad (2)$$

$$w_e^\bullet + w_s^\bullet = \rho \frac{\partial \psi}{\partial \epsilon} : \dot{\epsilon} + \rho \frac{\partial \psi}{\partial \alpha} \cdot \dot{\alpha}, \quad (3)$$

In Equations (2) and (3), the notation $()^\bullet$ means that the variation of $()$ is path-dependent. These different energies are therefore not state functions *a priori*. We define the corresponding energy variations by

$$w_x = \int_{t_0}^t w_x^\bullet d\tau,$$

where x symbolizes the energy type (for example, deformation, elastic, stored) while t_0 and t are respectively the times of initial and current states. Similarly, the dissipated energy variation is defined by

$$w_d = \int_{t_0}^t d_1 d\tau.$$

Temperature variations within the sample are governed by the heat equation. Combining the first and second principles of thermodynamics, we derive its following local expression

$$\rho C \dot{T} + \text{div} q = d_1 + r_e + w_{\text{tec}}^\bullet + w_{\text{tmc}}^\bullet, \quad (4)$$

where C stands for the specific heat capacity at constant ϵ and α , while q is the heat influx vector. Assuming a Fourier's law of heat conduction ($q = -k \text{ grad } T$, where k is the constant isotropic conduction coefficient), the left hand side of Equation (4) becomes a partial derivative operator applied to the temperature field T .

Heat sources responsible for temperature variations within the specimen are gathered on the right hand member of Equation (4). They correspond to the intrinsic dissipation d_1 , the external volume heat supply r_e , and thermomechanical couplings w_{tec}^\bullet and w_{tmc}^\bullet . The term w_{tec}^\bullet represents the famous thermoelastic coupling (that is, Lord Kelvin's term), and w_{tmc}^\bullet represents all other possible thermomechanical couplings associated with interactions between the temperature and microstructure. In the general case, the sum $w_{\text{tec}}^\bullet + w_{\text{tmc}}^\bullet$ reads

$$w_{\text{tec}}^\bullet + w_{\text{tmc}}^\bullet = \rho T \frac{\partial^2 \psi}{\partial T \partial \epsilon} : \dot{\epsilon} + \rho T \frac{\partial^2 \psi}{\partial T \partial \alpha} \cdot \dot{\alpha}.$$

In the following, we will assume that the thermoelastic part of the behaviour remains linear and isotropic. We will also neglect all other thermomechanical couplings. This latter hypothesis is justified by the fact that temperature variations remain relatively small, thus inducing no microstructure variation (no self-induced annealing). We thus introduce w_{h}^\bullet as the overall heat source defined by $w_{\text{h}}^\bullet = d_1 + w_{\text{tec}}^\bullet$ and verifying

$$\rho C \dot{\theta} - k \Delta \theta = w_{\text{h}}^\bullet + r_e \quad (5)$$

with θ being the local temperature difference $T - T_0$, where T_0 is the current room temperature.

In this setting, the stored energy ratio can then be defined by

$$F_w = \frac{w_s}{w_{\text{in}}} = \frac{w_{\text{def}} - w_e - w_d}{w_{\text{def}} - w_e} = \frac{w_s}{w_s + w_d}.$$

The difference $w_{\text{def}} - w_e$ represents the inelastic work w_{in} . In the case of plastic hardening at finite strain, the elastic energy generally remains very low relative to the deformation energy so that $F_w \approx 1 - w_d/w_{\text{def}}$.

As often supposed in the literature, the stored energy ratio is considered as a constant regardless of the initial hardening state and the loading path (typically $F_w = 0.1$). In this particular case, the ratio of the stored energy rate is also constant and equal to the stored energy ratio itself

$$\dot{F}_w = \frac{d}{dt} \left(\frac{w_s}{w_{in}} \right) = \frac{w_{in} \dot{w}_s - w_{in}^{\bullet} w_s}{w_{in}^2} = 0 \Rightarrow F_w^{\bullet} = \frac{\dot{w}_s}{w_{in}^{\bullet}} = F_w. \quad (6)$$

Conversely, if the stored energy ratio changes, it is no longer equal to the stored energy rate ratio.

Experimental procedure

The experimental tests involved performing, at room temperature, displacement-controlled tensile loading at constant velocity ($v_0 = 250 \mu\text{m s}^{-1}$). The following constant dimensions were allocated to the gauge part of the specimen: the gauge length L_0 and depth D_0 were 50 and 0.3 mm, respectively. Conversely, the width $W_0(X)$ might depend on the longitudinal Lagrangian coordinate X to force, for example, the development of the strain localization around a given cross-section. We chose $W_0 = 12.5$ mm for samples with constant W_0 . The depth was set at a particularly low measurement to ensure the plane stress hypothesis and legitimate the hypothesis of uniform strain in the depth direction. A simultaneous record of infrared and visible images was performed on each side of the sample surface $L \times W$ during tensile straining. [Figure 1](#) illustrates the experimental set-up designed for this purpose. It involves a MTS hydraulic testing machine (frame: 100 kN, load cell: 25 kN), a Cedip Jade III infrared camera and a Camelia 8M high resolution CCD camera. The optical axis of both cameras was set perpendicularly to the frame of the testing machine, and it remained fixed during the test.

The main characteristics of the two cameras are given in [Table 1](#). The chemical composition of the material tested is given in [Table 2](#).

Each camera was controlled by a separate computer. A specific electronic device was designed to synchronise the frame grabbing of the two cameras. The principle of this device is as follows. A frequency generator is used to produce the trigger signal of the master camera. This frequency is divided or multiplied by an integer factor to generate the trigger signal of the slave camera. Each time an image acquisition is completed, the analogical signals provided by the machine sensors are digitised, and the

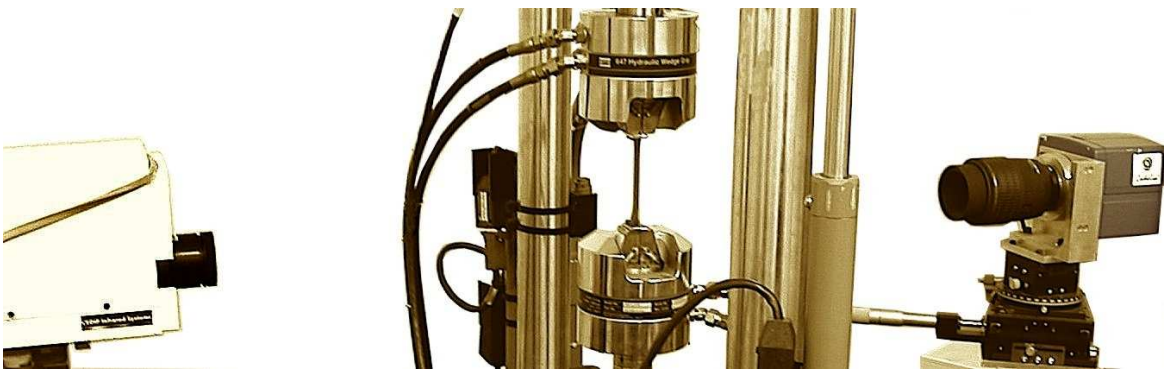


Figure 1. Experimental set-up.

| | image size (pix) | scale factor ($\mu\text{m pix}^{-1}$) | frame rate (Hz) |
|--------------------------------|------------------|--|-----------------|
| IR: Cedip Jade III | 320×240 | 524 | 25 |
| CCD: Camelia 8M (binning: 4×4) | 875×575 | 136 | 5 |

Table 1. Main camera characteristics.

time is given, on each acquisition computer, by a common clock having a 0.1 ms period. Using this device, we estimated that the synchronisation error between the two cameras was less than 0.05 ms.

Data processing

This section briefly reviews the numerical processing principles developed to determine the different energy fields. We focus particularly on the validation procedure improvements to check the data processing.

Visible images. The surface of the specimen observed by the visible CCD camera is speckled with white paint in order to obtain a random pattern defining the local optical signature of each material surface element (MSE). A classical digital image correlation algorithm allowed us to determine the in-plane components of the displacement field on a regular rectangular grid. The in-plane velocity and strain components were derived from the displacement data by a numerical differentiation method based on a local polynomial approximation of the displacement field [Wattrisse et al. 2001a]. The local time fitting of displacement fields involves a 2nd order polynomial while the local space fitting is associated with a coupled 1st order polynomial of the two in-plane coordinates. The choice of the approximation zone *AZ* is very important in the differentiation process. The optimized *AZ* depends on the signal-to-noise ratio and the amplitude of the sought derivatives. The image processing remains then relevant as long as the localization zone is greater than *AZ*. The camera resolution led us to choose centred *AZ* spanning around ± 2.5 s by ± 3.5 mm by ± 3.5 mm. In such conditions, the incertitude on strain measurements was estimated at 5×10^{-4} [Wattrisse et al. 2001a].

Using the kinematical data obtained by the DIC algorithm, we constructed the local stress distribution, assuming a quasistatic, plane stress, isochoric transformation. For each acquisition time *t*, corresponding to an applied load *F(t)*, the tensile component of the stress tensor σ_{xx} was assumed to be homogeneously distributed over each cross-section $S(X, t) = W(X, t) \times D(X, t)$ of the specimen

$$\sigma_{xx}(X, t) = \frac{F(t)}{S_0(X) \cdot \exp(-\varepsilon_{xx}(X, t))} \tag{7}$$

| | C | Mn | P | S | Si | Al | N | Ti |
|-------|-------|------|-------|-------|-------|------|-------|------|
| % (w) | 0.003 | 0.15 | 0.007 | 0.007 | 0.007 | 0.02 | 0.003 | 0.06 |

Table 2. Chemical composition of the tested steel [Béranger et al. 1994].

where ϵ henceforth represents the Hencky strain, derived from displacement fields after a standard polar decomposition of the transformation gradient tensor. The independence of ϵ_{xx} from the Y coordinate (that is, the width direction) was confirmed by the experimental observations while the independence from the Z coordinate (that is, the depth direction) was legitimated by the small sample thickness. The two other stress components (shear: σ_{xy} and contraction: σ_{yy}) were computed by integrating the momentum equations [Wattrisse et al. 2001b]

$$\sigma_{xy}(X, Y, t) = -\frac{\partial \sigma_{xx}(x(X, Y, t), t)}{\partial x} y(X, Y, t),$$

$$\sigma_{yy}(X, Y, t) = \frac{\partial^2 \sigma_{xx}(x(X, Y, t), t)}{\partial x^2} \cdot \left(\frac{y(X, Y, t)^2}{2} - \frac{W_0^2(X) \exp(-\epsilon_{xx}(X, t))}{8} \right).$$

The deformation energy locally developed by the material $w_{\text{def}}^\bullet(X, Y, t)$ was then estimated using Equation (2)

$$w_{\text{def}}^\bullet = \sigma_{xx} D_{xx} + 2\sigma_{xy} D_{xy} + \sigma_{yy} D_{yy}.$$

Thermal images. The infrared camera records the thermal radiations of the observed scene. Using the pixel calibration protocol described in [Honorat et al. 2005], we deduced the temperature variations of the specimen induced by the mechanical loading. To filter thermal data, local least-squares approximation of temperature fields was performed using the same set of polynomials as the one already used for visible images. In standard conditions, we estimated that the peak-to-peak thermal noise was about 200 mK (that is, before data filtering) and the range of the thermal noise dropped to 20 mK for standard filtering parameters. Moreover, the order of magnitude of the spatial resolution (in terms of pixel size) was about 0.4 mm and the temporal resolution was considered to be equal to the 0.04 s.

By integrating the heat Equation (4) over the depth of the sample [Chrysochoos and Louche 2000], and defining the mean thermal disequilibrium over the thickness between the sample and its surroundings by $\Theta = \theta - \theta^{\text{ref}}$, we obtained the following 2D differential equation:

$$\rho C \dot{\Theta} - k \left(\frac{\partial^2 \Theta}{\partial x^2} + \frac{\partial^2 \Theta}{\partial y^2} \right) + \rho C \frac{\Theta}{\tau_{\text{th}}^{2\text{D}}} = w_{\text{h}}^\bullet. \quad (8)$$

The external heat supply r_e defined in Equation (4) is here taken into account by monitoring the uniform temperature variations $\theta^{\text{ref}} = T^{\text{ref}} - T_0^{\text{ref}}$ of an unloaded reference specimen of the same geometry placed near the specimen in the field of view of the IR camera

$$r_e = \rho C \dot{\theta}^{\text{ref}} + \frac{\theta^{\text{ref}}}{\tau_{\text{th}}^{2\text{D}}}.$$

The parameter $\tau_{\text{th}}^{2\text{D}}$ represents a time constant characterising heat losses by convection and radiation between the sample surfaces and the surroundings, and $\dot{\Theta} = \frac{\partial \Theta}{\partial t} + v \cdot \text{grad} \Theta$ is the particular time derivative of Θ , v representing the velocity vector.

Table 3 presents the different thermophysical parameters used in the heat source computations.

An overall estimate of the incertitude on heat sources should take into account every possible error sources associated with: (i) temperature accuracy, (ii) knowledge of the thermophysical parameters, (iii) relevance of the thermal modelling (heat exchanges, source distribution), and its identification, (iv)

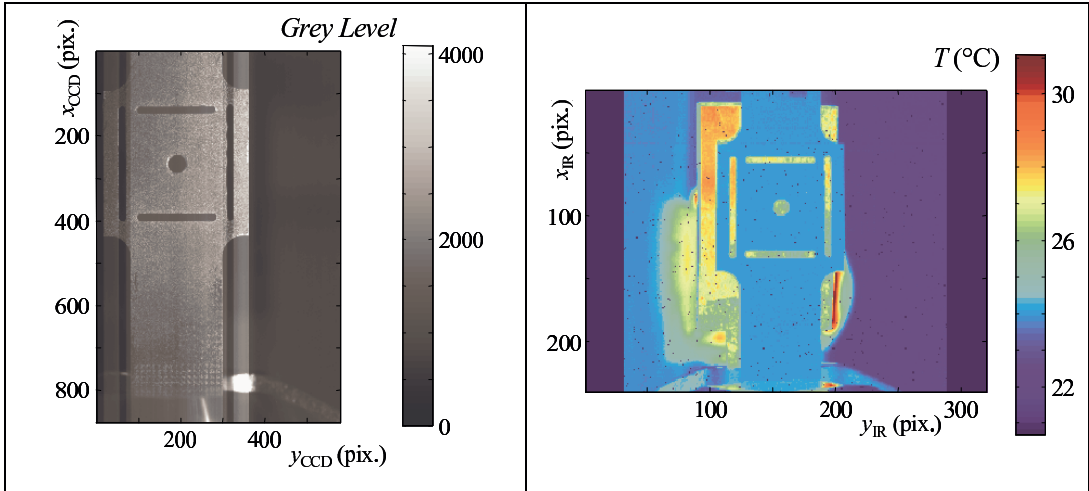


Figure 2. Calibration target: visible image (left); IR image (right).

mapping between kinematic and thermal data, (v) image processing robustness (derivation of discrete noisy temperature fields). Limiting the error analysis to items (i) and (v), we estimated the incertitude by computing the heat sources distribution from set of IR images of a nonloaded specimen. The parasitic sources derived from the image processing then gave an order of magnitude of the relative incertitude on the heat source fields. We found about a mean value of parasitic heat source of about $1.8 \times 10^{-5} \text{ W mm}^{-3}$ for a standard deviation of $2.7 \times 10^{-5} \text{ W mm}^{-3}$.

In the case of localized flow, the time derivative should take the convective term $v \cdot \text{grad } \Theta$ induced by the material flow into account. Once more, in accordance with the plane stress assumption, the smallness of the sample thickness and the high thermal diffusivity of the tested material, the depth-wise averaged temperatures were assumed to remain close to the surface temperatures. This enabled us to compute the convective term using the kinematic and thermal data.

Reference speckle and infrared images were mapped using a calibration target (see [Figure 2](#)). Comparing the visible and infrared images of the target, we determined the rigid body movements and the scale factor ratios between the two cameras.

For each acquisition time, the thermal data given by the IR camera (measured in the current, deformed configuration) were linearly interpolated spatiotemporally using the positions of the deformed configuration given by the DIC computation. This operation allowed us to track material particles associated with the DIC mesh, and it thus enabled us to compute temperature variations in the Lagrangian configuration. [Figure 3](#) illustrates the distribution of temperature variations after 50 s of loading. [Figure 3](#) (left) gives

| ρ (kg m^{-3}) | C ($\text{J kg}^{-1} \text{K}^{-1}$) | k ($\text{W m}^{-1} \text{K}^{-1}$) | $\lambda 10^6$ (K^{-1}) | T_0^{ref} (K) | $\tau_{\text{th}}^{2\text{D}}$ (s) |
|-------------------------------|--|---|------------------------------------|------------------------|------------------------------------|
| 7800 | 480 | 60 | 12.5 | 293 | 32 |

Table 3. Thermophysical parameters.

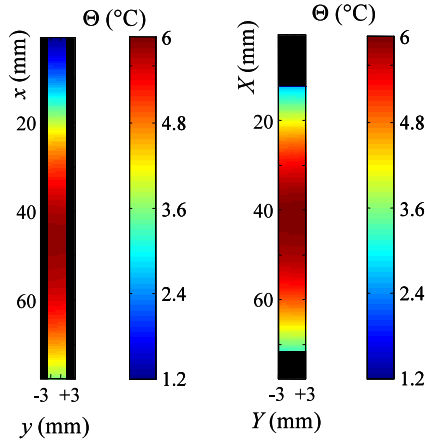


Figure 3. Thermal images represented in the current configuration (O, x, y) (left) and in the reference configuration (O, X, Y) (right).

the temperature distribution in the current (Eulerian) configuration, while Figure 3 (right) shows the temperature distribution brought back in the reference (Lagrangian) configuration. Both figures are associated with a material zone, initially of 6 mm width, centred on the gauge part of the specimen.

To illustrate the temperature patterns throughout the test, we plotted, in a single diagram, the time course of the temperature longitudinal profile captured in the middle of the sample width ($y = Y = 0$ mm). Figure 4 represents variations in the profile throughout the test, in the Eulerian configuration ($\theta(x, y = 0, t)$, Figure 4, left) and in the Lagrangian configuration ($\theta(X, Y = 0, t)$, Figure 4, right). In these figures, the horizontal axis represents the time while the vertical axis represents the sample longitudinal axis in the current configuration (x in Figure 4 (left)) or in the initial configuration (X in Figure 4 (right)). A conventional stress versus time curve was also superimposed in order to link the local thermal data to the overall mechanical loading. The paths of three MSE (named **A**, **B** and **C**) were plotted to illustrate the material flows. Element **A** is quite specific as it is the fracture point. Naturally, as material particles remain fixed in the Lagrangian configuration, their paths are simple horizontal lines in Figure 4 (right).

We can observe that the level curves in Figure 4 (left) appear to be noisier than those of Figure 4 (right). This is simply due to the fact that the Eulerian representation was here constructed without any temperature filtering (crude data) unlike the Lagrangian one.

The temperature time and spatial partial derivatives were then computed using a local polynomial fitting technique [Moreau et al. 2004]. Naturally, the spatial derivation was performed with respect to the current deformed state. In the Lagrangian configuration, the particular derivative of the temperature is equal to the partial time derivative $\dot{\Theta} = \partial\Theta/\partial t(X, Y = 0, t)$ and can thus be easily computed. Figure 5 (top left) gives the particular time derivative of the temperature during the test, while Figure 5 (top right) shows changes in the convective term $v(x, y = 0, t) \cdot \text{grad } \Theta(x, y = 0, t)$. This latter term can here reach up to 50% of the particular derivative and thus cannot be neglected in the time derivation. To more easily understand the particular distribution of the convective terms during strain localization, the longitudinal

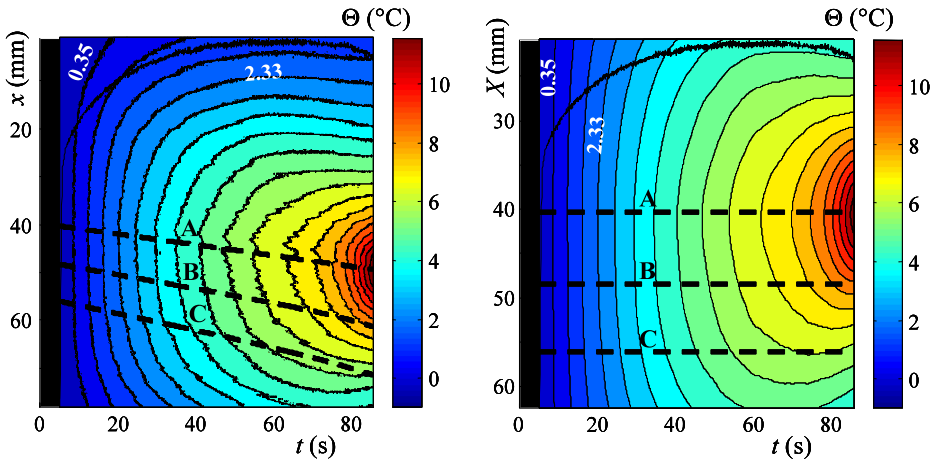


Figure 4. Temperature profile variations during a tensile test performed on a IF–Ti steel represented ($\delta\theta$ between two consecutive level curves: 0.66 K). The deformed configuration (O, x, t) (left); and the reference configuration (O, X, t) (right).

thermal gradients $\partial\Theta/\partial x(X, Y = 0, t = 80)$ and the longitudinal velocity profile $v_x(X, Y = 0, t = 80)$ were plotted in [Figure 5](#) (bottom) as functions of the Lagrangian coordinate X . The strain localization zone is characterized by a high strain rate in the necking region and inversion of the thermal gradient induced by the combined effect of the concentration of heat sources and heat diffusion.

Data processing validation

DIC and IRT image processing algorithms have been widely presented and checked in previous works [[Wattrisse et al. 2001a](#); [Chrysochoos and Louche 2000](#)]. In what follows, we tested the reliability of the energy balance construction by comparing the overall heat sources w_h^\bullet with the mechanical energy rate w_{def}^\bullet developed by the material. Indeed, as the elastic deformation energy and the heat induced by thermoelastic coupling remain small in plasticity, the plastic work and dissipated energy must be of the same order of magnitude.

Hereafter, the validation mainly deals with the local deformation energy. As it is experimentally impossible to impose a heterogeneous distribution of mechanical energy on a structure, we chose to check the image processing through numerical tests. A displacement-controlled tensile test was simulated using a finite element code (Cast3M) and a Prager elastoplastic model with linear kinematic hardening. We performed a three-dimensional computation in order to account for the triaxiality effect in the development of the neck. Furthermore, to obtain localized stress and strain patterns consistent with the development of necking, we used an initial geometry corresponding to an already necked specimen. To facilitate the three-dimensional FE computation (no need of remeshing due to mesh distortion), we chose to simulate the straining on a thicker specimen of about 2 mm (to be compared with the 0.3 mm of the real specimens). We were thus able to reproduce localized flow using a simple and thus easily identifiable homogeneous model. [Figure 6](#) shows the initial geometry of the specimen used in the computation, measured by a

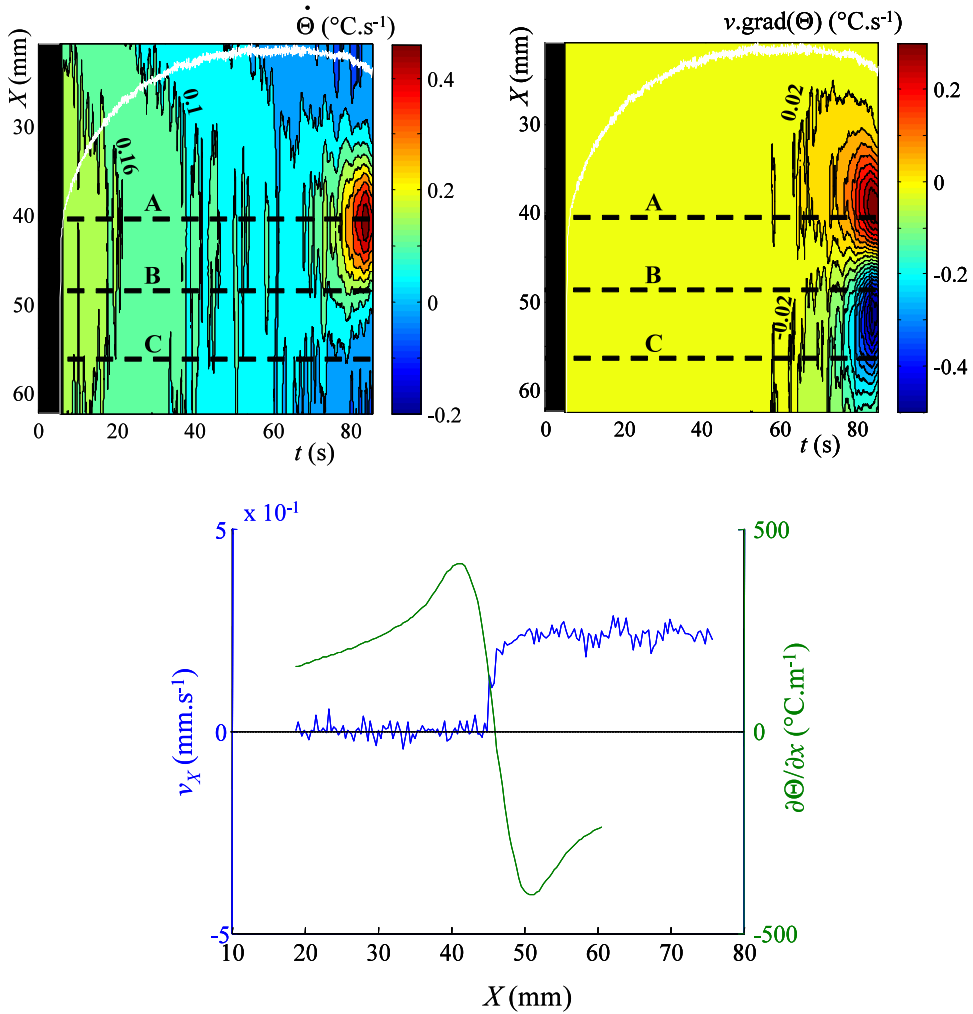


Figure 5. Time derivation of the temperature field: particular derivative (0.06 K s^{-1}) between two level curves (top left); convective term (0.04 K s^{-1}) between two level curves (top right); and Ox profiles of the longitudinal thermal gradient and velocity at $t = 80$ s (bottom).

3-axis measurement machine. It clearly highlights the presence of the neck in the middle of the sample $(X, Y) = (0, 0)$.

The material properties used in the FE computation are given in Table 4. The upper side of the sample was fixed and a vertical displacement was imposed on its lower side. We deduced the mechanical energy rate distribution $w_{\text{def}}^{\bullet}(X, Y, Z, t)$ from the stress and strain-rate patterns given by the computation. By averaging this quantity over the specimen depth, we obtained the time patterns of the 2D distribution of the mechanical energy rate $w_{\text{def}}^{\bullet}(X, Y, t) = 1/D_0 \int_{-D_0/2}^{D_0/2} w_{\text{def}}^{\bullet}(X, Y, Z, t) dZ$. We also extracted the in-plane components of the displacement vector of points located on the sample surface to analyze.

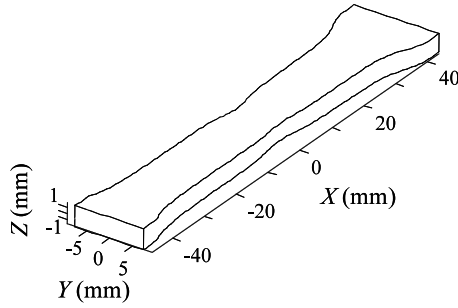


Figure 6. Initial geometry of the specimen used in the finite element computation.

The numerical displacement fields represented the kinematic data obtained by DIC. They were used to compute the in-plane strain-rate $D(X, Y, t)$ and the stress tensors $\sigma(X, Y, t)$ with the data processing described in the previous section.

In Figure 7, we compared the distribution of w_{def}^\bullet given by the finite element simulations with those associated with the depthwise-averaged tensor product $\sigma : D$. Figure 8 shows the longitudinal profile of the w_{def}^\bullet and $\sigma : D$ integrated over slides of equivalent length ($\Delta X = 0.827 \text{ mm}$) $D_0 \times W_0(X) \times \Delta X$ of the specimen. As expected, the mechanical energy was localized in the neck. Even in this highly heterogeneous triaxial situation, we obtained a satisfactory correlation on the distribution, and quantitatively good estimates of the mechanical energy rate amplitudes.

Experimental results

DIC analysis reveals variations in various kinematical data such as the Eulerian strain-rate tensor D , and the acceleration vector γ . Figure 9 (left) shows the spatiotemporal diagram of the strain-rate component $D_{xx}(X, Y = 0, t)$, and Figure 9 (right) illustrates variations in the axial components of the acceleration vector $\gamma_X(X, Y = 0, t)$. All the results presented here correspond to the same displacement-controlled test, with a conventional strain-rate $\dot{\epsilon}_c$ of about of $5 \times 10^{-3} \text{ s}^{-1}$, performed on IF-Ti steel.

At the beginning of the test, D_{xx} was lower than $\dot{\epsilon}_c$ throughout the test section of the specimen because of the finite elastic stiffness of the testing machine frame. The early and steady narrowing of the level curves indicates that the necking region (located around point A) gradually concentrates before the maximum load is reached. This also means that the gauge part of the sample is no longer uniformly strained (and stressed), thus complicating extraction of the material response and consequently its modelling and identification.

Figure 9 (right) shows that the acceleration amplitude is negligible compared with the gravity acceleration, which is classically ignored in the case of quasistatic tensile tests. This experimental result and the

| Young modulus E (MPa) | Poisson ratio ν | Yield stress σ_0 (MPa) | Hardening modulus H (MPa) |
|-------------------------|---------------------|-------------------------------|-----------------------------|
| 210000 | 0.3 | 315 | 5200 |

Table 4. Material properties used for the finite element computation.

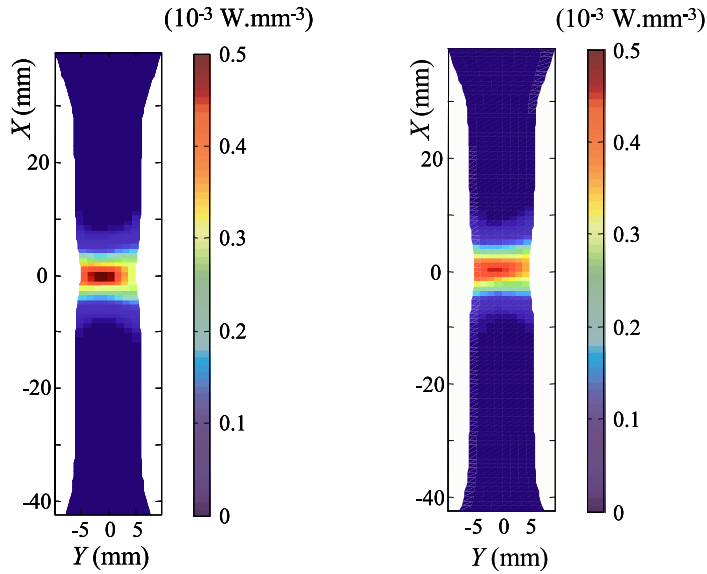


Figure 7. Distribution of the deformation energy rate given by the finite element code (left) and computed as $\sigma : D$ (right).

sample geometry (thin specimen) confirms the hypotheses used in the stress calculation (see section on *Data processing*). At the beginning of the test, uniform acceleration profiles of “significant” level (that is, 10^{-5} m s^{-2}) can be observed. They are induced by a combination of the mechanical gaps and finite stiffness of the testing machine and the time fitting of displacement data. As it was already shown in a

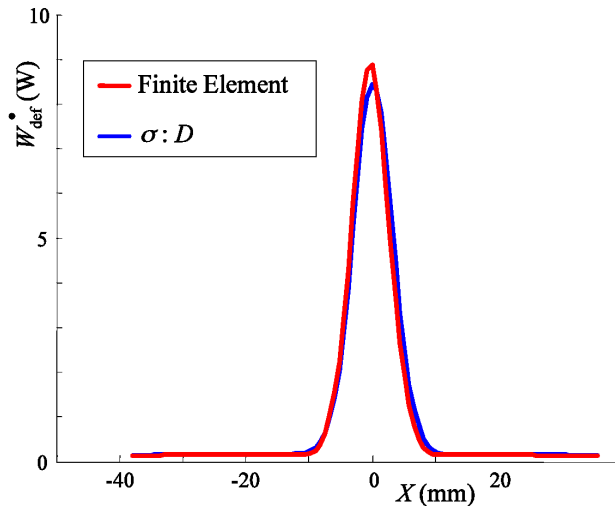


Figure 8. Longitudinal profiles of the deformation energy rate received by equally-spaced, 0.827 mm width, slices of the sample derived from kinematical data with the equilibrium equations (computed as $\sigma : D$) and directly computed within the FE code.

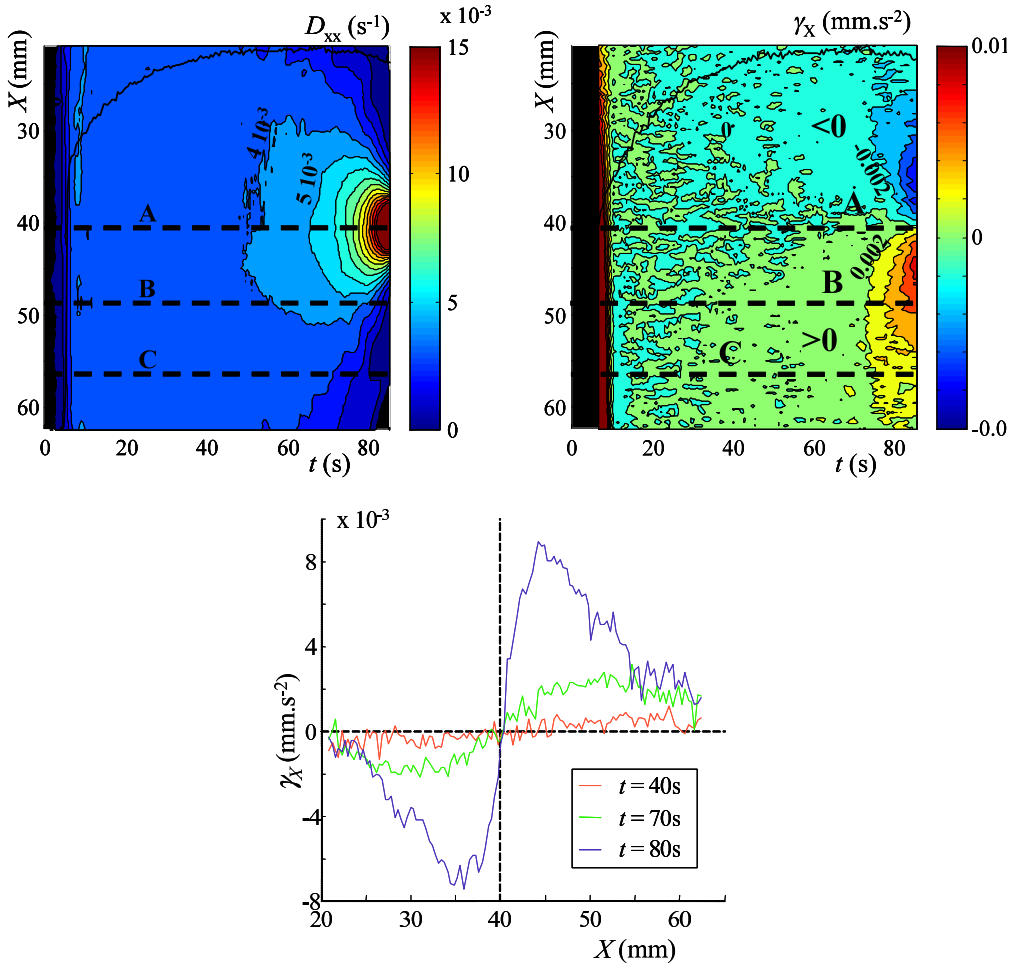


Figure 9. Spatiotemporal variation in $D_{xx}(X, 0, t)$, 10^{-3} s^{-1} between two level curves (left) and $\gamma_X(X, 0, t)$, 0.002 mm s^{-2} between two level curves (right). Different profiles of g_x acceleration at $t = 40, 70, 80 \text{ s}$ are also shown (bottom).

previous paper [Wattrisse et al. 2001b], a steady concentration of level curves of positive and negative acceleration, distributed on both sides of the necking zones, can be clearly observed. The section where the fracture will take place (that is, MSE A) is then characterized by zero acceleration. This cross-section can be easily predetermined on the basis of Figure 9 (bottom) where acceleration profiles captured at time $t = 40 \text{ s}, 70 \text{ s},$ and 80 s intersect in a given cross-section. It seems like the specimen knew where the crack would take place a long time before the crack occurred. Moreover, the odd distribution of acceleration profiles with respect to MSE A is consistent with a symmetric strain distribution centred on the neck.

The heterogeneity of the specimen response can also be clearly observed in Figure 10, illustrating variations in the longitudinal distribution of tensile stress throughout the test.

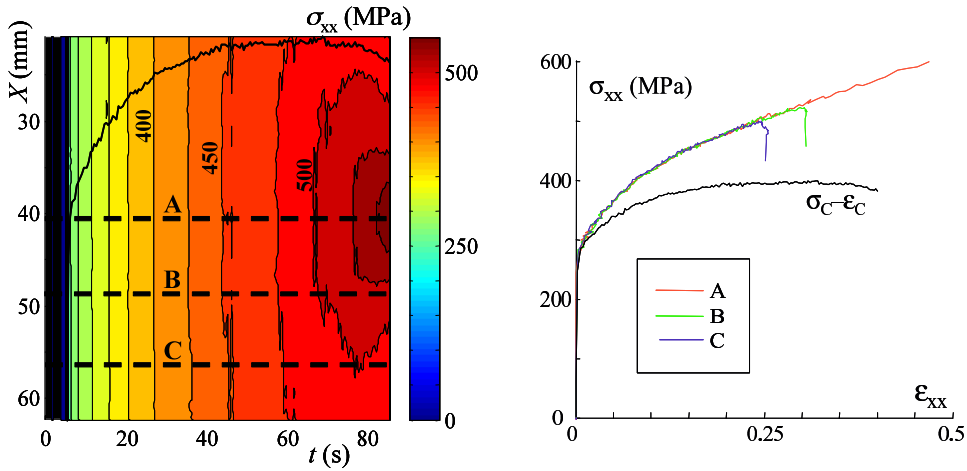


Figure 10. Mechanical response. Time duration of $\sigma_{xx}(X, 0, t)$, 25 MPa between two level curves (left); stress-strain diagrams of the three MSE **A**, **B** and **C** with the conventional stress-strain response of the specimen (right).

As expected, the stress was rather homogeneous in the sample gauge part at the beginning of the test. A stress concentration appeared in the necking zone as localization developed. Combining the local stress and strain measurements, we plotted the local stress-strain correspondence in the MSE, denoted **A**, **B** and **C** (Figure 10, right). All the curves describe a unique path at the beginning of hardening. Nevertheless, as the imposed macroscopic strain ϵ_C increased, the stress amplitudes at **A**, **B** and **C** rapidly diverged. Once the structure started to soften, σ_{xx} decreased in sections in which the strain-rate vanished (**B** and **C**), while it continued to increase in the section of the current necking zone. It thus seems that softening of the sample, translated by a nonmonotone load-elongation curve, was induced by heterogeneous hardening accompanied by elastic unloading in cross-sections outside the localization zone.

Figure 11 (left) presents the spatiotemporal distribution of the deformation energy rate during strain hardening. The data processing was stopped before the rapid growth of localized necking in order to consider relatively low temperature, stress and strain-rate gradients. In the late stages of localization, the spatial resolution of the method was not sufficient to catch the high thermal and high kinematical gradients. Investigations require then to change the scale of observation or the data processing parameters. The gradual narrowing of level curves observed in Figure 11 (left) again highlights the progressive development of localization, but now in terms of deformation energy.

The different terms on the left-hand side of Equation (8) were successively calculated to estimate the overall heat source. Figure 11 (right) shows heat source variations along the longitudinal axis of the sample. Again, the contour plot revealed progressive narrowing of the level curves. We obviously attributed the concentration of heat sources to the development of dissipative mechanisms due to localization of hardening and damage.

The intrinsic dissipation was deduced from Equation (8). For simplicity, we assumed a linear, isotropic thermoelastic behavior. The quantity w_{tec}^\bullet was determined using the computed stress data and following

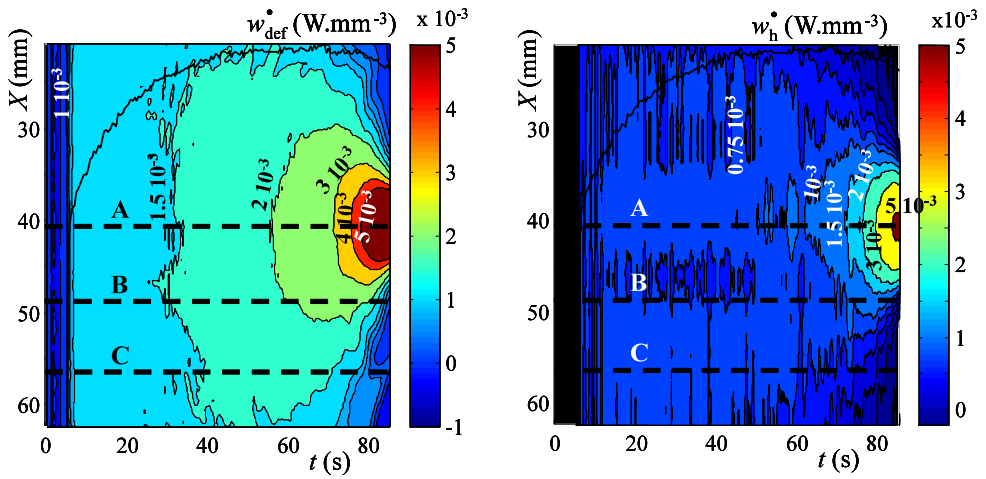


Figure 11. Spatiotemporal variations in energy rates: $w_{\text{def}}^*(X, 0, t)$ (left); $w_h^*(X, 0, t)$ (right).

the approximated definition

$$w_{\text{tec}}^* \approx -\lambda T_0 \text{tr}(\dot{\sigma}), \tag{9}$$

where λ stands for the linear thermal dilatation coefficient. The validity of this approximation is mainly due to the smallness of λ . In previous papers we already underlined that Equation (9) holds true as long as

$$\frac{9K\lambda^2}{\rho C} T \approx \frac{9K\lambda^2}{\rho C} T_0 \ll 1,$$

where K is the bulk elastic modulus.

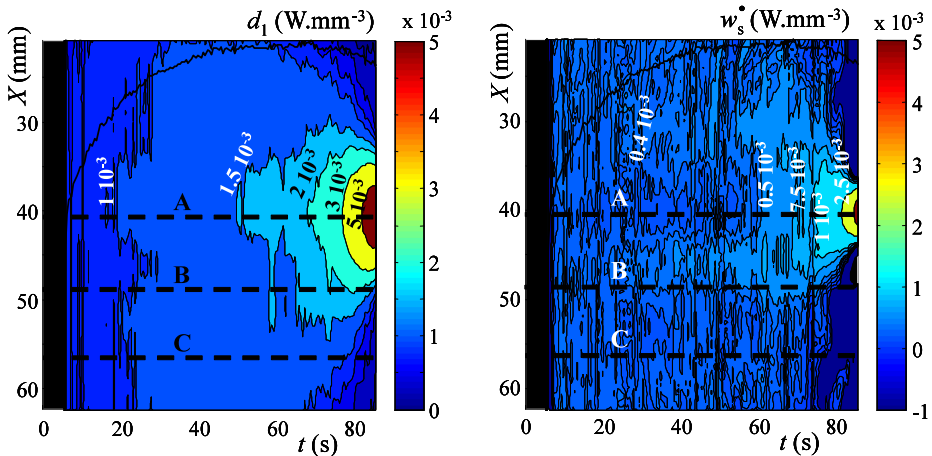


Figure 12. Spatiotemporal variations in $d_1(X, 0, t)$ (left); $w_s^*(X, 0, t)$ (right).

During elastic loading and in the early stages of plasticity, the intrinsic dissipation was very small (see [Figure 12](#), left), and then increased with hardening. The same trends could be observed in the d_1 and w_h^\bullet patterns.

According to Equations (1)–(3), the stored energy rate was then estimated by the relation

$$w_s^\bullet = w_{\text{def}}^\bullet - d_1 - w_e^\bullet.$$

The local evaluation of stored energy rate obtained via [Equation \(6\)](#) must be considered with caution owing to the successive approximations. Nevertheless, we decided to show a detailed picture of the stored energy rate in [Figure 12](#) (right). The distribution appears to be heterogeneous and concentrated in the vicinity of the necking region.

The energy distributions ($w_{\text{def}}(X, Y, t)$, $w_d(X, Y, t)$, $w_s(X, Y, t)$, ...) were computed by integrating the corresponding energy rates over time. [Figure 13](#) (left) shows variations in the different energies involved in the energy balance at MSE *A*, *B* and *C* for the same macroscopic applied load. These energies are plotted with respect to the local strain reached at each point. The test heterogeneity is noted by the fact that the curves corresponding to the three points are not identical: the trends are fairly similar, but the energy levels reached are not the same (they increase as they get close to the localization zone). Note that we did not observe significant decrease of the stored energy for large strain, particularly in the necking zone. We have to mention that this last result then differs from findings recently published by [Oliferuk and Maj \[2007\]](#) who observed strong decreases of the (overall) stored energy at the maximum load defining the famous Considère instability point. It is worth noting that the method developed to estimate the dissipated energy used an electrical analogy, these authors tuning in Joule’s effects to determine the dissipated heat, the adjustment of the electrical power being controlled by the thermal response.

[Figure 13](#) (right) shows variations in the Taylor–Quinney stored energy ratio F_w with the local strain reached at MSE *A*, *B* and *C*. We observed negative values of the stored energy at the beginning of strain hardening induced by an underestimate of the thermoelastic source intensity. This poor estimate was associated with disputable values of standard thermoelastic constants extracted from the literature.

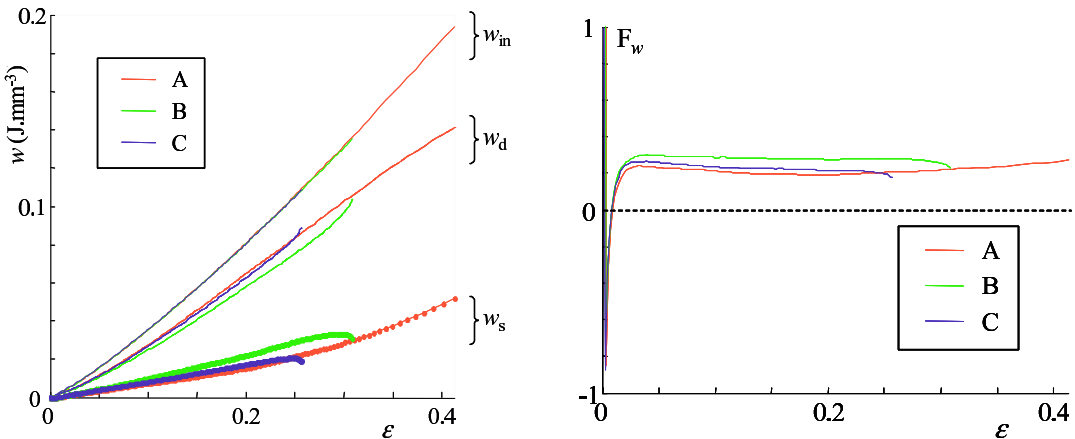


Figure 13. Energy balance at MSE *A*, *B* and *C*: variations in the stored w_s , dissipated w_d and inelastic w_{in} energies (left); local Taylor–Quinney ratio F_w (right).

This effect on the stored energy ratio is here naturally amplified by the computation of a ratio involving two small noisy amounts of energy. Once the anelastic energy becomes significant, the F_w variations are fairly similar from one point to another: rapid increase during the first stages of hardening, until a maximum is reached ($F_w \approx 0.25$), followed by a small decrease ($F_w \approx 0.2$). Further tests on different steel grades are under way to confirm these trends. Contrary to most literature results, in this study the Taylor–Quinney coefficient distributions were determined on relatively small volume elements (typically around 1 mm^3). The main advantage is to reduce the effects of localization on the energy balance. In return, the signal-to-noise-ratios are lower, so well controlled data processing techniques are required.

Conclusions

In conclusion, we designed a set-up that combined DIC with IRT. Many technical difficulties were overcome and the first results presented in this paper are encouraging. Both imaging techniques gave similar spatial distributions and temporal patterns concerning heat sources and mechanical energy rates throughout the test (especially during strain localization). The capabilities of the imaging techniques allowed us to check the local quasistatic character of strain localization. Despite the low acceleration intensities, the kinematical image processing showed that the cross-section where the crack initiated was early characterized during hardening by zero acceleration, dividing the gauge part of the sample into two parts where the accelerations were positive and negative, respectively. The combination of thermal and kinematical data illustrated that heat involved by matter convection could represent up to more than 50% of the overall heat sources. This underlined the necessity of combining DIC with IRT as soon as localization occurs.

However, several metrology problems remain. The next stage is to increase the signal-to-noise ratios of both cameras to obtain more reliable quantitative results. Moreover, the data processing will be improved in order to better account for the last localization stages. The fine knowledge of the material parameters involved in the heat conduction equation, and their possible variations with the material state, are critical for the heat source computation. Collaborations are under way with several research teams, within the framework of the French National Research Agency program, to better characterize the influence of hardening and damage on the thermophysical properties of materials.

Local determination of dissipated and stored energies is essential to test the reliability of the constitutive equations proposed in thermomechanical formalism of plasticity and damage. This could give rise to a method for separating hardening from damage, inasmuch as the latter is purely dissipative, unlike the first one. This should be performed on elastoplastic cohesive zone models used to numerically manage fracture in heterogeneous materials.

References

- [Batra and Chen 2001] R. C. Batra and L. Chen, “Effect of viscoplastic relations on the instability strain, shear band initiation strain, the strain corresponding to the minimum shear band spacing, and the band width in a thermoviscoplastic material”, *Int. J. Plasticity* **17**:11 (2001), 1465–1489.
- [Béranger et al. 1994] G. Béranger, G. Henry, and G. Sanz, *Le livre de l'acier*, Lavoisier, Paris, 1994.
- [Campagne et al. 2005] L. Campagne, L. Daridon, and S. Ahzi, “A physically based model for dynamic failure in ductile metals”, *Mech. Mat.* **37**:8 (2005), 869–886.
- [Chrysochoos and Louche 2000] A. Chrysochoos and H. Louche, “An infrared image processing to analyse the calorific effects accompanying strain localisation”, *Int. J. Eng. Sci.* **38**:16 (2000), 1759–1788.

- [Chrysochoos and Martin 1989] A. Chrysochoos and G. Martin, “Tensile test microcalorimetry for thermomechanical behaviour law analysis”, *Mat. Sci. Eng. A* **108** (1989), 25–32.
- [Chu et al. 1985] T. C. Chu, W. F. Ranson, and M. A. Sutton, “Applications of digital-image-correlation techniques to experimental mechanics”, *Exp. Mech.* **25**:3 (1985), 232–244.
- [Guduru et al. 2001] P. R. Guduru, A. T. Zehnder, A. J. Rosakis, and G. Ravichandran, “Dynamic full field measurements of crack tip temperatures”, *Eng. Fract. Mech.* **68**:14 (2001), 1535–1556.
- [Halphen and Nguyen 1975] B. Halphen and Q. S. Nguyen, “Sur les matériaux standards généralisés”, *J. de Méc.* **14**:1 (1975), 39–63.
- [Honorat et al. 2005] V. Honorat, S. Moreau, J.-M. Muracciole, A. Chrysochoos, and B. Wattrisse, “Calorimetric analysis of polymer behaviour using a pixel calibration of an I. R. F. P. A. camera”, *Int. J. Q. I. R. T.* **2**:2 (2005), 153–172.
- [Lemaitre and Chaboche 1985] J. Lemaitre and J.-L. Chaboche, *Mécanique des matériaux solides*, Dunod, Paris, 1985.
- [Lubliner 1991] J. Lubliner, “A simple model of generalized plasticity”, *Int. J. Solids Struct.* **28**:6 (1991), 769–778.
- [Mercier and Molinari 1998] S. Mercier and A. Molinari, “Steady-state shear band propagation under dynamic conditions”, *J. Mech. Phys. Solids* **46**:8 (1998), 1463–1495.
- [Moreau et al. 2004] S. Moreau, J.-M. Muracciole, and B. Wattrisse, “Experimental energy balance for thermoplastics stretched at room temperature”, in *Proceedings of the 12th I. C. E. M. Conference* (Bari, Italy), vol. CDROM 8, 2004.
- [Oliferuk and Maj 2007] W. Oliferuk and M. Maj, “Plastic instability criterion based on energy conversion”, *Mat. Sci. Eng. A* **462**:1-2 (2007), 363–366.
- [Rittel 1999] D. Rittel, “On the conversion of plastic work to heat during high strain rate deformation of glassy polymers”, *Mech. Mat.* **31**:2 (1999), 131–139.
- [Rosakis et al. 2000] P. Rosakis, A. J. Rosakis, G. Ravichandran, and J. Hodowany, “A thermodynamic internal variable model for the partition of plastic work into heat and stored energy in metals”, *J. Mech. Phys. Solids* **48**:3 (2000), 581–607.
- [Rusinek et al. 2007] A. Rusinek, R. Zaera, and J. Klepaczko, “Constitutive relations in 3-D for a wide range of strain rates and temperatures: application to mild steels”, *Int. J. Solids Struct.* **44**:17 (2007), 5611–5634.
- [Schmidt et al. 1945] A. O. Schmidt, O. W. Gilbert, and A. Boston, “Thermal balance method and mechanical investigation for evaluating machinability”, *Trans. A.S.M.E.* **67** (1945), 225–232.
- [Shenogin et al. 2002] S. V. Shenogin, G. W. Hohne, and E. F. Oleinik, “Thermodynamics of the pre-yield deformation behavior of glassy polymers: measurements with new deformation calorimeter”, *Thermochim. Acta* **391**:1-2 (2002), 13–23.
- [Taylor and Quinney 1934] G. I. Taylor and H. Quinney, “The latent energy remaining in a metal after cold working”, *P. Roy. Soc. Lond. A Mat.* **143**:849 (1934), 307–326.
- [Wattrisse et al. 2001a] B. Wattrisse, A. Chrysochoos, J.-M. Muracciole, and M. Némoz-Gaillard, “Analysis of strain localization during tensile tests by digital image correlation”, *Exp. Mech.* **41**:1 (2001), 29–39.
- [Wattrisse et al. 2001b] B. Wattrisse, A. Chrysochoos, J.-M. Muracciole, and M. Némoz-Gaillard, “Kinematic manifestations of localisation phenomena in steels by digital image correlation”, *Eur. J. Mech. A Solids* **20**:2 (2001), 189–211.
- [Zehnder et al. 1998] A. T. Zehnder, E. Babinsky, and T. Palmer, “Hybrid method for determining the fraction of plastic work converted to heat”, *Exp. Mech.* **38**:4 (1998), 295–302.

Received 18 Dec 2007. Revised 9 Jun 2008. Accepted 19 Jun 2008.

ANDRÉ CHRYSOCHOOS: chryso@lmgc.univ-montp2.fr

LMGC UM 5508 CNRS UMII, CC 081, Université Montpellier II, Place E. Bataillon, 34095 Montpellier Cedex, France

BERTRAND WATTRISSE: wattrisse@lmgc.univ-montp2.fr

LMGC UM 5508 CNRS UMII, CC 081, Université Montpellier II, Place E. Bataillon, 34095 Montpellier Cedex, France

JEAN-MICHEL MURACCIOLE: muracciole@lmgc.univ-montp2.fr

LMGC UM 5508 CNRS UMII, CC 081, Université Montpellier II, Place E. Bataillon, 34095 Montpellier Cedex, France

YVES EL KAÏM: elkaim@lmgc.univ-montp2.fr

LMGC UM 5508 CNRS UMII, CC 081, Université Montpellier II, Place E. Bataillon, 34095 Montpellier Cedex, France

PREDICTION OF DRYING SHRINKAGE BEYOND THE PORE ISODEFORMATION ASSUMPTION

OLIVIER P. COUSSY AND SÉBASTIEN BRISARD

The paper aims at assessing how, for a porous material whose pore size distribution is experimentally known, the variation in pore deformation with pore size might affect predictions of drying shrinkage. Unsaturated poroelasticity is first revisited in a general macroscopic thermodynamic framework irrespective of any morphology of the porous space. Saturation is shown to be a state function of capillary pressure governing the change in the solid-fluid interface energy; it can be experimentally obtained from a knowledge of pore size distribution only. Unsaturated poroelastic properties are then determined under three homogenization schemes: the standard Mori–Tanaka scheme, the self-consistent scheme, and the differential homogenization scheme extended to unsaturated conditions. Except for the Mori–Tanaka scheme, the function weighting the fluid pore pressure in the poroelastic constitutive equations is found to depart from the pore volume fraction the liquid occupies. As a result the pores do not deform uniformly. This departure roughly accounts for the difference in deformation between pores of different sizes and subjected to the same pressure, and it is found to significantly affect predictions of drying shrinkage, in particular for cement paste.

Drying shrinkage of water-infiltrated materials is relevant to many kinds of materials and disciplines: cement-based materials in civil engineering [Baroghel-Bouny et al. 1999], woods in the building industry [Santos 2000], plants in botany [Kozłowski and Pallardy 2002], soils in soil science [Chertkov 2002], gels in physical chemistry [Smith et al. 1995], vegetables in foods engineering [Ratti 1994], tissues in biomechanics [Gusnard and Kirschner 1977], etc. The mechanism of drying shrinkage is well known. When a porous material is subjected to an outer relative humidity lower than its initial inner relative humidity, the vapour thermodynamic imbalance forces the porous material to exchange water vapour with the outer atmosphere, so that the outer relative humidity progressively takes hold within the material. In turn liquid water simultaneously evaporates in order to maintain the vapour-liquid equilibrium. This causes the decrease of the degree of liquid saturation. The shrinkage of the porous material finally results from the lowering in liquid pressure induced by the desaturation process at the gas-liquid water interface. While the kinetics of drying is governed by transport phenomena [Mainguy et al. 2001], asymptotic drying shrinkage is governed by the outer relative humidity only, since, asymptotically, the value of the air pressure tends toward the atmospheric pressure.

The macroscopic modeling of drying shrinkage has been addressed by many authors; see [Bazant and Wittmann 1982; Coussy et al. 1998], for example. In the last decade the development of microporomechanics [Dormieux et al. 2006a] has provided new tools to assess the influence of microstructure

Keywords: drying, unsaturated, poroelasticity, homogenization scheme, pore size distribution, homogenization, drying shrinkage.

The authors acknowledge the support of the ATILH-CNRS research program “Porosity, Transport, Strength”.

upon drying shrinkage [Chateau and Dormieux 2002], and in particular the influence of pore shape and geometry changes [Chateau et al. 2003].

In most approaches to the mechanical behaviour of unsaturated porous materials, such as geomaterials [Lewis and Schrefler 1998; Hutter et al. 1999], and to the subsequent prediction of their drying shrinkage, the pore volume fraction occupied by a fluid is generally taken as the weighting factor for the fluid pore pressure in the constitutive equations. This approach results from the assumption that all pores (irrespective of their size) undergo the same deformation when subjected to the same pressure. Owing to their disparity in size and shape, this assumption is questionable.

The main goal of this paper is to draw attention to how the difference in deformation undergone by pores can quantitatively be taken into account in predicting drying shrinkage for a material whose pore size distribution is experimentally known. In Section 1 the poroelasticity of unsaturated porous solids is considered within a general macroscopic thermodynamic framework, irrespective of the morphology of the porous space. Saturation is shown to be a state function of the capillary pressure governing the change in the solid-fluid interface energy and can be determined from a knowledge of the pore size distribution only. In Section 2, following the methods developed in [Dormieux et al. 2006a], we determine the unsaturated poroelastic properties with the help of three homogenization schemes: the standard Mori–Tanaka scheme, the self-consistent scheme and the differential homogenization scheme extended to unsaturated conditions. Except for the Mori–Tanaka scheme, the function weighting the fluid pore pressure in the poroelastic constitutive equations departs from the pore volume fraction that the liquid occupies. Using these homogenization schemes and adopting the experimental pore size distribution appropriate for a typical cement paste, our analysis reveals (Section 3) that this departure may significantly affect predictions of drying shrinkage and ultimately of failure for a water-infiltrated material subjected to drying.

1. Unsaturated poroelasticity

Capillary pressure curve. Consider an element of a porous solid of overall volume V , with initial porosity ϕ_0 , so that its porous volume is $\phi_0 V$. For the time being, we assume that the porous solid is undeformable, so that its porosity remains constant. The porous solid, initially fully saturated by a wetting liquid denoted by subscript L, is progressively invaded by a nonwetting gas denoted by subscript G. At a given time the fractions of the porous volume $\phi_0 V$ occupied by the liquid and by the gas are S_L and S_G . We write

$$\phi_L = \phi_0 S_L, \quad \phi_G = \phi_0 S_G, \quad S_L + S_G = 1, \quad (1-1)$$

where ϕ_J is the partial porosity related to phase J ($=$ L or G).

Assuming no hysteresis, the first and second laws of thermodynamics combine to give the isothermal incremental free energy balance

$$\mu_L dn_L + \mu_G dn_G - dA = 0 \quad (1-2)$$

[Coussy 2004], where μ_J and n_J are respectively the chemical potential and the number of moles per unit of volume V relative to phase J, and A is the Helmholtz free energy of the whole matter contained in the volume V . The standard isothermal Gibbs–Duhem equality applied to phase J reads

$$\phi_J dp_J - n_J d\mu_J = 0,$$

where p_J is the pressure related to phase J. Let F denote the Helmholtz free energy of the system once the bulk phases L and G are removed. Owing to the additive character of energy we can write

$$F = A - (\mu_L n_L + \mu_G n_G - \phi_L p_L - \phi_G p_G). \quad (1-3)$$

The three last equations combine to give

$$p_L d\phi_L + p_G d\phi_G - dF = 0. \quad (1-4)$$

Since in (1-3) the bulk liquid and gas phases have been removed and the porous solid is assumed to be undeformable, the free energy F reduces to the surface energy of the interfaces between the phases and the solid matrix. Denoting by U the surface energy of these interfaces per unit volume, we write

$$F = \phi_0 U. \quad (1-5)$$

Substituting (1-1) and (1-5) in (1-4) we get

$$p_G - p_L = -\frac{dU}{dS_L}, \quad (1-6)$$

which shows that the liquid saturation S_L is a state function of the capillary pressure $p_G - p_L$. We write

$$S_L = \varpi (p_G - p_L), \quad (1-7)$$

where ϖ describes the so-called capillary curve. The macroscopic capillary curve can receive a simple microscopic interpretation at the pore scale. At that scale the mechanical equilibrium of the current gas-liquid interface is governed by the Laplace law according to

$$p_G - p_L = \frac{2\gamma_{GL}}{r}, \quad (1-8)$$

where γ_{LG} is the energy per unit of surface of the gas-liquid interface and r is the mean curvature radius. As illustrated in Figure 1 for a cement paste, standard porosimetry provides the cumulated porous volume fraction $S(r)$ of pores having a pore entry radius smaller than r . For a given value of the capillary pressure $p_G - p_L$, pores having an entry radius smaller than the one given by (1-8) will still remain filled with liquid, while pores with larger entry radius will be invaded by the gas. As a consequence we write

$$S_L = S(r). \quad (1-9)$$

Combining (1-8) and (1-9) we get

$$S_L = S\left(\frac{2\gamma_{GL}}{p_G - p_L}\right), \quad (1-10)$$

which provides an explicit determination of the capillary curve.

Unsaturated state equations of poroelasticity. Now consider a deformable porous solid and denote by σ_{ij} and ε_{ij} the stress and strain components. The free energy balance (1-4) is extended to

$$\sigma_{ij} d\varepsilon_{ij} + p_L d\phi_L + p_G d\phi_G - dF = 0, \quad (1-11)$$

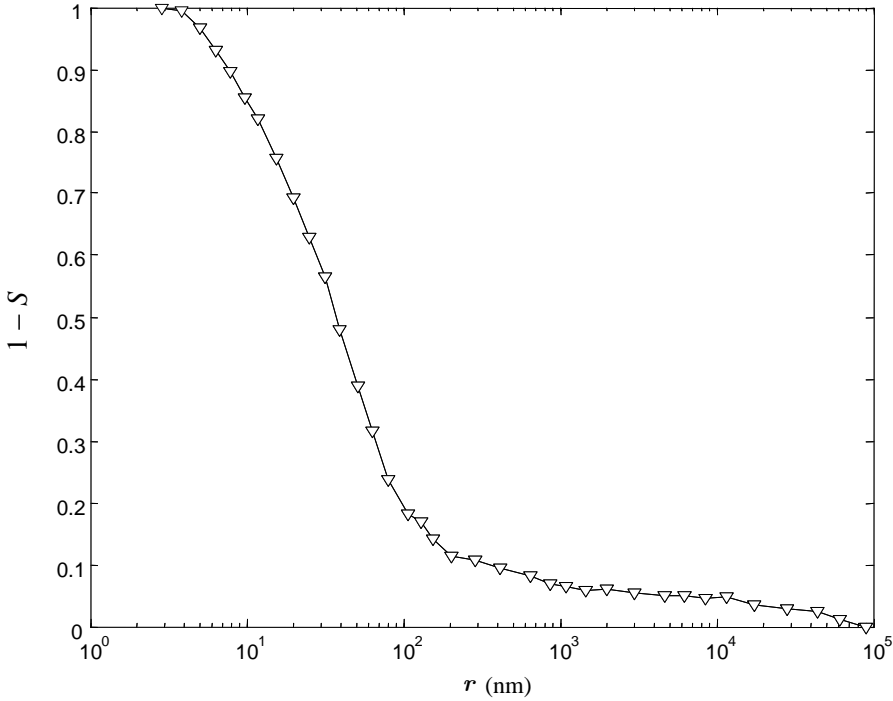


Figure 1. Cumulative pore volume fraction $1 - S(r)$ of pores having a pore entry radius greater than r for a typical cement paste (extracted from [Huang and Feldman 1985]).

where $\sigma_{ij} d\varepsilon_{ij}$ accounts for the strain work. Since the porous space now deforms, instead of (1-1) we write

$$\phi_L = \phi_0 S_L + \varphi_L, \quad \phi_G = \phi_0 S_G + \varphi_G, \quad S_L + S_G = 1, \quad (1-12)$$

where φ_J accounts for the change in the partial porosity ϕ_J due to deformation only. In contrast to the standard Eulerian configuration referring to the current deformed configuration, as recently introduced in [Coussy 2007], the saturation S_J can be regarded as a Lagrangian saturation related to phase J, since it refers to the undeformable configuration. More precisely, starting from full liquid saturation, $\phi_0 S_L V$ is the volume in the undeformable configuration whose solid walls will still be wetted in the current deformed configuration [Coussy 2007]. This is sketched in Figure 2 for two distinct current deformed configurations, the undeformed reference configuration pertaining to liquid saturated conditions at zero pressure p_L .

With regard to the undeformable porous solid, the free energy F of the deformable system obtained by removing the bulk phases L and G now splits into the surface energy $\phi_0 U$ associated with the interfaces, and the elastic energy Ψ_S stored in the deformable solid matrix. Accordingly expression (1-5) transforms into

$$F = \phi_0 U + \Psi_S. \quad (1-13)$$

Substitution of (1-12) and (1-13) into (1-11) yields

$$\sigma_{ij} d\varepsilon_{ij} + p_L d\varphi_L + p_G d\varphi_G - d\Psi_S - \phi_0((p_G - p_L) dS_L + dU) = 0. \quad (1-14)$$

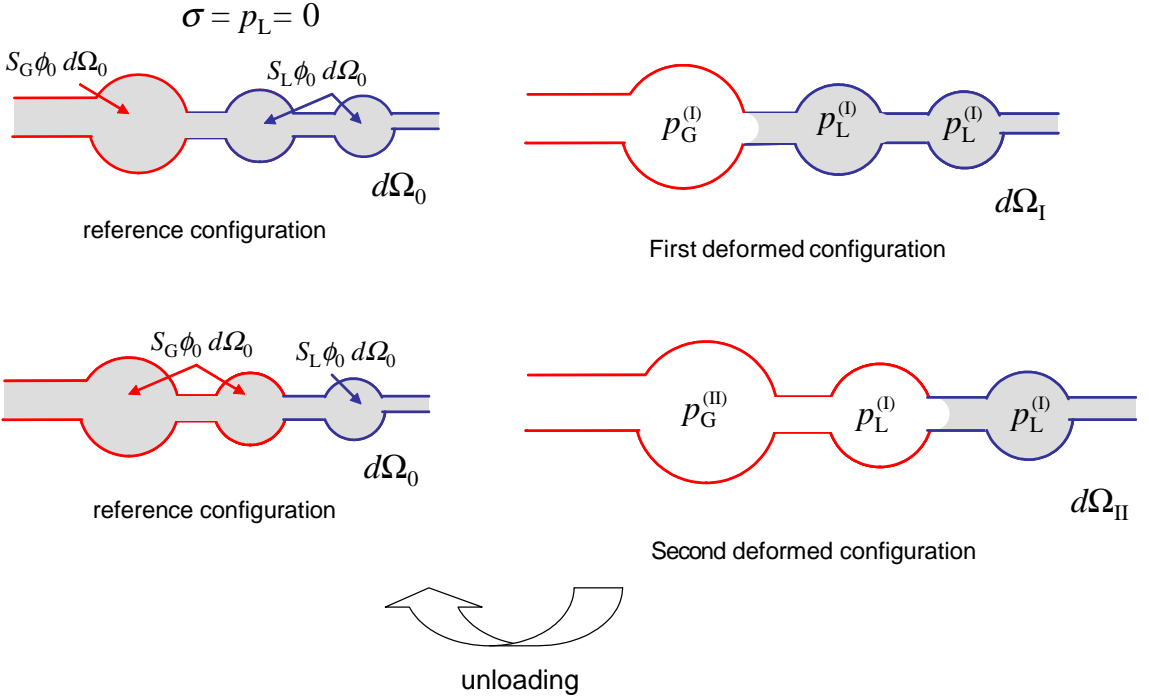


Figure 2. Unlike the case of disconnected porous networks, for a porous solid formed of connected pores embedded in a solid matrix the Lagrangian saturation S_J does not remain constant, since either fluid, the gas G or the liquid L, can invade the porous solid or recede from it.

When the porous solid is deformable, in contrast to the previous section, even when S_L is held constant, U can change because it is affected by the deformation of the solid-fluid interface. The change in U is then due to the work produced by the pressure difference exerted on the solid-fluid interface and made possible by the solid-fluid surface energy γ_{SG} or γ_{SL} . However, because of the low values of γ_{SG} and γ_{SL} compared to the elastic energy, as the liquid saturation changes during the process of invasion by the gas, the change in U is mainly due to the creation of new interfaces between the phases. According to the analysis of the previous section, U can then still be considered as a function of the liquid saturation S_L only. Conversely, if we assume infinitesimal elastic deformations of the porous solid, the elastic energy Ψ_S will be slightly affected by the variation dS_L of the liquid saturation. As a result, (1-14) allows us to conclude that (1-5) will still hold to a good approximation, while the free energy balance related to the deformable porous solid obtained by removing the interfaces is

$$\sigma_{ij} d\varepsilon_{ij} + p_L d\phi_L + p_G d\phi_G - d\Psi_S = 0, \tag{1-15}$$

from which we derive

$$\sigma_{ij} = \frac{\partial \Psi_S}{\partial \varepsilon_{ij}}, \quad p_L = \frac{\partial \Psi_S}{\partial \phi_L}, \quad p_G = \frac{\partial \Psi_S}{\partial \phi_G}. \tag{1-16}$$

Letting $W_S = \sigma_{ij}\varepsilon_{ij} + \varphi_L p_L + \varphi_G p_G - \Psi_S$ be the Legendre transform of Ψ_S with regard to φ_J we alternatively get

$$\varepsilon_{ij} = \frac{\partial W_S}{\partial \sigma_{ij}}, \quad \varphi_L = \frac{\partial W_S}{\partial p_L}, \quad \varphi_G = \frac{\partial W_S}{\partial p_G}. \quad (1-17)$$

In the context of both infinitesimal deformation, and restricting to linear isotropic poroelasticity, the expression of the elastic energy of the solid matrix, $W_S = \Psi_S$, is

$$W_S = \frac{1}{2K} (\sigma + b_L p_L + b_G p_G)^2 + \frac{1}{2N_{LL}} p_L^2 + \frac{1}{N_{LG}} p_L p_G + \frac{1}{2N_{GG}} p_G^2 + \frac{1}{4G} s_{ij} s_{jk}, \quad (1-18)$$

where $\sigma = \sigma_{kk}/3$ and s_{ij} represent the mean stress and the components of the deviatoric stress tensor. Letting $\varepsilon = \varepsilon_{kk}$ be the volumetric strain and substituting (1-18) in (1-17) we finally get

$$\sigma = K\varepsilon - b_L p_L - b_G p_G, \quad s_{ij} = 2G e_{ij}, \quad (1-19)$$

$$\varphi_L = b_L \varepsilon + p_L / N_{LL} + p_G / N_{LG}, \quad (1-20)$$

$$\varphi_G = b_G \varepsilon + p_L / N_{LG} + p_G / N_{GG}. \quad (1-21)$$

K and G are therefore identified as the bulk modulus and the shear modulus of the dry porous solid with no internal pore pressures.

When $p_L = p_G$ we must retrieve the saturated case so that we have [Coussy 2004]

$$b_L + b_G = b = 1 - \frac{K}{k_S} \quad \text{and} \quad \frac{1}{N_{LL}} + \frac{2}{N_{LG}} + \frac{1}{N_{GG}} = \frac{1}{N} = \frac{b - \phi_0}{k_S}, \quad (1-22)$$

where b and N are the poroelastic properties of the porous solid with uniform pore pressure, while k_S is the bulk modulus of the solid matrix assumed to be homogeneous. Using mesoscopic-macroscopic considerations [Coussy 1991; 2007] or more refined upscaling methods [Dormieux et al. 2006a], it can further be shown that

$$\frac{1}{N_{JJ}} + \frac{1}{N_{LG}} = \frac{b_J - \phi_0 S_J}{k_S}. \quad (1-23)$$

Provided that k_S is known, these relations are independent of the porous solid considered. In contrast, separate expressions for the poroelastic properties K , G , b_J and N_{JK} as functions of the porosity ϕ_0 and saturation S_J require specific information on the porous solid considered.

2. Estimates of the unsaturated poroelastic properties

Pore isodeformation. The first relation in (1-22) shows that b_L and b_G are not independent and allows us to introduce a Bishop-like parameter χ (see [Bishop and Blight 1963]) depending on the liquid saturation S_L and such that

$$b_L = b\chi(S_L), \quad b_G = b(1 - \chi(S_L)). \quad (2-1)$$

The explicit determination of function $\chi(S_L)$ requires additional information. One assumption, sometimes made implicitly [Coussy 2004] or explored explicitly [Chateau and Dormieux 2002; Dormieux et al. 2006a; Coussy 2007], is the isodeformation of the porous volumes respectively occupied by the

liquid and the gas in the absence of any pore pressure. This assumption amounts to writing

$$\frac{\varphi_L}{\phi_0 S_L} = \frac{\varphi_G}{\phi_0 S_G} \quad \text{when } p_L = p_G = 0. \quad (2-2)$$

Substitution of (2-2) in (1-20) and (1-21) provides

$$\frac{b_L}{S_L} = \frac{b_G}{S_G}. \quad (2-3)$$

Substituting (2-3) in (1-22)₁ we finally get the simple identifications

$$b_J = bS_J, \quad \chi = S_L. \quad (2-4)$$

From (1-23) and (2-4) it can easily be shown that the porous volumes occupied by the liquid and the gas would still deform equally if they are subjected to the same pressure.

Mori–Tanaka and self-consistent schemes. When the pore isodeformation assumption is relevant, we are left with the determination of the bulk modulus K and the shear modulus G as functions of the porosity ϕ_0 . This can be achieved using upscaling procedures, the details of which we cannot go into here; we will only recall well known results in view of their further application to the analysis of drying shrinkage. (For a comprehensive and fruitful application of micromechanics to porous materials, see [Dormieux et al. 2006a].)

Since the overall volumetric strain ε is the averaged volumetric strain, we can write, letting ε_S be the volumetric strain of the solid matrix,

$$\varepsilon = (1 - \phi_0)\varepsilon_S + \varphi_L + \varphi_G. \quad (2-5)$$

The variation of volume $\varphi_J V$ of a spherical void of initial volume $\phi_0 S_J V$, which is embedded within an elastic matrix with k and g as bulk and shear moduli and which is subjected to the pore pressure p_J , can be expressed in the form

$$\frac{\varphi_J}{\phi_0 S_J} = \left(1 + \frac{3k}{4g}\right)\varepsilon_0 + \frac{3}{4g}p_J, \quad (2-6)$$

where ε_0 is the volumetric strain prescribed at infinity. If the spherical void is replaced by a spherical solid inclusion with k_S as bulk modulus, the volumetric strain ε_S of the latter is given by

$$\varepsilon_S = \frac{3k + 4g}{3k_S + 4g}\varepsilon_0. \quad (2-7)$$

Substitution of (2-6) for $J = L$ and G and (2-7) into (2-5) yields ε_0 in the form

$$\varepsilon_0 = \frac{3k_S + 4g}{3\phi_0 k_S + 4g} \times \frac{4g}{3k + 4g} \left(\varepsilon - \frac{3\phi_0}{4g} (S_L p_L + S_G p_G) \right). \quad (2-8)$$

In turn, substituting (2-8) in (2-6) we recover the constitutive equations (1-20) and (1-21) of unsaturated poroelasticity and the associated relations (1-22)–(1-23), but now with the benefit of the new relations

$$b_J = bS_J, \quad K = (1 - \phi_0) \frac{4k_S g}{3\phi_0 k_S + 4g}, \quad (2-9)$$

$$\frac{1}{N_{JJ}} = S_J^2 \left(\frac{1}{N} - \frac{3\phi_0}{4g} \right) + \frac{3\phi_0 S_J}{4g}, \quad \frac{1}{N_{LG}} = S_L S_G \left(\frac{1}{N} - \frac{3\phi_0}{4g} \right).$$

These relations agree with those given in [Dormieux et al. 2006a].

The homogenization schemes differ by the choice of the embedding medium with elastic properties k and g . The Mori–Tanaka scheme consists in choosing the solid matrix as the embedding medium, that is, $k = k_S$ and $g = g_S$ so that $\varepsilon_0 = \varepsilon_S$. The self-consistent scheme consists in choosing as the embedding medium the porous solid whose poroelastic properties we seek, that is, $k = K$ and $g = G$. The determination of the relations providing the missing relation involving the shear modulus G is much less straightforward, since it corresponds to prescribing at infinity the deviatoric strain instead of the volumetric strain [Dormieux et al. 2006a]. We limit ourselves to recalling the final result:

$$G = (1 - \phi_0) \frac{(9k + 8g) g_S}{9k \left(1 + \frac{2}{3}\phi_0 g_S/g\right) + 8g \left(1 + \frac{3}{2}\phi_0 g_S/g\right)}. \quad (2-10)$$

In view of the explicit determination of K and G , further calculations lead us to rewrite (2-9) and (2-10), with $k = K$ and $g = G$, in the more convenient form

$$\frac{K}{k_S} = 1 - \phi_0 (1 + 3K/4G), \quad \frac{G}{g_S} = 1 - 5\phi_0 \frac{1 + 4G/3K}{3 + 8G/3K}. \quad (2-11)$$

Letting

$$A = \frac{4G}{3K} = 2 \frac{1 - 2\nu}{1 + \nu}, \quad a_S = \frac{4g_S}{3k_S} = 2 \frac{1 - 2\nu_S}{1 + \nu_S}, \quad (2-12)$$

where ν and ν_S stand for the Poisson coefficients, relations (2-11) combine to give

$$2(1 - \phi_0) \left(A - \frac{\phi_0}{1 - \phi_0} \right)^2 + (3 - \phi_0 - (2 - 5\phi_0)a_S) \left(A - \frac{\phi_0}{1 - \phi_0} \right) - 3 \frac{1 - 2\phi_0}{1 - \phi_0} a_S = 0. \quad (2-13)$$

Retaining the solution of (2-13) that matches the solution $\phi_0/(1 - \phi_0)$ for $k_S \rightarrow \infty$, namely $a_S = 0$, we finally find

$$A - \frac{\phi_0}{1 - \phi_0} = \frac{(2 - 5\phi_0)a_S - 3 + \phi_0 + \sqrt{(3 - \phi_0 - (2 - 5\phi_0)a_S)^2 + 24a_S(1 - 2\phi_0)}}{4(1 - \phi_0)}, \quad (2-14)$$

which in turn can be substituted in the relation $\frac{K}{g_S} = \frac{4G}{3Ag_S}$ and in (2-11)₂.

Beyond pore isodeformation. The first relation in (2-9), which is identical to (2-4), holds irrespective of the choice of the embedding medium. This is because the standard upscaling schemes considered here are all based on the solution to the problem of a single inclusion embedded in an infinite medium. Therefore neither absolute length scales (since the embedding medium is infinite) nor relative ones (since

only one inclusion is considered at a time) can be introduced. Thus no scale effect associated with the pore size distribution can arise from these models.

With a view toward accounting for this scale effect, we start by considering only two sizes of pores, denoted respectively by subscript G (the larger pores occupied by the nonwetting gas) and subscript L (smaller pores occupied by the wetting liquid), in conformity with the analysis we carried out in the first section.

We assume scale separation between the smaller and larger pores, which is clearly a convenient over-simplification, since the smallest pores occupied by the gas have a size comparable with the largest pores occupied by the liquid. Within this assumption, the larger pores are embedded in a porous matrix. At this scale we have

$$b_G = 1 - \frac{K}{\kappa_G} \quad \text{and} \quad \frac{1}{N_{GG}} = \frac{b_G - \phi_0 S_G}{\kappa_G}, \quad (2-15)$$

where κ_G is the bulk modulus of the porous solid matrix consisting of the original solid matrix and of the smaller pores forming the porous volume at pressure p_L . The porosity ϕ_0^G of this porous solid matrix is the ratio of the porous volume at pressure p_L to the overall volume from which we remove the porous volume at pressure p_G . Accordingly we write

$$\phi_0^G = \frac{\phi_0 S_L}{1 - \phi_0 S_G}. \quad (2-16)$$

From the general relation (1-23) the other poroelastic properties are then derived in sequence:

$$b_L = b - b_G = \frac{K}{\kappa_G} - \frac{K}{k_S}, \quad \frac{1}{N_{LG}} = \frac{b_G - \phi_0 S_G}{k_S} - \frac{b_G - \phi_0 S_G}{\kappa_G}, \quad \frac{1}{N_{LL}} = \frac{b_L - \phi_0 S_L}{k_S} - \frac{1}{N_{LG}}. \quad (2-17)$$

The assessment of the poroelastic properties requires the determination of κ_G . One might think of carrying out this determination using the Mori–Tanaka scheme of the previous section, the expression for κ_G being derived from (2-9)₂ in the form

$$\kappa_G = (1 - \phi_0^G) \frac{4k_S g_S}{3\phi_0^G k_S + 4g_S}. \quad (2-18)$$

However, when combining (2-16)–(2-18) it can easily be checked that relations (2-9) are preserved. This is an unexpected result, for it can be shown that the two populations of pores do not sustain the same volumetric strain when subjected to the same pressure. However, the macroscopic shrinkage turns out to be equal to what it would be had the pore isodeformation assumption been valid ($b_J = b_{S_J}$).

Turning our attention to the self-consistent scheme, we have

$$\begin{aligned} K &= (1 - \phi_0) \frac{4k_S G}{3\phi_0 k_S + 4G}, \quad \kappa_G = (1 - \phi_0^G) \frac{4k_S \gamma_G}{3\phi_0^G k_S + 4\gamma_G} \\ \gamma_G &= (1 - \phi_0^G) \frac{(9\kappa_G + 8\gamma_G) g_S}{9\kappa_G (1 + \frac{2}{3}\phi_0^G g_S / \gamma_G) + 8\gamma_G (1 + \frac{3}{2}\phi_0^G g_S / \gamma_G)}, \end{aligned} \quad (2-19)$$

where γ_G is the shear modulus associated with κ_G . Explicit expressions for κ_G and γ_G can be obtained with the help of the same procedure that led to (2-14) in the previous section. Substituting (2-19) in

(2-15)₁ and (2-17)₁, while also using (2-16), we obtain

$$b_G = b + \phi_0 S_L \frac{1 - G/\gamma_G}{\phi_0 + 4G/3k_S} \quad \text{and} \quad b_L = b S_L - \phi_0 S_L \frac{1 - G/\gamma_G}{\phi_0 + 4G/3k_S}. \quad (2-20)$$

Since $b = 1 - K/k_S$, the expression (2-19) for K can be combined with (2-20) to give the following expression for the quantity χ of (2-1):

$$\chi = S_L \left(\frac{G/\gamma_G + 4G/3k_S}{1 + 4G/3k_S} \right). \quad (2-21)$$

As a consequence, relation (2-4) no longer holds, so the pores do not undergo the same deformation. The remaining poroelastic properties N_{JK} are obtained by substituting (2-18)–(2-20) into (2-17).

The third homogenization scheme we consider is the differential scheme. Whereas the use of this scheme is well known to provide assessments of the elastic properties K and G of a porous solid, to the authors' knowledge it has never been used for the assessment of unsaturated poroelastic properties, and thereby for the prediction of the drying shrinkage of a porous solid. As sketched in Figure 2, the original idea of the differential scheme consists in progressively introducing the porosity by infinitesimal volume fractions according to an iterative procedure. At a given stage of the iterative procedure the porosity $\phi_0 S$ has already been introduced. The next step consists in removing a new volume fraction df_0 out of the porous solid and replacing it by the same volume of pores. Since the fraction $\phi_0 S$ of df_0 already consisted of pores, the incremental porosity $\phi_0 dS$ finally created is given by

$$df_0 = \frac{\phi_0 dS}{1 - \phi_0 S}. \quad (2-22)$$

The removal of the fraction df_0 has transformed the current bulk and shear moduli κ and γ in the new moduli $\kappa + d\kappa$ and $\gamma + d\gamma$, which can be computed as those of a porous solid of porosity df_0 whose solid matrix has κ and γ as bulk and shear moduli. Accordingly, in the left-hand side of (2-9)₂ and in (2-10), we replace K and G by $\kappa + d\kappa$ and $\gamma + d\gamma$, while on the right-hand side we replace k_S , g_S and ϕ_0 by κ , γ and df_0 . Retaining only the terms of main order with regard to the infinitesimals $d\kappa$, $d\gamma$ and dS , we finally get

$$\frac{d\kappa}{\kappa} = -\frac{\phi_0 dS}{1 - \phi_0 S} \left(1 + \frac{1}{a} \right), \quad \frac{d\gamma}{\gamma} = -\frac{5\phi_0 dS}{1 - \phi_0 S} \times \frac{1 + a}{3 + 2a}, \quad (2-23)$$

where

$$a = \frac{4\gamma}{3\kappa}. \quad (2-24)$$

Integrating (2-23) and (2-24) over κ from k_S to κ_G , over a from $a_S = \frac{4g_S}{3k_S}$ to $a_G = \frac{4g_G}{3k_G}$, and over S from 0 to $S_L = 1 - S_G$, we obtain

$$\frac{\kappa_G}{k_S} = \frac{|1 - a_G|^{5/3} a_S}{|1 - a_S|^{5/3} a_G}; \quad \frac{|1 - a_G|^5}{1 + a_G} = \frac{|1 - a_S|^5}{1 + a_S} (1 - \phi_0 S_L)^6, \quad (2-25)$$

while the expression for the shear modulus γ_G is known through the relation $\gamma_G = 3\kappa_G a_G/4$. The overall properties are then determined by simultaneously letting $\kappa_G = K$ and $a_G = 4G/3K$ in (2-25)₁ and letting $a_G = 4G/3K$ and $S_L = 1$ in (2-25)₂. For a given value of $S_L = 1 - S_G$ the unsaturated poroelastic properties b_J and N_{JK} are then derived by combining relations (2-25), (2-15) and (2-17).

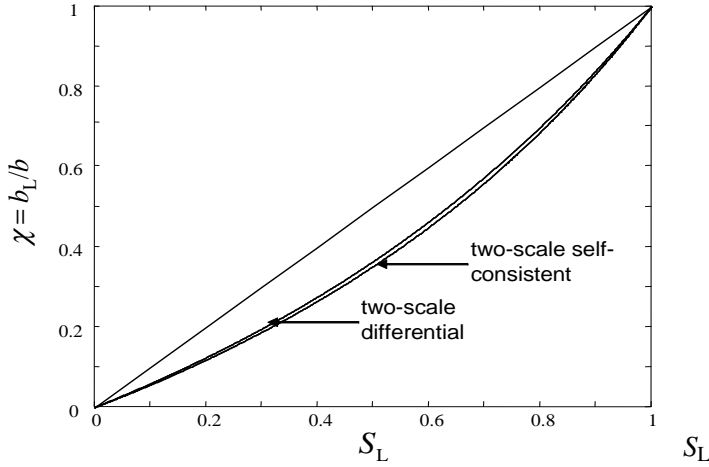


Figure 3. Bishop parameter $\chi = b_L/b$ versus liquid saturation S_L under three homogenization schemes. For the two-step self-consistent and differential schemes, χ departs from S_L , showing that pores of different sizes do not deform equally under these schemes.

Although no simple expression can be obtained when χ is defined by (2-1), as was the case for the previous two-scale self-consistent scheme, one can show that small and large pores do not undergo the same deformation when subjected to the same pressure; yet, in contrast with Mori–Tanaka’s approach, (2-4) no longer holds. Figure 3 plots $\chi = b_L/b$ against S_L for the Mori–Tanaka scheme ($\chi = b_L/b = S_L$), for the two-scale self-consistent scheme and for the differential scheme. For the two-scale self-consistent and differential schemes the ratio $\chi = b_L/b$ exhibits a significant lower value than the liquid saturation S_L , which is the expression of b_L/b associated with pore isodeformation. In Figure 4 we plot the proelastic

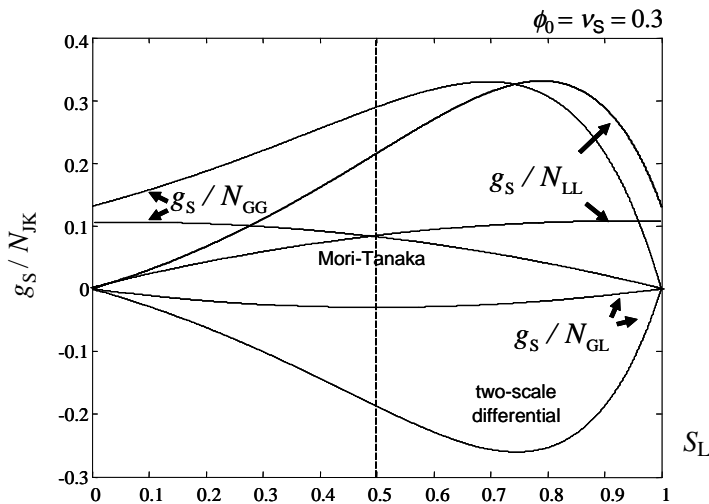


Figure 4. Poroelastic coupling property N_{JK} normalized by solid matrix shear modulus g_S versus liquid saturation for Mori–Tanaka and two-scale differential schemes.

coupling properties g_S/N_{JK} against S_L for the Mori–Tanaka scheme and the differential scheme. In the differential scheme, the coupling properties g_S/N_{JK} are not symmetric with regard to the line $S_L = \frac{1}{2}$. The properties g_S/N_{LL} and g_S/N_{GL} reach their maximum value for larger values than $S_L = \frac{1}{2}$, and the opposite holds for g_S/N_{GG} .

3. Drying shrinkage

The drying history of a water-infiltrated porous material is driven by Kelvin's law

$$p_L - p_{\text{atm}} = \frac{RT}{v_L} \ln h_R, \quad (3-1)$$

where R is the constant of ideal gases, T the temperature and v_L the water molar volume; h_R is the relative humidity, that is, the ratio p_V/p_{VS} of the vapour pressure p_V and the saturating vapour pressure p_{VS} that would prevail over liquid water at atmospheric pressure. At equilibrium the air pressure within the porous material is atmospheric. With $p_{\text{atm}} = p_G$ Kelvin's law and (1-10) combine to give

$$S_L = S \left(-\frac{2v_L \gamma_{GL}/RT}{\ln h_R} \right). \quad (3-2)$$

This captures the fact that the pore entry radius, and consequently the liquid saturation, have to adjust to the current relative humidity in order for the confined liquid water to remain in thermodynamic equilibrium with the current vapour pressure imposed by the current relative humidity. This is accompanied by a depressurization of liquid water, which in turn provokes drying shrinkage.

Consider then a stress-free drying process so that $\sigma = 0$, starting from a reference initial state where the porous material is saturated ($S_L = 1$), the pore pressure is atmospheric ($p_L = p_{\text{atm}}$) and the relative humidity is 100%. With regard to a zero pore pressure state, the deformation ε_0 related to the drying initial state is provided by substituting the initial conditions in (1-19) with $b_L = b$ and $b_G = 0$ so that

$$\varepsilon_0 = \frac{bp_{\text{atm}}}{K}. \quad (3-3)$$

When the relative humidity is lowered below 100% and the gas pressure is maintained at atmospheric pressure p_{atm} a drying shrinkage $\varepsilon_{\text{drying}} = \varepsilon - \varepsilon_0$ is observed and an associated extra elastic energy W_{drying} is stored whose respective intensity are obtained by substituting $\sigma = 0$, (2-1), (3-1) and (3-3) in (1-19) and (1-18). We get

$$\varepsilon_{\text{drying}} = -\frac{b\chi}{K} (p_G - p_L) = \frac{b\chi}{K} \frac{RT}{v_L} \ln h_R; \quad W_{\text{drying}} = \left(\frac{b^2\chi^2}{2K} + \frac{1}{2N_{LL}} \right) \left(\frac{RT}{v_L} \ln h_R \right)^2. \quad (3-4)$$

Under a specific homogenization scheme, the poroelastic properties b , K , χ and N_{LL} are then known as functions of the porosity ϕ_0 , the current liquid saturation S_L and the matrix elastic properties k_S and g_S . In addition the current liquid saturation S_L is known through (3-2) as a function of the current relative humidity h_R and the pore size distribution. Adopting for the latter the data reported in Figure 1, we have plotted $\varepsilon_{\text{drying}}$ and W_{drying} against S_L in Figure 5, top, for the various homogenization schemes explored in this paper. Drying shrinkage exhibits a maximum in absolute value, achieved when the decrease in liquid pressure induced by the decrease in relative humidity is exactly compensated by a decrease in the

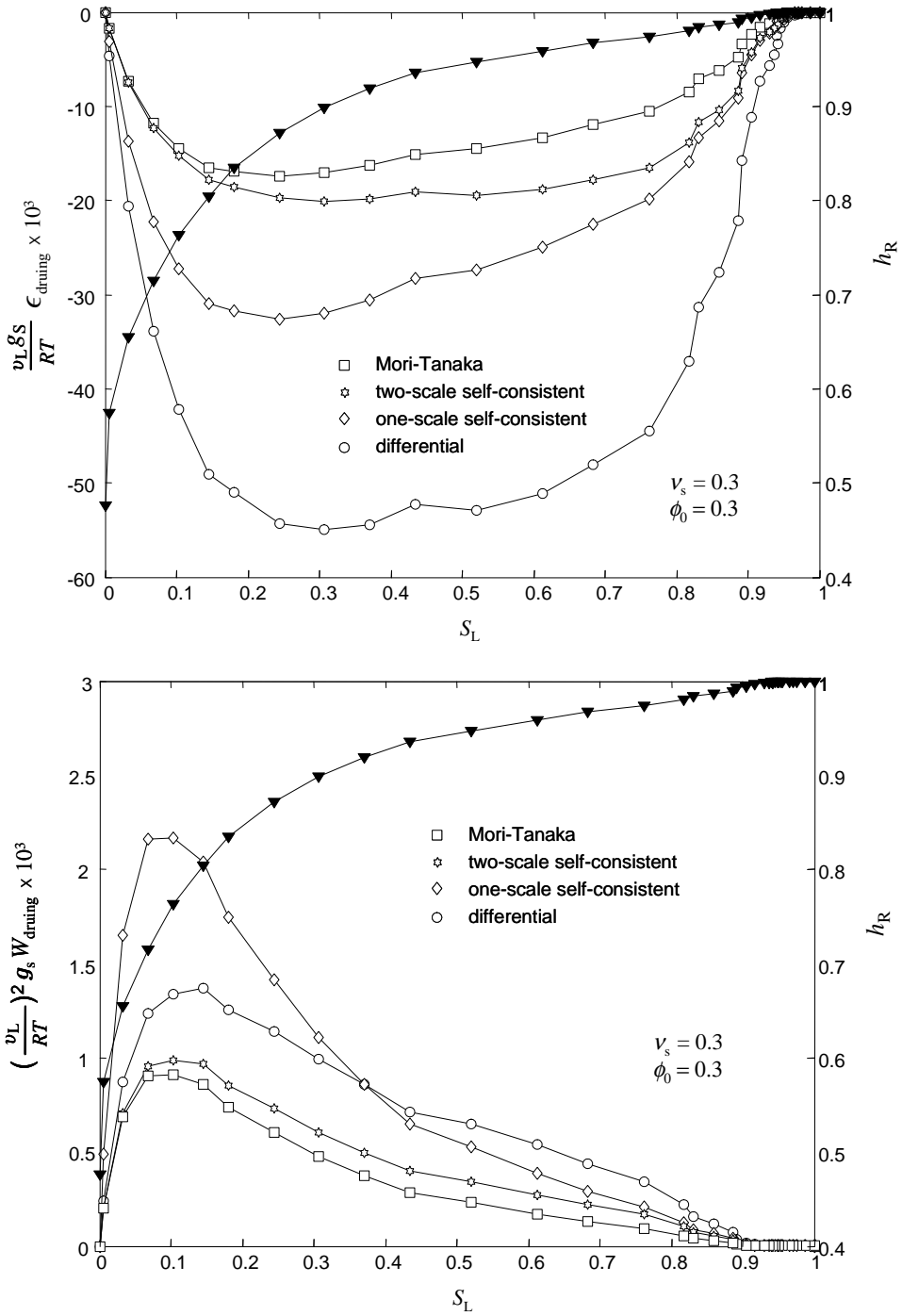


Figure 5. Normalized drying shrinkage (top) and elastic energy stored in the solid matrix during drying (bottom) versus liquid saturation, under various homogenization schemes. The cumulative volume fraction is assumed to depend on the pore entry radius as in Figure 1.

still wetted porous volume. In view of (3-4), this maximal drying shrinkage occurs when

$$\frac{p_G - p_L}{\chi} = -\frac{d}{d\chi} (p_G - p_L). \quad (3-5)$$

As shown in Figure 6, this condition allows the graphical determination of the capillary pressure associated with maximal drying shrinkage.

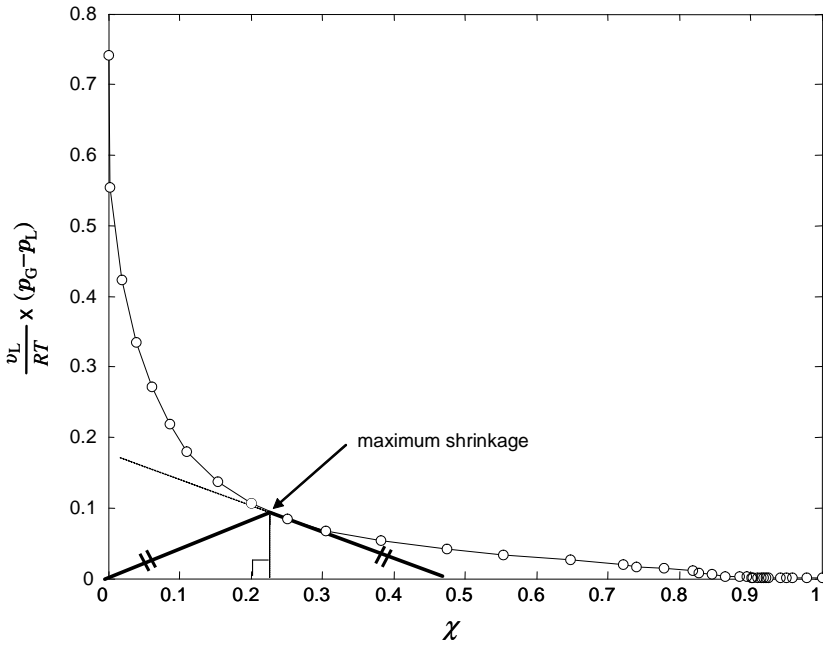


Figure 6. Graphical determination of the capillary pressure associated with maximal drying shrinkage under the differential homogenization scheme. The cumulative volume fraction is assumed to depend on the pore entry radius as in Figure 1.

As seen in the top half of Figure 5 the drying shrinkage predicted with the two-scale self-consistent scheme is less significant than the one predicted with the one-scale self-consistent scheme. Actually the bulk modulus K is the same for the two schemes while b_L is smaller for the two-scale self-consistent scheme (see Figure 7). In the bottom half of Figure 5 we see that the homogenization scheme affects significantly the elastic energy stored in the solid matrix during the drying process. As a result the difference of deformation of pores having a different size can significantly affect the strength of a porous material subjected to drying if its fracture is brittle and governed by a threshold in the stored elastic energy.

4. Discussion

The preceding results show that different homogenization schemes can dramatically affect the numerical values of macroscopic shrinkage and the elastic energy stored in the material. The strength of a porous material subjected to drying (if its fracture is brittle, and governed by a threshold in stored elastic energy)

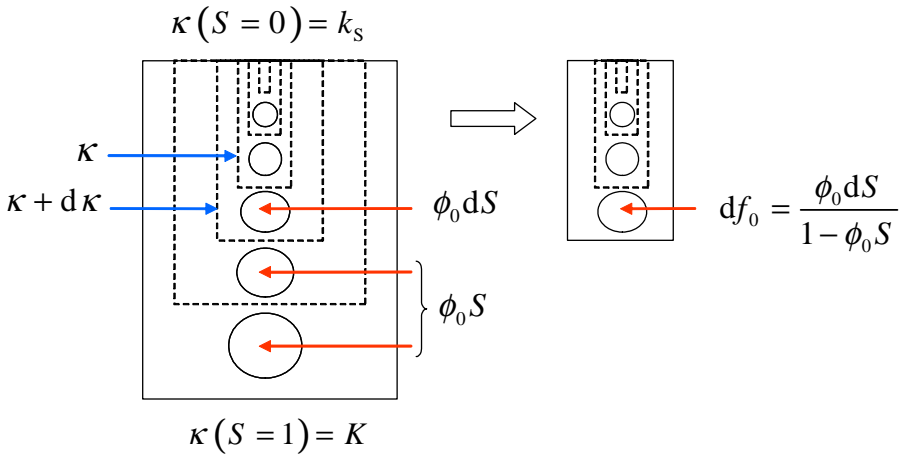


Figure 7. Sketch of the differential homogenization scheme for determining unsaturated poroelastic properties.

could therefore be significantly affected as well. This should not come as a surprise, since the choice of a homogenization scheme is already paramount to the values of the effective moduli of a composite medium. Whether neglecting pore size distribution leads to an over- or underestimation of macroscopic shrinkage is not clear, as can be seen in Figure 5; in fact, the numerical values found in the present study should not be taken for granted.

Indeed, it should be emphasized that the pore-size distribution has only been taken into account very crudely in this paper, and the assumption has been made that scale separation prevails between gas- and liquid-filled pores, *at each stage of the drying process* (i.e., for all values of liquid saturation). This strong assumption cannot be true of a continuous pore size distribution, since, as already stated, the largest liquid-filled pores are of size comparable with the smallest gas-filled pores. Even if the experimental data used in this paper (Figure 1) only imply continuity of the pore entry radii distribution (not the pore-size distribution), it is in fact well-known that the pore-size distribution is indeed continuous in cementitious materials, which effectively rules out our assumption.

Finally, it might be rightly argued that some of the homogenization schemes presented here were used outside their well-documented range of applicability. Although both the Mori–Tanaka and the differential schemes were designed for composites with bound inclusions as considered here, the self-consistent estimate is usually associated with polycrystals [Kröner 1977]; the application of this scheme to cementitious materials should therefore be considered with care. Regarding the differential scheme, as has been argued elsewhere [Norris 1985], the pore-size distribution implicitly taken into account is that of a large number of well separated families of spherical pores, which is not in contradiction with the assumption made previously, but for the fact that the basic differential scheme used here requires each family of pores to represent the same volume fraction.

The situation clearly calls for a clarification of the subtle effects of the pore-size distribution on the macroscopic properties of a porous medium. Although some attempts have been made towards this end [Bilger et al. 2007], this is still, to the authors’ knowledge, an open question.

5. Concluding remarks

Unsaturated poroelasticity and homogenization schemes have been combined to reveal the effects of nonuniform pore deformation upon the mechanical behaviour of a porous material subjected to drying. This analysis can easily be extended to other confined phase transitions such as freezing [Coussy 2005; Coussy and Monteiro 2007] or drying-induced crystallization of sea salts [Coussy 2006].

However, the present analysis is based upon the assumption that scale separation prevails between the liquid- and gas-invaded pores. This assumption is in contradiction with the fact that the size of the largest pores occupied by the liquid is comparable with the size of the smallest pores occupied by the gas. This study should therefore only be considered as a first attempt at taking the pore-size distribution into account for the estimation of macroscopic shrinkage. The reliability of this estimate would greatly benefit from an upscaling method explicitly integrating pore-size distribution.

In addition, the internal stresses generated by the drying process might induce progressive cracking of the solid matrix. This in turn could significantly alter the conclusions drawn from a reversible, poroelastic, analysis. Further research is thereby needed to assess the effects of the size of the pores upon the ultimate strength of porous materials subjected to confined phase transitions, microporomechanics [Dormieux et al. 2006b] being the appropriate tool for this issue.

References

- [Baroghel-Bouny et al. 1999] V. Baroghel-Bouny, M. Mainguy, T. Lassabatère, and O. Coussy, “Characterization and identification of equilibrium and transfer moisture properties for ordinary and high-performance cementitious materials”, *Cement Concrete Res.* **29**:8 (1999), 1225–1238.
- [Bazant and Wittmann 1982] Z. Bazant and F. Wittmann (editors), *Mathematical modeling of creep and shrinkage of concrete*, Wiley, New York, 1982.
- [Bilger et al. 2007] N. Bilger, F. Auslender, M. Bornert, H. Moulinec, and A. Zaoui, “Bounds and estimates for the effective yield surface of porous media with a uniform or a nonuniform distribution of voids”, *Eur. J. Mech. A Solids* **26**:5 (2007), 810–836.
- [Bishop and Blight 1963] A. W. Bishop and G. E. Blight, “Some aspects of effective stress in saturated and partly saturated soils”, *Géotechnique* **13** (1963), 177–197.
- [Chateau and Dormieux 2002] X. Chateau and L. Dormieux, “Micromechanics of saturated and unsaturated porous media”, *Int. J. Numer. Anal. Methods Geomech.* **26**:8 (2002), 831–844.
- [Chateau et al. 2003] X. Chateau, L. Dormieux, and Y. Xu, “Influence of geometry changes on drying-induced strains in a cracked solid”, *C. R. Mécanique* **331**:10 (2003), 679–686.
- [Chertkov 2002] V. Y. Chertkov, “Modelling cracking stages of saturated soils as they dry and shrink”, *Eur. J. Soil Sci.* **53**:1 (2002), 105–118.
- [Coussy 1991] O. Coussy, *Mécanique des milieux poreux*, Technip, Paris, 1991. In French; translated as *Mechanics of porous continua*, Wiley, New York, 1995.
- [Coussy 2004] O. Coussy, *Poromechanics*, Wiley, Chichester, 2004.
- [Coussy 2005] O. Coussy, “Poromechanics of freezing materials”, *J. Mech. Phys. Solids* **53**:8 (2005), 1689–1718.
- [Coussy 2006] O. Coussy, “Deformation and stress from in-pore drying-induced crystallization of salt”, *J. Mech. Phys. Solids* **54**:8 (2006), 1517–1547.
- [Coussy 2007] O. Coussy, “Revisiting the constitutive equations of unsaturated porous solids using a Lagrangian saturation concept”, *Int. J. Numer. Anal. Methods Geomech.* **31**:15 (2007), 1675–1694.
- [Coussy and Monteiro 2007] O. Coussy and P. Monteiro, “Unsaturated poroelasticity for crystallization in pores”, *Comput. Geotech.* **34**:4 (2007), 279–290.

- [Coussy et al. 1998] O. Coussy, R. Eymard, and T. Lassabatère, “Constitutive modeling of unsaturated drying deformable materials”, *J. Eng. Mech. (ASCE)* **124**:6 (1998), 658–667.
- [Dormieux et al. 2006a] L. Dormieux, D. Kondo, and F.-J. Ulm, *Microporomechanics*, Wiley, Chichester, 2006.
- [Dormieux et al. 2006b] L. Dormieux, J. Sanahuja, and S. Maghous, “Influence of capillary effects on strength of non-saturated porous media”, *C. R. Mécanique* **334**:1 (2006), 19–24.
- [Gusnard and Kirschner 1977] D. Gusnard and R. H. Kirschner, “Cell and organelle shrinkage during preparation for scanning electron-microscopy: effects of fixation, dehydration and critical-point drying”, *J. Microsc.* **110** (1977), 51–57.
- [Huang and Feldman 1985] C.-y. Huang and R. F. Feldman, “Dependence of frost resistance on the pore structure of mortar containing silica fume”, *ACI J.* **82**:5 (1985), 740–743.
- [Hutter et al. 1999] K. L. Hutter, L. Laloui, and L. Vulliet, “Thermodynamically based mixture models of saturated and unsaturated soils”, *Mech. Cohes. Frict. Mater.* **4**:4 (1999), 295–338.
- [Kozłowski and Pallardy 2002] T. T. Kozłowski and S. G. Pallardy, “Acclimation and adaptive responses of woody plants to environmental stresses”, *Bot. Rev.* **68**:2 (2002), 270–334.
- [Kröner 1977] E. Kröner, “Bounds for effective elastic moduli of disordered materials”, *J. Mech. Phys. Solids* **25**:2 (1977), 137–155.
- [Lewis and Schrefler 1998] R. W. Lewis and B. A. Schrefler, *The finite element method in the static and dynamic deformation and consolidation of porous media*, 2nd ed., Wiley, Chichester, 1998.
- [Mainguy et al. 2001] M. Mainguy, O. Coussy, and V. Baroghel-Bouny, “Role of air pressure in drying of weakly permeable materials”, *J. Eng. Mech. (ASCE)* **127**:6 (2001), 582–592.
- [Norris 1985] A. N. Norris, “A differential scheme for the effective moduli of composites”, *Mech. Mater.* **4**:1 (1985), 1–16.
- [Ratti 1994] C. Ratti, “Shrinkage during drying of foodstuffs”, *J. Food Eng.* **23**:1 (1994), 91–105.
- [Santos 2000] J. A. Santos, “Mechanical behaviour of Eucalyptus wood modified by heat”, *Wood Sci. Technol.* **34**:1 (2000), 39–43.
- [Smith et al. 1995] D. M. Smith, G. W. Scherer, and J. M. Anderson, “Shrinkage during drying of silica gel”, *J. Non-Cryst. Solids* **188**:3 (1995), 191–206.

Received 14 Dec 2007. Accepted 8 Nov 2008.

OLIVIER P. COUSSY: Olivier.Coussy@mail.enpc.fr

Université Paris-Est, UR Navier, École des Ponts, 6-8 Av. Blaise Pascal, Cité Descartes, F-77455 Marne-la-Vallée Cedex 2, France

SÉBASTIEN BRISARD: Sebastien.Brisard@mail.enpc.fr

Université Paris-Est, UR Navier, École des Ponts, 6-8 Av. Blaise Pascal, Cité Descartes, F-77455 Marne-la-Vallée Cedex 2, France

ON THE METHOD OF VIRTUAL POWER IN CONTINUUM MECHANICS

GIANPIETRO DEL PIERO

Dedicated to the memory of Paul Germain

The method of virtual power is generally used to produce balance equations for nontraditional continua such as continua with various types of microstructure. Here I show that the expression of the internal power can be deduced from that of the external power using a general invariance requirement due to Noll and a generalized version of Cauchy's tetrahedron theorem. In other words, the measures of deformation and stress, as well as the balance equations, are determined by the expression chosen for the external power and by the invariance assumptions. A pair of examples taken from the literature shows that both ingredients are essential for defining a specific class of continua.

1. Introduction

In classical continuum mechanics, the balance laws of linear and angular momentum determine an integral identity, the equation of virtual power. On this identity is based the weak formulation of the problems of motion and of equilibrium. A more recent choice is to consider the equation of virtual power as primitive, and to deduce from it the balance equations. A reason for this alternative approach is the difficulty met in formulating generalized versions of the balance equations, appropriate to nonclassical continua. The method of virtual power of [Germain \[1972\]](#) is just a formalization of this new approach, and his papers [[Germain 1973a](#); [1973b](#)] show how the method applies to some specific nonclassical continua, such as second-gradient and micropolar continua.

To describe nonclassical continua within the classical approach is possible, at the price of introducing supplementary balance equations. For example [Capriz \[1989\]](#) introduces just one supplementary equation, the equation of balance of micromomentum, and with it he succeeds in describing a variety of microstructures.

Both approaches have some inconvenience. When introducing new balance equations, it is not clear how many they should be, and which should be their motivation. For example, [Capriz \[1989, Section 8\]](#) says that the supplementary equation is just a plausible form for the balance of micromomentum, justified by the analogy with the form of the balance equation of linear momentum. Clearly, we are far from the status of fundamental laws of mechanics attributed to the classical balance equations.

Even worse, when some form of the equation of virtual power is assumed as primitive, the expressions of the external and internal power are not really arbitrary, due to the general belief that the classical

Keywords: foundations of continuum mechanics, virtual power, continua with microstructure, materials with microstructure. This research has been supported by the Research Project *Mathematical Models for Materials Science* — PRIN 2005 of the Italian Ministry for University and Scientific Research.

balance equations should hold anyway. Thus, there is a tacit preselection of the possible forms of the equation, in spite of its claimed status of a postulate.

Here I follow a different approach, first proposed by Noll [1974] in the context of classical continua. This approach is based on two assumptions: one on the nature of the external actions, and one on the invariance of the power under changes of observer. The first assumption determines the expression of the external power, and the two assumptions together determine the balance equations. With the aid of the divergence theorem, one obtains an expression of the internal power, which consists of a volume integral involving the inner products of a certain number of internal forces by the corresponding generalized deformations.¹

In this way, the two basic assumptions determine a class of continua. Within each class, specific materials can be characterized by ad hoc constitutive assumptions, such as the existence of energy functions, of dissipation potentials, or of explicit relations between internal forces and generalized deformations. Both assumptions are essential in determining a class of continua. Indeed, the same form of the external power can be associated with different invariance requirements, dictated by the differences in the physical nature of the microstructures. This determines different classes of continua, each one with its own set of balance equations. An example is provided in Section 5.

Besides the genuine balance equations coming from the invariance requirements, it is sometimes convenient to introduce pseudo-balance equations, which are just mathematical devices transforming area integrals into volume integrals. Though in the literature the two are frequently mixed together, it is important to keep them separated. A clear distinction between genuine balance equations and pseudo-balance equations is one of the purposes of the present work.

Section 2 shows the application of the proposed method to classical continua. In the next sections 3 and 4 the same method is applied to higher-gradient continua and to micropolar continua, respectively. The last section deals with two examples taken from strain-gradient plasticity. In them, the expression of the external power is the same as in a micropolar continuum, while the invariance assumptions are different.

2. Classical continua

For a classical continuum occupying a three-dimensional region Ω of space, the assumed system of external actions is a pair (b, s) formed by distance actions and contact actions, where $b = b(x)$ is a vector field on Ω representing the volume density of the body forces, and $s = s(\Pi, x)$ is a system of vector fields acting at the boundary points x of each part Π of Ω , and representing the surface density of the contact force at x .² The corresponding expression for the external power is

$$\mathcal{P}(\Pi, v) = \int_{\Pi} b \cdot v \, dV + \int_{\partial\Pi} s \cdot v \, dA, \quad (1)$$

¹Instead of making assumptions on the nature of the external actions, one might fix a priori the structure of the set \mathcal{V} of the virtual displacements, and then define the external power by duality, that is, by identifying it with the most general continuous linear functional on \mathcal{V} . Whether or not to consider forces as primitive is a philosophical matter which has long been debated, see Jammer [1957]. Whatever is the preference, the procedure proposed here applies.

²Throughout this paper it is assumed that all vector, or tensor, fields defined on the pair (Π, x) are bounded almost everywhere, and that the associated fluxes $\int_{\partial\Pi} s \, dA$ are additive on regions Π with pairwise disjoint interiors [Gurtin and Martins 1976].

where $v = v(x)$ is a field of virtual displacements³ on Ω , and dV and dA are the volume measure and the area measure, respectively.

The assumption that \mathcal{P} is invariant under changes of observer is expressed by the equation

$$\mathcal{P}(\Pi, v) = \mathcal{P}(\Pi, v + c + \omega \times x), \quad (2)$$

to be satisfied for all parts Π of Ω and for all pairs of vectors c, ω . In view of the linear dependence of \mathcal{P} on v , this is true if and only if

$$\mathcal{P}(\Pi, c) = \mathcal{P}(\Pi, \omega \times x) = 0 \quad (3)$$

for all Π and for all vectors c, ω . That is, if and only if

$$\int_{\Pi} b \, dV + \int_{\partial\Pi} s \, dA = 0, \quad \int_{\Pi} x \times b \, dV + \int_{\partial\Pi} x \times s \, dA = 0, \quad (4)$$

for all Π . These are the balance equations of the linear and angular momentum for a classical continuum.⁴

It has been proven by Noll [1959] that, as a consequence of the first balance equation, s depends on Π only through the exterior unit normal n to $\partial\Pi$ at x : $s(\Pi, x) = s(n, x)$. Then from Cauchy's tetrahedron theorem one deduces the linearity of the dependence on n , that is, the existence of a tensor T , the Cauchy stress, such that $s(n, x) = T(x)n$, as well as the local forms of the two balance equations

$$\operatorname{div} T + b = 0, \quad T = T^T. \quad (5)$$

Substituting into the expression of the external power and using the divergence theorem, one gets

$$\begin{aligned} \mathcal{P}(\Pi, v) &= - \int_{\Pi} \operatorname{div} T \cdot v \, dV + \int_{\partial\Pi} T n \cdot v \, dA \\ &= \int_{\Pi} T \cdot \nabla^s v \, dV, \end{aligned}$$

where $\nabla^s v$ is the symmetric part of ∇v . The integral on the right is the internal power. In it, the integrand function is the product of an internal force, the Cauchy stress, by a virtual generalized deformation, the symmetric part of the gradient of the virtual displacement. As we see, in the proposed approach there are not indeed two virtual powers, internal and external, but rather a single virtual power with two different expressions, the first of which is given a priori, while the second is deduced from the first, using the invariance assumptions.

³The fields v are usually called *virtual velocities*. I prefer *displacements*, since *velocity* suggests that physical time is somehow involved. For consistency, I should then speak of *virtual work* instead of *virtual power*, but the latter term is so generally used that I don't dare go back to the more classical *virtual work*.

⁴In the past, an obstacle against this very simple deduction of the balance equations was that the power of the inertial actions is not indifferent under Galilean changes of observer. Following ideas originating from Mach's criticism of the Newtonian concept of absolute space, Noll [1974] ruled out this difficulty by regarding inertia as the distance action "between the bodies in the solar system and the masses occupying the rest of the universe" [Truesdell and Noll 1965, Sect. 18]. Accordingly, the power of the inertial actions has to be invariant under changes of observer involving the whole universe, and not just a part of it, as it is tacitly done when referring the change of observer to our planet or to a part of it, instead of to the whole universe.

3. Higher-gradient continua

Let us add to the virtual power (1) two terms in ∇v

$$\mathcal{P}(\Pi, v) = \int_{\Pi} (b \cdot v + B \cdot \nabla v) dV + \int_{\partial\Pi} (s \cdot v + S \cdot \nabla v) dA, \quad (6)$$

one associated with a second-order tensor field $B = B(x)$ of body double-forces, and one with a system $S = S(\Pi, x)$ of second-order tensor fields representing surface double-tractions. From the divergence theorem

$$\int_{\Pi} B \cdot \nabla v dV = - \int_{\Pi} \operatorname{div} B \cdot v dV + \int_{\partial\Pi} B n \cdot v dA. \quad (7)$$

it follows that the power generated by a field of body double-forces is equal to the power generated by a field $b^* = -\operatorname{div} B$ of body forces, plus a system $s^* = B n$ of surface tractions. Because the term $B \cdot \nabla v$ does not change substantially the expression of the virtual power, for simplicity it will be neglected in what follows.

For the power generated by S , we first observe that the invariance assumption (3)₁ still yields the balance equation (4)₁, so Equation (6) reduces to

$$\mathcal{P}(\Pi, v) = \int_{\Pi} T \cdot \nabla v dV + \int_{\partial\Pi} S \cdot \nabla v dA. \quad (8)$$

Then from the invariance assumption (3)₂ one gets

$$\int_{\Pi} T^w dV + \int_{\partial\Pi} S^w dA = 0, \quad (9)$$

with T^w and S^w the skew-symmetric parts of T and S , respectively. This equation is formally identical with Equation (4)₁, except for the fact that the integrand functions are now second-order tensors. Using Noll's theorem on the dependence on the normal and Cauchy's tetrahedron theorem⁵ one deduces the existence of a third-order tensor \mathbb{T}^w , skew-symmetric with respect to the first two indices, such that

$$S^w = \mathbb{T}^w n, \quad S_{ij}^w = \mathbb{T}_{ijk}^w n_k. \quad (10)$$

Substituting Equation (10) into the balance equation (9) and then using the divergence theorem one gets

$$\operatorname{div} \mathbb{T}^w + T^w = 0, \quad \mathbb{T}_{ijk,k}^w + T_{ij}^w = 0. \quad (11)$$

Therefore,

$$\begin{aligned} \int_{\partial\Pi} S^w \cdot \nabla v dA &= \int_{\partial\Pi} \mathbb{T}^w n \cdot \nabla v dA \\ &= \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v + \operatorname{div} \mathbb{T}^w \cdot \nabla v) dV \\ &= \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v - T^w \cdot \nabla v) dV, \end{aligned}$$

⁵For an extension of Cauchy's theorem to second-order tensors see Appendix A.

and the expression (8) of the virtual power reduces to

$$\mathcal{P}(\Pi, v) = \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v + T^S \cdot \nabla v) dV + \int_{\partial \Pi} S^S \cdot \nabla v dA. \tag{12}$$

As we see, the right-hand side has not yet the form of a single volume integral, except in the special case of S skew-symmetric, which will be considered separately at the end of this section. To transform the area integral into a volume integral, we introduce the field of the body forces \tilde{B} defined by

$$\tilde{B}(x) = - \lim_{\delta \rightarrow 0} \frac{\int_{\partial \mathcal{B}(x, \delta)} S^S dA}{V(\mathcal{B}(x, \delta))}, \tag{13}$$

where $\mathcal{B}(x, \delta)$ is the three-dimensional ball of radius δ centered at x , and \tilde{B} is symmetric because S^S is symmetric.⁶ By integration over a part Π of the body, we get the pseudo-balance equation

$$\int_{\Pi} \tilde{B} dV + \int_{\partial \Pi} S^S dA = 0. \tag{14}$$

I point out that this is not a proper balance equation, since it does not come from an invariance assumption. It comes instead from the definition (13), which is instrumental to transforming a surface integral into a volume integral. In fact, again from Noll’s and Cauchy’s theorems one deduces the existence of a third-order tensor \mathbb{T}^S , symmetric with respect to the first two indices, such that $S^S = \mathbb{T}^S n$, and substituting into the pseudo-balance equation and using the divergence theorem, the local pseudo-balance equation

$$\operatorname{div} \mathbb{T}^S + \tilde{B} = 0 \tag{15}$$

follows. Then the area integral in (12) transforms as

$$\begin{aligned} \int_{\partial \Pi} S^S \cdot \nabla v dA &= \int_{\partial \Pi} \mathbb{T}^S n \cdot \nabla v dA \\ &= \int_{\Pi} (\mathbb{T}^S \cdot \nabla \nabla v + \operatorname{div} \mathbb{T}^S \cdot \nabla v) dV \\ &= \int_{\Pi} (\mathbb{T}^S \cdot \nabla \nabla v - \tilde{B} \cdot \nabla v) dV, \end{aligned}$$

and after setting

$$\mathbb{T} := \mathbb{T}^S + \mathbb{T}^w, \tag{16}$$

one has $S = \mathbb{T} n$, $\operatorname{div} \mathbb{T} + T^w + \tilde{B} = 0$, and the virtual power (12) takes the form

$$\mathcal{P}(\Pi, v) = \int_{\Pi} (\mathbb{T} \cdot \nabla \nabla v + (T^S - \tilde{B}) \cdot \nabla v) dV. \tag{17}$$

We see that the internal forces are the third-order tensor \mathbb{S} and the second-order tensor $T^S - \tilde{B}$, and that the generalized deformations are the first two gradients of v . If we agree to label a class of continua after the highest gradient of v which appears among the generalized deformations, then the classical

⁶For the regularity assumptions on S^S ensuring the existence almost everywhere of the limit \tilde{B} , see Gurtin and Martins [1976, Theorem 7].

continuum is a first-gradient continuum, and the continuum defined by (6) and (3) is a second-gradient continuum.

A special subclass of second-gradient continua is the one in which the external action S is skew-symmetric. As said before, in this case the area term in (12) vanishes, and therefore there is no pseudo-balance equation. Moreover, the power $S \cdot \nabla v$ can be given the form $m \cdot \omega$, where m and ω are the vectors associated with S and with twice the skew-symmetric part of ∇v , respectively:

$$m_i = \frac{1}{2} e_{ijk} S_{kj}, \quad \omega_i = e_{ijk} v_{k,j}. \tag{18}$$

Because ω measures the local rotation at x , the vector m can be identified with a surface couple. Then, since S skew-symmetric implies $\mathbb{T} = \mathbb{T}^w$ and $\mathbb{T}^s = 0$, the first product in (17) takes the form

$$\mathbb{T} \cdot \nabla \nabla v = \mathbb{T}^w \cdot \nabla \nabla v = M \cdot \nabla \omega, \quad \mathbb{T}_{ijk} v_{i,jk} = M_{ij} e_{ihk} v_{h,kj} = M_{ij} \omega_{i,j}, \tag{19}$$

where M is the second-order tensor associated with \mathbb{T}^w ⁷

$$M_{ij} = \frac{1}{2} e_{ihk} \mathbb{T}_{khj}^w. \tag{20}$$

In conclusion, for S skew-symmetric the two expressions (6), (17) of the virtual power simplify into

$$\int_{\Pi} b \cdot v \, dV + \int_{\partial \Pi} (s \cdot v + m \cdot \omega) \, dV, \quad \int_{\Pi} (M \cdot \nabla \omega + T^s \cdot \nabla v) \, dV. \tag{21}$$

The foregoing analysis can be easily generalized to higher-gradient continua. Notice that for all such continua there are no invariance assumptions besides the classical assumptions (3).

4. Micropolar continua

In micropolar continua, with each point x is associated a finite set of vectors $\alpha \mapsto d^\alpha = d^\alpha(x)$, called the *directors*, each with fixed length and variable orientation. The virtual power consists of the two terms appearing in Equation (1) plus two extra terms, representing the power of the body and surface director forces $\beta^\alpha, \sigma^\alpha$, both multiplied by virtual changes v^α of the directors ⁸

$$\mathcal{P}(\Pi, v, v^\alpha) = \int_{\Pi} (b \cdot v + \beta^\alpha \cdot v^\alpha) \, dV + \int_{\partial \Pi} (s \cdot v + \sigma^\alpha \cdot v^\alpha) \, dA. \tag{22}$$

While the first of the invariance axioms (3) remains unchanged, the second now requires the invariance of the virtual power under simultaneous rigid rotations of the body and of the directors. That is,

$$\mathcal{P}(\Pi, c, 0) = \mathcal{P}(\Pi, \omega \times x, \omega \times d^\alpha) = 0, \tag{23}$$

⁷In view of Equation (18)₁, M can be seen as the unique second-order tensor such that Me is the vector associated with $\mathbb{T}^w e$ for all vectors e .

⁸Here and in the following, the repeated indices α are summed.

for all Π and for all vectors c, ω . The corresponding balance equations are

$$\begin{aligned} \int_{\Pi} b \, dV + \int_{\partial\Pi} s \, dA &= 0, \\ \int_{\Pi} (x \times b + d^\alpha \times \beta^\alpha) \, dV + \int_{\partial\Pi} (x \times s + d^\alpha \times \sigma^\alpha) \, dA &= 0, \end{aligned} \quad (24)$$

respectively. From the first equation still follows the existence of a Cauchy stress T such that $s = Tn$ and $\operatorname{div} T + b = 0$, and by substitution into the second equation one gets

$$\int_{\Pi} (T^w + (\beta^\alpha \otimes d^\alpha)^w) \, dV + \int_{\partial\Pi} (\sigma^\alpha \otimes d^\alpha)^w \, dA = 0. \quad (25)$$

Using again Noll's theorem on the dependence of the normal and Cauchy's tetrahedron theorem, one deduces the existence of a third-order tensor field \mathbb{T}_{ijk}^w , skew-symmetric with respect to the first two indices, such that $(\sigma^\alpha \otimes d^\alpha)^w = \mathbb{T}^w n$, from which one gets the following local form of the second balance equation

$$\operatorname{div} \mathbb{T}^w + T^w + (\beta^\alpha \otimes d^\alpha)^w = 0. \quad (26)$$

Then,

$$\begin{aligned} \int_{\Pi} T^w \cdot \nabla v \, dV &= - \int_{\Pi} (\operatorname{div} \mathbb{T}^w + (\beta^\alpha \otimes d^\alpha)^w) \cdot \nabla v \, dV \\ &= \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v - (\beta^\alpha \otimes d^\alpha)^w \cdot \nabla v) \, dV - \int_{\partial\Pi} (\sigma^\alpha \otimes d^\alpha)^w \cdot \nabla v \, dA, \end{aligned}$$

and the expression (22) of the virtual power transforms as

$$\begin{aligned} \mathcal{P}(\Pi, v, v^\alpha) &= \int_{\Pi} (T \cdot \nabla v + \beta^\alpha \cdot v^\alpha) \, dV + \int_{\partial\Pi} \sigma^\alpha \cdot v^\alpha \, dA \\ &= \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v + (T^s - (\beta^\alpha \otimes d^\alpha)^w) \cdot \nabla v + \beta^\alpha \cdot v^\alpha) \, dV \\ &\quad + \int_{\partial\Pi} (\sigma^\alpha \cdot v^\alpha - (\sigma^\alpha \otimes d^\alpha)^w \cdot \nabla v) \, dA \\ &= \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v + T^s \cdot \nabla v + \beta^\alpha \cdot (v^\alpha - \nabla^w v \, d^\alpha)) \, dV \\ &\quad + \int_{\partial\Pi} \sigma^\alpha \cdot (v^\alpha - \nabla^w v \, d^\alpha) \, dA. \end{aligned}$$

The vectors

$$\varphi^\alpha = v^\alpha - \nabla^w v \, d^\alpha \quad (27)$$

represent the local relative rotations between the directors and the body. As in the preceding section, to transform the area integral into a volume integral we introduce a pseudo-balance equation. We define

$$\tilde{\beta}^\alpha(x) = - \lim_{\delta \rightarrow 0} \frac{\int_{\partial B(x, \delta)} \sigma^\alpha \, dA}{V(B(x, \delta))}, \quad (28)$$

so that

$$\int_{\Pi} \tilde{\beta}^\alpha dV + \int_{\partial\Pi} \sigma^\alpha dA = 0. \tag{29}$$

By the tetrahedron theorem, for each α there is a second-order tensor Σ^α such that

$$\sigma^\alpha = \Sigma^\alpha n, \quad \text{div } \Sigma^\alpha + \tilde{\beta}^\alpha = 0. \tag{30}$$

Thus,

$$\begin{aligned} \int_{\partial\Pi} \sigma^\alpha \cdot \varphi^\alpha dA &= \int_{\partial\Pi} \Sigma^\alpha n \cdot \varphi^\alpha dA \\ &= \int_{\Pi} (\text{div } \Sigma^\alpha \cdot \varphi^\alpha + \Sigma^\alpha \cdot \nabla \varphi^\alpha) dV \\ &= \int_{\Pi} (-\tilde{\beta}^\alpha \cdot \varphi^\alpha + \Sigma^\alpha \cdot \nabla \varphi^\alpha) dV, \end{aligned}$$

and the virtual power takes the form

$$\mathcal{P}(\Pi, v, v^\alpha) = \int_{\Pi} (\mathbb{T}^w \cdot \nabla \nabla v + T^s \cdot \nabla v + \Sigma^\alpha \cdot \nabla \varphi^\alpha + (\beta^\alpha - \tilde{\beta}^\alpha) \cdot \varphi^\alpha) dV. \tag{31}$$

For this class of micropolar continua, the measures of stress are the tensors \mathbb{T}^w , T^s , Σ^α and $(\beta^\alpha - \tilde{\beta}^\alpha)$. They are subject to the balance equations (5)₁, (26) and to the pseudo-balance equation (30)₂. The measures of deformation are ∇v , φ^α and their first gradients. In particular, if the external actions β^α and σ^α are balanced, that is, if Equation (29) is satisfied by $\tilde{\beta}^\alpha = \beta^\alpha$ for all Π , the last term in Equation (31) disappears.

5. Two examples from strain-gradient plasticity

For continua described by internal state variables one can still assume a virtual power of the form (22), where now v^α are virtual variations of the state variables. They may be scalars, vectors, or tensors of any order, according to the nature of the corresponding state variable. The physical nature of the state variables also determines the appropriate invariance assumptions under changes of observer. Thus, in a sense, the invariance assumptions are a part of the definition of a state variable.

In strain-gradient plasticity, a configuration of a body is characterized by the deformation gradient F and by a state variable, the plastic distortion, identified with the plastic part of the strain gradient in its decomposition $F = F^e F^p$.⁹ A virtual variation of plastic distortion is a second-order tensor, which I will denote by L .¹⁰ The assumed expression of the virtual power is

$$\mathcal{P}(\Pi, v, L) = \int_{\Pi} b \cdot v dV + \int_{\partial\Pi} (s \cdot v + S \cdot L) dA. \tag{32}$$

It includes a contact action S associated with L , while the corresponding distance action is neglected. While for a micropolar continuum the virtual power is assumed to be invariant under simultaneous rigid

⁹Usually, a plastic continuum is regarded as a classical continuum subject to appropriate constitutive assumptions. Strain-gradient plasticity deals with a continuum with a microstructure described by the plastic distortion.

¹⁰It is usually supposed that L is traceless, but I shall ignore this restriction.

rotations of the body and of the directors, here the virtual power is assumed to be invariant under rigid body rotations, associated with a null change of plastic distortion.¹¹ Then we have the conditions

$$\mathcal{P}(\Pi, c, 0) = \mathcal{P}(\Pi, \omega \times x, 0) = 0, \tag{33}$$

and the resulting balance equations are the equations (4) of the classical continuum instead of the equations (24) of the micropolar continuum. But there is a third condition, which comes from the invariance of power under a rigid rotation of the intermediate configuration.¹² It requires that the virtual power associated with a rigid virtual rotation of plastic strain be zero

$$\mathcal{P}(\Pi, 0, W) = 0, \tag{34}$$

for all skew-symmetric tensors W .

From the first two invariance assumptions one again deduces the existence of a tensor field T such that $s = Tn$, $\text{div } T + b = 0$ and $T = T^T$, so that (32) becomes

$$\mathcal{P}(\Pi, v, L) = \int_{\Pi} T \cdot \nabla^s v \, dV + \int_{\partial\Pi} S \cdot L \, dA. \tag{35}$$

From (34) it follows that

$$\int_{\partial\Pi} S^W \, dA = 0, \tag{36}$$

and from here, using again Noll's and Cauchy's theorems, one deduces the existence of a third-order tensor field \mathbb{T}^W , skew-symmetric with respect to the first two indices, such that $S^W = \mathbb{T}^W n$ and $\text{div } \mathbb{T}^W = 0$. Therefore,

$$\int_{\partial\Pi} S^W \cdot L \, dA = \int_{\partial\Pi} \mathbb{T}^W n \cdot L \, dA = \int_{\Pi} \mathbb{T}^W \cdot \nabla L \, dV,$$

and the virtual power reduces to

$$\mathcal{P}(\Pi, v, L) = \int_{\Pi} (\mathbb{T}^W \cdot \nabla L + T \cdot \nabla^s v) \, dV + \int_{\partial\Pi} S^S \cdot L \, dA. \tag{37}$$

Once again, to eliminate the area integral a pseudo-balance equation is required. By defining a symmetric tensor \tilde{B} as in (13), we deduce the existence of a third-order tensor \mathbb{T}^S , symmetric with respect to the

¹¹A change of observer transforms the deformation gradient F into QF , with Q a proper rotation. The equation $QF = QF^e F^p$ then suggests that we choose the intermediate configuration determined locally by F^p in such a way that F^e and F^p transform into QF^e and F^p , respectively. With this choice, the intermediate configuration is left unchanged by a change of observer. Then $L = 0$, and the condition Equation (33)₂ follows. But there is nothing wrong in allowing the intermediate configuration to rotate, see the following footnote.

¹²A rigid rotation R of the intermediate configuration transforms F^p into RF^p , leaving F unchanged. This corresponds to $v(x) = 0$ and $L(x) = W$ constant and skew-symmetric. Because this is just a change of orientation of an auxiliary configuration, the power involved must be zero.

first two indices, such that $S^S = \mathbb{T}^S n$ and $\operatorname{div} \mathbb{T}^S + \tilde{B} = 0$. Then the area integral transforms as

$$\begin{aligned} \int_{\partial\Pi} S^S \cdot L \, dA &= \int_{\partial\Pi} \mathbb{T}^S n \cdot L \, dA \\ &= \int_{\Pi} (\mathbb{T}^S \cdot \nabla L + \operatorname{div} \mathbb{T}^S \cdot L) \, dV \\ &= \int_{\Pi} (\mathbb{T}^S \cdot \nabla L - \tilde{B} \cdot L^S) \, dV, \end{aligned} \quad (38)$$

and the virtual power takes the form

$$\mathcal{P}(\Pi, v, L) = \int_{\Pi} (\mathbb{T} \cdot \nabla L + T \cdot \nabla^S v - \tilde{B} \cdot L^S) \, dV, \quad (39)$$

with $\mathbb{T} = \mathbb{T}^S + \mathbb{T}^W$, $S = \mathbb{T}n$, and $\operatorname{div} \mathbb{T} + \tilde{B} = 0$. This is the only case considered in this paper, in which a third balance equation appears. As discussed above, the supplementary equation is motivated by the invariance of the virtual power under a rigid rotation of the intermediate configuration. The local form of this balance equation is $\operatorname{div} \mathbb{T}^W = 0$, while $\operatorname{div} \mathbb{T}^S + \tilde{B} = 0$ is the pseudo-balance equation used to transform the area integral in (37) into a volume integral.

Notice that if the external action S^S is self-balanced, that is, if Equation (14) is satisfied by $\tilde{B} = 0$ for every Π , the term $\tilde{B} \cdot L^S$ disappears from (39). This occurs, for example, if S is skew-symmetric. But in this case there is no pseudo-balance equation, and the virtual power (39) further reduces to

$$\mathcal{P}(\Pi, v, L) = \int_{\Pi} (\mathbb{T}^W \cdot \nabla L + T \cdot \nabla^S v) \, dV. \quad (40)$$

Let me discuss two examples taken from Gurtin and Anand [2005] and Gurtin [2004]. In the former, the authors assume the external power (32) and an expression of the internal power which, in the present notation, takes the form

$$\mathcal{P}_i(\Pi, v, L) = \int_{\Pi} (\mathbb{T} \cdot \nabla L + T \cdot \nabla^S v + (T^p - T) \cdot L) \, dV. \quad (41)$$

They also assume that L and T^p are symmetric, and that \mathbb{T} is symmetric with respect to the first two indices. By equating (32) and (41), they obtain the equation of virtual power, from which they deduce the classical balance equations (5), plus the relation $S = \mathbb{T}n$ and the microforce balance equation

$$\operatorname{div} \mathbb{T} + T - T^p = 0. \quad (42)$$

If we compare (39) and (41), we see that they coincide if $T^p - T = -\tilde{B}$, that is, if $\operatorname{div} \mathbb{T} + \tilde{B} = 0$. Due to the symmetry of \tilde{B} and \mathbb{T} , this is true if and only if our pseudo-balance equation $\operatorname{div} \mathbb{T}^S + \tilde{B} = 0$ holds. The expression (39) is slightly more general than (41), since it does not require the symmetry of L , T^p , and \mathbb{T} . These symmetry assumptions can be regarded as restrictions defining a special subclass of continua. This example shows that there is no need of assuming an expression for the internal power, since (41) follows from (32) under the appropriate invariance assumptions.

In [Gurtin 2004] the external power is taken as in (32) and the internal power is assumed to be

$$\mathcal{P}_i(\Pi, v, L) = \int_{\Pi} (P \cdot \text{curl } L + T \cdot \nabla^s v + (T^p - T) \cdot L) dV, \quad (43)$$

respectively. The tensor L is assumed to be nonsymmetric with the purpose of accounting for the power dissipated by the Burgers tensor, whose virtual change is measured by $\text{curl } L$. From the equation of virtual power he obtains the classical balance equations (5), the microforce balance equation

$$-(\text{curl } P^T)^T + T - T^p = 0, \quad (44)$$

and the relation $S = \mathbb{T} n$, with

$$\mathbb{T}_{ijk} = e_{kjh} P_{hi}. \quad (45)$$

Then

$$\text{div } \mathbb{T} = -(\text{curl } P^T)^T = T^p - T, \quad \mathbb{T} \cdot \nabla L = P \cdot \text{curl } L, \quad (46)$$

and the expressions (43) and (39) of the internal power coincide if $T^p - T = -\tilde{B}$, that is, if $\text{div } \mathbb{T} + \tilde{B} = 0$. But this is exactly what is required by the third balance equation $\text{div } \mathbb{T}^w = 0$ and by the pseudo-balance equation $\text{div } \mathbb{T}^s + \tilde{B} = 0$. Thus, (43) is the expression of the internal power for the subclass of continua obtained from the expression (32) of the virtual power, the invariance assumptions (33), (34), and the supplementary assumption (45).

Appendix A: Proof of the Cauchy theorem for second-order tensors

Here I give a quick proof of Cauchy's theorem for the balance equation

$$\int_{\Pi} B dV + \int_{\partial\Pi} S dA = 0, \quad (A.1)$$

where $B = B(x)$ and $S = S(\Pi, x)$ are second-order tensor fields. For every fixed vector e , the balance Equation (4)₁ holds with $b = Be$ and $s = Se$. Then by Cauchy's theorem there is a second-order tensor $T(e)$ such that

$$Se = T(e)n, \quad \text{div } T(e) + Be = 0. \quad (A.2)$$

These equations show that the dependence of T on e is linear. That is, there is a third-order tensor \mathbb{T} such that¹³

$$\mathbb{T}[e] = T(e) \quad \mathbb{T}_{ijk}e_j = T(e)_{ik}. \quad (A.3)$$

Then from Equation (A.2),

$$\begin{aligned} Se &= (\mathbb{T}[e])n & S_{ij}e_j &= \mathbb{T}_{ijk}e_j n_k, \\ \text{div } (\mathbb{T}[e]) + Be &= 0 & \mathbb{T}_{ijk,k}e_j + B_{ij}e_j &= 0, \end{aligned} \quad (A.4)$$

and, from the arbitrariness of e ,

$$\begin{aligned} S &= \mathbb{T}n & S_{ij} &= \mathbb{T}_{ijk}n_k, \\ \text{div } \mathbb{T} + B &= 0 & \mathbb{T}_{ijk,k} + B_{ij} &= 0. \end{aligned} \quad (A.5)$$

¹³The choice of the summed index in (A.3)₂ is arbitrary.

Acknowledgements

I thank P. Bisegna and M. Frémond for stimulating observations.

References

- [Capriz 1989] G. Capriz, *Continua with microstructure*, vol. 35, Springer tracts in natural philosophy, Springer-Verlag, New York, 1989.
- [Germain 1972] P. Germain, “Sur l’application de la méthode des puissances virtuelles en mécanique des milieux continus”, *C. R. Acad. Sci. A-B* **274** (1972), A1051–A1055.
- [Germain 1973a] P. Germain, “La méthode des puissances virtuelles en mécanique des milieux continus, I: Théorie du second gradient”, *J. Mécanique* **12** (1973), 235–274.
- [Germain 1973b] P. Germain, “The method of virtual power in continuum mechanics, II: microstructure”, *SIAM J. Appl. Math.* **25**:3 (1973), 556–575.
- [Gurtin 2004] M. E. Gurtin, “A gradient theory of small-deformation isotropic plasticity that accounts for the Burgers vector and for dissipation due to plastic spin”, *J. Mech. Phys. Solids* **52**:11 (2004), 2545–2568.
- [Gurtin and Anand 2005] M. E. Gurtin and L. Anand, “A theory of strain-gradient plasticity for isotropic, plastically irrotational materials, I: small deformations”, *J. Mech. Phys. Solids* **53**:7 (2005), 1624–1649.
- [Gurtin and Martins 1976] M. E. Gurtin and L. C. Martins, “Cauchy’s theorem in classical physics”, *Arch. Ration. Mech. Anal.* **60**:4 (1976), 305–324.
- [Jammer 1957] M. Jammer, *Concepts of force; a study in the foundations of dynamics*, Harvard University Press, 1957. Reprinted by Dover, 1999.
- [Noll 1959] W. Noll, “The foundations of classical mechanics in the light of recent advances in continuum mechanics”, pp. 266–281 in *The axiomatic method. With special reference to geometry and physics. Proceedings of an International Symposium held at the Univ. of Calif., Berkeley, Dec. 26, 1957-Jan. 4, 1958*, edited by L. Henkin et al., Studies in Logic and the Foundations of Mathematics, North-Holland Publishing, Amsterdam, 1959.
- [Noll 1974] W. Noll, “La mécanique classique, basée sur un axiome d’objectivité”, pp. 47–56 in *The Foundations of Continuum Mechanics and Thermodynamics. Selected Papers of W. Noll*, Springer-Verlag, Berlin, 1974.
- [Truesdell and Noll 1965] C. Truesdell and W. Noll, “The non-linear field theories of mechanics”, pp. 1–602 in *Handbuch der Physik, Bd III/3*, Springer-Verlag, Berlin, 1965.

Received 20 Dec 2007. Accepted 7 Mar 2008.

GIANPIETRO DEL PIERO: dlpgpt@unife.it

Dipartimento di Ingegneria, Università di Ferrara, Via Saragat 1, 44100 Ferrara, Italy

NUMERICAL EXPLORATION OF THE DANG VAN HIGH CYCLE FATIGUE CRITERION: APPLICATION TO GRADIENT EFFECTS

FELIX HOFMANN, GRATIELA BERTOLINO, ANDREI CONSTANTINESCU AND MOHAMED FERJANI

The objective of this paper is to show that a number of key features of the Dang Van high cycle fatigue criterion can be observed using simple polycrystalline computational models.

This paper presents a series of numerical computations for an inclusion consisting of 156 grains embedded in a homogeneous matrix. The grains are modeled using a polycrystalline single slip elasto-plastic model, whilst the matrix is considered as elastic. As expected the numerical simulations confirm the theoretical prediction on which the Dang Van fatigue criterion is based, that if a large enough number of grains is considered under uniform loading, a grain with the least favourable lattice orientation will always be present. This grain will constitute the weakest link in the assembly and thus its fatigue life largely determines the fatigue life of the bulk material.

Next the question of stress-gradients in the high cycle fatigue regime is addressed. An example of stress gradients appears around notches as they create stress concentrations in structures. It is a well known problem that fatigue criteria have to be locally arranged using stress-factors or critical distances in order to give satisfactory predictions. The work presented here shows that an analysis of the problem at the grain scale explains the apparent discrepancy when using classical fatigue criteria. The discussion is based on a numerical model of single slip crystal plasticity and the Dang Van fatigue criterion.

1. Introduction

Initially fatigue criteria were purely phenomenological, relying directly on the interpretation of experimental results at the macroscopic scale. Starting with the pioneering paper of [Orowan \[1939\]](#) on grain plasticity, the possibility of including grains scale effects within fatigue criteria was recognised. One of the fatigue models including grain level phenomena in a macroscopic fatigue criterion is the Dang Van–Papadopoulos criterion [[Dang Van 1993](#); [Papadopoulos 1994](#); [1995](#); [Dang Van and Papadopoulos 1999](#)]. It states that fatigue does not occur if all grains reach an elastic shakedown state. In order to estimate the stress-strain state at the meso scale a simple homogenisation scheme of a plastic inclusion in an elastic matrix is considered. The keypoint of the homogenisation scheme is the assumption that any macroscopic material point includes all possible grain lattice orientations. On this basis a simple set of macroscopic formulae provide an estimate of the fatigue limit.

Since the initial Orowan grain models numerous refinements have been proposed and today complex polycrystalline grain models are available [[Asaro 1983](#); [Kothari and Anand 1998](#)], as well as the computational power required for calculations comparable to experimental observations. Recent studies

Keywords: high cycle fatigue, fatigue criterion, polycrystalline plasticity, stress gradient, notch.

Part of the work presented in this paper was carried out by Felix Hofmann during a summer internship at the LMS funded by the CNRS.

of face centered cubic (FCC) crystal plasticity and fatigue predictions are, for example, presented in [Saanouni and Abdul-Latif 1996; Bennet and McDowell 2003; Manonkul and Dunne 2004]. Based on these models low cycle fatigue crack nucleation has been studied in [Dunne et al. 2007]. One can remark that the polycrystalline models discussed in the preceding references are rather complex. However this complexity will fade by the application of a phenomenological fatigue criterion.

This work revisits the Dang Van criterion (DVC) using a simple numerical polycrystalline model, with the specific focus of interpreting the fatigue limit as a shakedown limit for each grain. As such the proposed modeling will refine the initial closed-form homogenisation scheme.

Initially the proposed model will be used to illustrate the correct functioning of the criterion when considering a group of grains in a representative material. When the applied stress is homogeneous at the level of the representative volume element, the results of the classical closed-form homogenisation scheme are obtained. However, when a stress gradient is applied, the homogenisation assumptions are no longer valid and the numerical results illustrate the stress distribution in the grains. If the fatigue criterion is interpreted as a shakedown limit for each grain, one obtains a natural explanation of the “gradient effect” observed classically in fatigue experiments [Taylor 1999; Adib and Pluvinage 2003; Naik et al. 2005].

In our particular case, we will focus, for convenience and simplicity, on a two dimensional plane strain model of an austenitic steel with FCC crystal structure. However since we will show that the underlying homogenisation assumption breaks down in the case of a steep stress gradient, the specific case considered here (a two dimensional model and FCC structure) does not limit the generality of the observations. In fact the results and conclusions can be readily extended to three dimensions, as well as other crystallographic structures.

First a short overview of the basic assumptions of the Dang Van fatigue theory based on the shakedown concept will be given. The next section outlines the models and the computations of the simulated experiments. Finally the results are presented, firstly in the case of uniform loading, illustrating the DVC and secondly demonstrating the effect of applying a stress gradient.

2. Main assumptions of the Dang Van fatigue criterion

The fatigue analysis presented next is based on the DVC as presented in [Dang Van 1993; Papadopoulos 1995; Dang Van and Papadopoulos 1999].

Let us consider a structure under cyclic mechanical loading. Its fatigue lifetime will be determined by a number of mechanical fields: elastic and plastic strains, stresses, etc. computed over each cycle. The underlying hypothesis is that, after a short initial period of a few cycles, the mechanical response of the structure is stabilized, meaning that the fields will evolve in closed loops.

Fatigue phenomena can then be characterized at three scales:

- (i) the microscopic scale of dislocations, which are the underlying elements of plastic deformation, persistent slip bands, and elastoplastic strains;
- (ii) the mesoscopic scale of grains, where fatigue and damage phenomena are concentrated either at the grain boundary or in the interior;
- (iii) the macroscopic scale of the structure, at which loads are applied and industrial design is performed.

An inspection of the three scales during a cyclic loading would lead to different observations determined by the fatigue regime (see [Figure 1](#)):

- In the low cycle fatigue regime, physical observations at both macroscopic and mesoscopic scale show extensive plastic strains. Moreover homogenisation theory shows that strains and stresses at the two scales tend to be closer to each other with increasing plastic strain. This can be translated into saying that the higher the applied load, the more similar mesoscopic and macroscopic scales will behave.
- In the high cycle fatigue regime, two fatigue domains corresponding to finite and infinite lifetime can be considered. Physical observations at the macroscopic scale show that structures are macroscopically in an elastic shakedown state. At the mesoscopic scale of the grains, it is now commonly accepted that elastic shakedown occurs only in the case of infinite lifetime. If lifetime is finite, some grains will be oriented such that they can not reach an elastic shakedown state, but will experience a plastic shakedown or ratcheting state leading to failure after a finite number of cycles. The stress concentration due to this mesoscopic failure marks the initiation of a macroscopic crack associated with failure on the macroscopic scale.

Focusing on the case of high cycle fatigue, one can imagine a case where only one misoriented grain is subject to plastic slip. Then a simple homogenisation scheme of a plastic inclusion in an elastic matrix can be used to derive closed-form relations between mesoscopic and macroscopic fields.

Examples of possible homogenisation assumptions are [\[Cano et al. 2004\]](#):

- Lin–Taylor supposes strain equality: $\boldsymbol{\epsilon} = \boldsymbol{E}$. This is the hypothesis of the initial Dang Van or Papadopoulos fatigue criterion.
- Sachs supposes stress equality: $\boldsymbol{\sigma} = \boldsymbol{\Sigma}$
- Kröner assumes $\boldsymbol{\sigma} = \boldsymbol{\Sigma} - \mathbb{C} : (\mathbb{I} - \mathbb{P} : \mathbb{C}) : \boldsymbol{\epsilon}^p$, where \mathbb{C} and \mathbb{P} are respectively the fourth rank elastic moduli and Hill tensors and $\boldsymbol{\sigma}$ and $\boldsymbol{\Sigma}$ are respectively the mesoscopic and macroscopic stresses. In

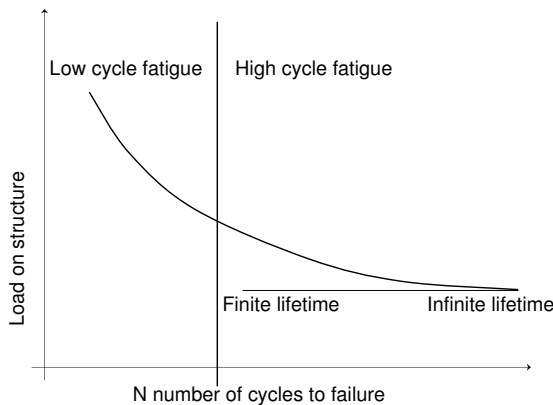


Figure 1. Illustration of high and low cycle fatigue regimes on a Wohler diagram.

the particular case of an idealized spherical inclusion, \mathbb{P} reads

$$\mathbb{P} = \frac{A}{3K} \mathbb{J} + \frac{B}{2\mu} \mathbb{K}, \quad \text{with} \quad A = \frac{3K}{3K + 4\mu}, \quad \text{and} \quad B = \frac{6}{5} \frac{K + 2\mu}{3K + 4\mu},$$

where $\mathbb{J} = \frac{1}{3} \mathbf{I} \otimes \mathbf{I}$ and $\mathbb{K} = \mathbb{I} - \mathbb{J}$ with \mathbb{I} the fourth rank identity tensor.

If, in all the cases, the same elastic behaviour at the mesoscopic and the macroscopic scale is assumed, the relation between mesoscopic and macroscopic fields can be written in the general form

$$\boldsymbol{\sigma} = \boldsymbol{\Sigma} - \mathbb{C}^* : \boldsymbol{\varepsilon}^p = \boldsymbol{\Sigma} + \boldsymbol{\rho}^*,$$

where $\boldsymbol{\rho}^*$ should be interpreted as a mesoscopic residual stress field.

The particular case of each model is obtained depending on the form of \mathbb{C}^* :

- for Lin–Taylor’s model, $\mathbb{C}^* = \mathbb{C}$;
- for Sachs model, $\mathbb{C}^* = 0$;
- for Kröner’s scheme, $\mathbb{C}^* = \mathbb{C} : (\mathbb{I} - \mathbb{P} : \mathbb{C})$.

Assuming only one active slip system generates the plastic strain

$$\boldsymbol{\varepsilon}^p = \frac{1}{2} \sum_s \gamma^s (\mathbf{m}^s \otimes \mathbf{n}^s + \mathbf{n}^s \otimes \mathbf{m}^s) = \sum_s \gamma^s \boldsymbol{\alpha}^s, \quad \text{with} \quad \boldsymbol{\alpha}^s = \frac{1}{2} (\mathbf{m}^s \otimes \mathbf{n}^s + \mathbf{n}^s \otimes \mathbf{m}^s).$$

The mesoscopic shear and normal stress for a slip system s with slip plane normal \mathbf{n}^s and slip direction \mathbf{m}^s can then be expressed as $\sigma_n^s = (\boldsymbol{\sigma} : \mathbf{n}^s \otimes \mathbf{n}^s)$ and $\tau^s = \boldsymbol{\sigma} : \mathbf{m}^s \otimes \mathbf{n}^s$.

Using the previous definitions we can define a series of fatigue criteria.

For an *individual grain*, when considering all slip systems, we have infinite lifetime if and only if

$$\max_s \max_t (\tau^s(t) + a\sigma_n^s(t)) < b,$$

where a and b are material constants and s is an index of the slip plane.

Looking at the macroscopic assembly of a number of grains g , this infinite lifetime criterion can be extended to

$$\max_g \max_s \max_t (\tau^s(t) + a\sigma_n^s(t)) < b,$$

where \max_g refers to the most critical grain.

Under the assumption that the grain orientations statistically cover all directions [Papadopoulos 1994; 1995], this can be simplified to

$$\max_t (\tau(t) + a\sigma^H(t)) < b,$$

which is the classical DVC formulation. Here τ is the Tresca norm of the mesoscopic shear and $\sigma^H = 1/3 \text{tr} \boldsymbol{\sigma}$ is the hydrostatic stress.

We can now recall that it is currently accepted that the Papadopoulos formulation of the criterion provides practically equivalent predictions to the initial Dang Van formulation and can thus be expressed as $k^* + a\sigma_{\max}^H < b$, where k^* denotes the smallest radius of a hypersphere encompassing the stress path and σ_{\max}^H is the maximal hydrostatic stress.

It is common to represent the stress path in a mesoscopic shear $\tau(t)$ versus mesoscopic hydrostatic stress $\sigma^H(t)$ diagram as schematically drawn in Figure 2. The line defined by the criterion is the frontier

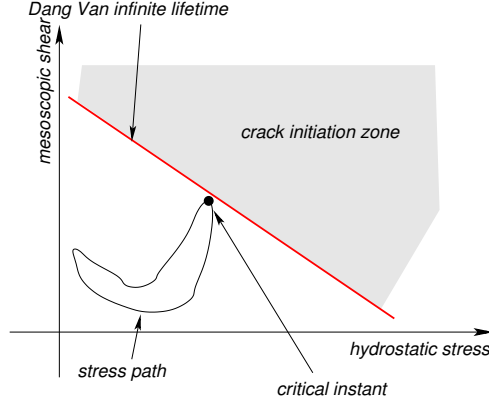


Figure 2. Illustration of the Dang Van criterion (DVC) in the τ, P plane.

between infinite life and fatigue. Component life for a load path contained completely below the DVC line will be infinite. If any point of the load path is located above the DVC line, fatigue will occur.

The parameters a and b are generally obtained from the torsion and bending fatigue limits, t_∞ and f_∞ respectively, as

$$a = \frac{t_\infty - \frac{1}{2}f_\infty}{\frac{1}{3}f_\infty}, \quad b = t_\infty.$$

3. The experiments and their modeling

The discussion will be based on the simulation of three different setups (see [Figure 3](#)):

- (i) a repeated tensile experiment (in plain strain) on a *box* specimen

$$\boldsymbol{\sigma} = \sigma_{\max} \frac{1 + \sin(t)}{2} (\mathbf{e}_y \otimes \mathbf{e}_y + \nu \mathbf{e}_z \otimes \mathbf{e}_z), \quad t \in \mathbb{R};$$

- (ii) an alternated shear experiment (in plain strain) on a *box* specimen

$$\boldsymbol{\sigma} = \tau_{\max} \sin(t) (\mathbf{e}_x \otimes \mathbf{e}_y + \mathbf{e}_y \otimes \mathbf{e}_x), \quad t \in \mathbb{R};$$

- (iii) a repeated tensile experiment (in plain strain) on a *notched* specimen

$$\boldsymbol{\sigma} = \sigma_{\max} \frac{1 + \sin(t)}{2} (\mathbf{e}_y \otimes \mathbf{e}_y + \nu \mathbf{e}_z \otimes \mathbf{e}_z), \quad t \in \mathbb{R}.$$

Shear alternated loading was chosen instead of repeated loading, as repeated loading required an excessively large number of cycles to reach an elastic shakedown. This can be easily explained considering that onset of first mesoscopic plasticity is reached at a macroscopic shear of $\tau \approx 0.75\tau_Y$ where τ_Y is the yield limit in shear [[Dang Van and Papadopoulos 1999](#)]. Therefore the maximal macroscopic shear in repeated loading is $\tau_{\max} = 2\tau \approx 1.5\tau_Y \gg \tau_Y$.

All experiments were modeled using both a homogeneous specimen with a macroscopic homogenised constitutive law and a specimen with a polygrain inclusion embedded within a homogenised elastic

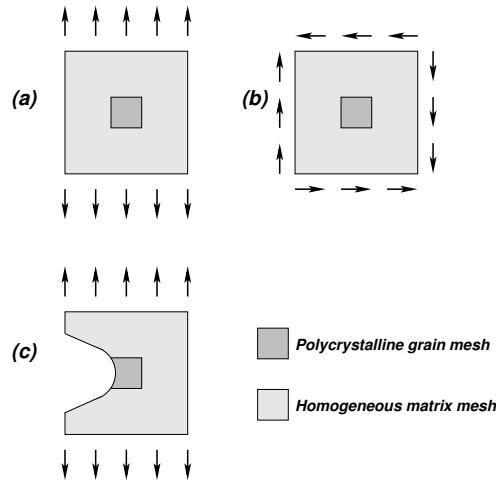


Figure 3. Illustration of the three different loading systems and the corresponding meshes which have been analysed.

material matrix (see Figure 3). The grain inclusion and the constitutive models for the grains will be described in the next section.

For simplicity the models were limited to two dimensions with plane strain condition. Loading was applied in displacement control at the mesh boundaries.

Care was taken to always remain macroscopically in the elastic shakedown state. However at the mesoscopic scale, some grains experience yielding and reach either a plastic or an elastic shakedown state after a few cycles.

Once the stabilized stress and strain fields had been obtained, the stress results of the FEM computations were post processed in the following ways:

- (DV1) Computation of the DVC for each slip system in each grain (slip system projection): The DVC for each slip system is computed using the precise knowledge of the grain orientation and the slip systems of the grain. Thus one can precisely compute the mesoscopic shear and hydrostatic stress on each slip system, and compute the inequality in [Dang Van 1993] in each case in order to determine the most critical grain and slip direction.
- (DV2) Computation of the Dang Van fatigue criterion in each grain: The computation of the criterion in each grain is done using the classical algorithms of the DVC with the mean stress field computed over each grain as an input value.
- (DV3) Computation of the Dang Van fatigue criterion for the homogenous structure: The computation of the criterion is performed using the classical algorithms of the criterion with the stress field computed from a homogenous elastic structure submitted to the same load. In this case only the hot spot of the structure, the most critical point, is plotted in the mesoscopic shear-hydrostatic stress diagram.

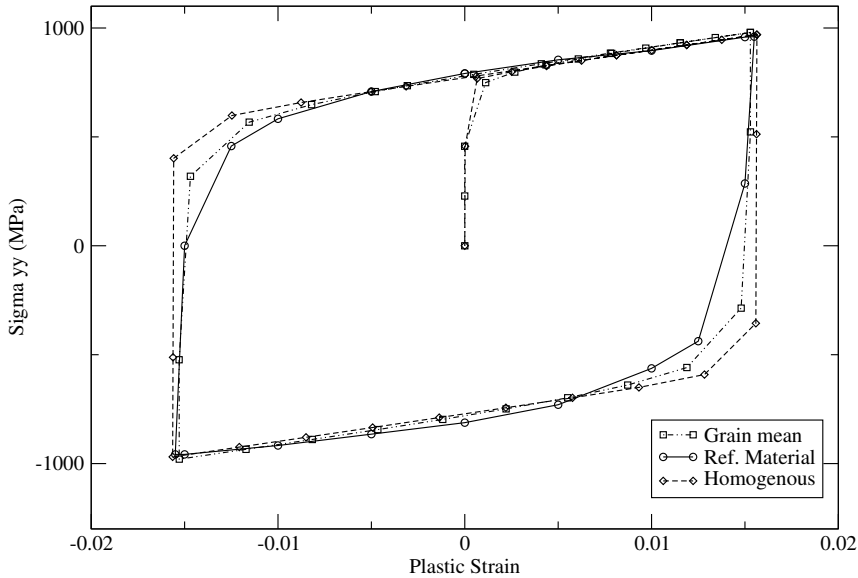


Figure 4. Macroscopic and mesoscopic material behaviour compared with experimental data averaged over a number of grains in a cyclic axial loading simulation.

4. The material

The material considered in this study corresponds to an austenitic steel with FCC crystallographic structure. The material behaviour during a cyclic tensile test is displayed in [Figure 4](#) and has been modeled at the macroscopic scale using an elastoplastic constitutive law with linear kinematic hardening.

At the mesoscopic scale the material was modeled based on a map of grains and lattice orientations obtained from experimental observation. The grain boundaries were mapped by image processing of electronic microscopy images. Lattice orientations were found by orientation imaging microscopy using electron back scattering diffraction and were assigned in the form of Euler angles for each grain.

In this study a map of 156 grains was used (see [Figure 5](#)). Each grain was individually meshed in two dimensions using linear 3 noded elements.

The material behaviour at the mesoscopic or grain scale is captured by a simplified phenomenological elastoplastic constitutive law with linear kinematic hardening. It assumes that plastic deformation is primarily caused by crystallographic slip, which applies to most cubic crystals and some hexagonal-close-packed crystals as discussed by [Weng \[1983\]](#) or by [Kowalczyk and Gambin \[2004\]](#).

We assumed that the applied stress resolved along the slip direction on the slip plane (to give a shear stress) initiates and controls the extent of plastic deformation. Yield begins on a given slip system when the shear stress on this system reaches a critical value, the critical resolved shear stress, independent of the tensile stress or any other normal stress on the lattice plane [[Bertolino et al. 2007](#)]. For FCC lattice structure this assumption is acceptable, however in less symmetric lattices, there may be some dependence on the hydrostatic stress.

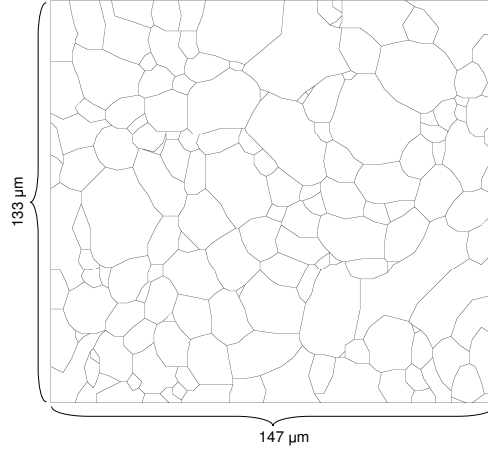


Figure 5. Grain contours of the microstructure used in this study.

For each grain g , the local yield criterion $f_g(\sigma_g)$ is obtained by the Schmid law. The individual yield stress σ_g^c depends on the active slip (gliding) system s

$$\sigma_g^c = \min_s \frac{\tau_0}{F_g^s},$$

where τ_0 denotes the critical resolved shear stress (a material parameter) and F_g^s the Schmid factor computed for each slip system using the lattice orientation provided for each grain. All twelve slip systems s of the FCC crystal structure were considered for the computation of σ_g^c and are given in [Table 1](#). For clarity [Figure 6](#) shows one of the FCC slip planes with its associated slip directions.

The macroscopic yield stress σ_y , the Young's modulus E , and the macroscopic kinematic hardening modulus H were identified from a macroscopic tensile test. The identification of the mesoscopic material parameters was carried out using a square grain inclusion (see [Figure 5](#)) in a square elastoplastic matrix (see [Figure 3](#)). The parameters were then adjusted to match the numerical homogenisation, that is, the averaged response over all the grains, with the macroscopic behaviour. The complete set of macroscopic material parameters is presented in [Table 2](#).

It is important to note that, although the identification of the material parameters was carried out in the plastic regime, it was ensured that loading during the actual simulations presented next was such that the matrix always remained in the elastic regime.

| Slip plane | Slip direction |
|------------|---------------------------------|
| (1 1 1) | [1 -1 0], [-1 0 1], [0 1 -1] |
| (1 -1 1) | [0 -1 -1], [1 0 -1], [1 1 0] |
| (-1 -1 1) | [0 -1 -1], [1 0 1], [-1 -1 0] |
| (-1 1 1) | [0 1 -1], [1 0 1], [-1 -1 0] |

Table 1. FCC slip planes and the associated slip systems.

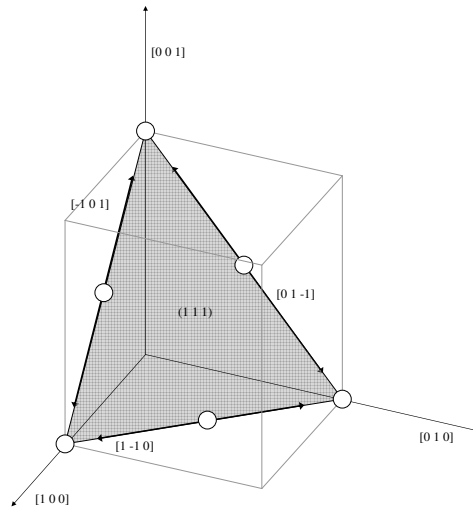


Figure 6. FCC crystal structure with one slip plane and its associated slip directions.

All finite element stress computations were performed using the object oriented finite element toolbox [Cast3M 2008] which includes a number of preprocessing, solving, and postprocessing routines. Parts of the postprocessing procedures were performed using MATLAB.

5. Results and discussion

The presentation of the numerical results is carried out by plotting the critical instant found for each grain on a Dang Van plot (see Figure 7). The critical grain instances are computed by post processing scheme 1 from the stabilised stress trajectory in each grain once a shakedown state has been reached (see Figure 2).

First, to validate the proposed approach, it was essential to verify that for tensile and shear loading, performed on the box specimen, the projection of mesoscopic shear onto crystal slip systems (post processing method DV1) leads to similar results as the DVC (post processing DV3).

To ensure that the number of grains in the grain mesh constitutes a representative sample, two further grain inclusions with randomly generated lattice orientations based on the same grain structure were generated. Figures 8 and 9 represent plots of mesoscopic shear versus hydrostatic stress of the grain critical instances for these three different distributions of lattice orientations (O_1, O_2, O_3), when subjected to macroscopic shear and tensile loading respectively.

| E (GPa) | ν | σ_Y (MPa) | H (MPa) | a | b (MPa) |
|-----------|-------|------------------|-----------|------|-----------|
| 210. | 0.3 | 670. | 9500. | 0.45 | 260. |

Table 2. The macroscopic material parameters: E is Young’s modulus, ν Poisson ratio, σ_Y the yield limit, H the kinematic hardening modulus, and a and b Dang Van fatigue parameters.

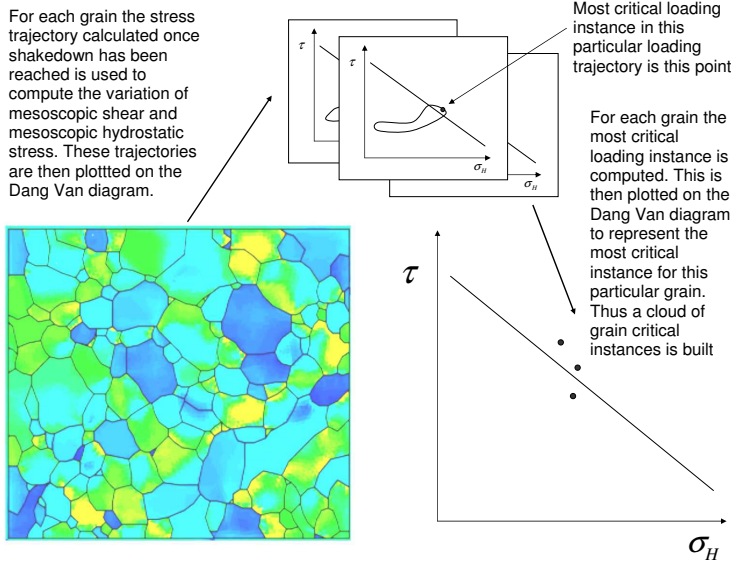


Figure 7. Plotting scheme of grain critical instances.

5.1. No stress gradient: box specimen. In each simulation, the structure was subjected to 5 cycles and the results from the last cycle were plotted. In both cases the homogeneous elastic solution has also been represented (corresponding to postprocessing DV3). One can easily remark that the clouds of grain

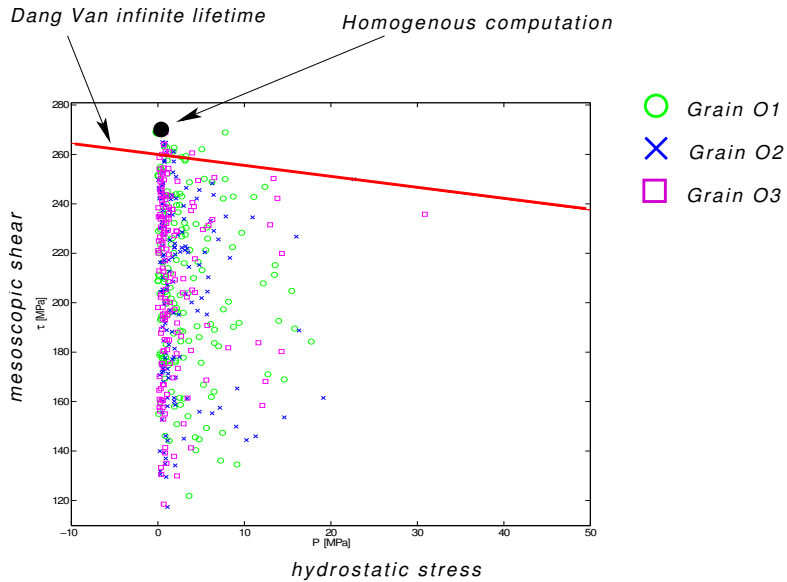


Figure 8. Three clouds of grain critical instances plotted in a Dang Van diagram of mesoscopic shear versus mesoscopic hydrostatic stress for different lattice orientations (O_1 , O_2 , O_3) in the case of alternated shear loading.

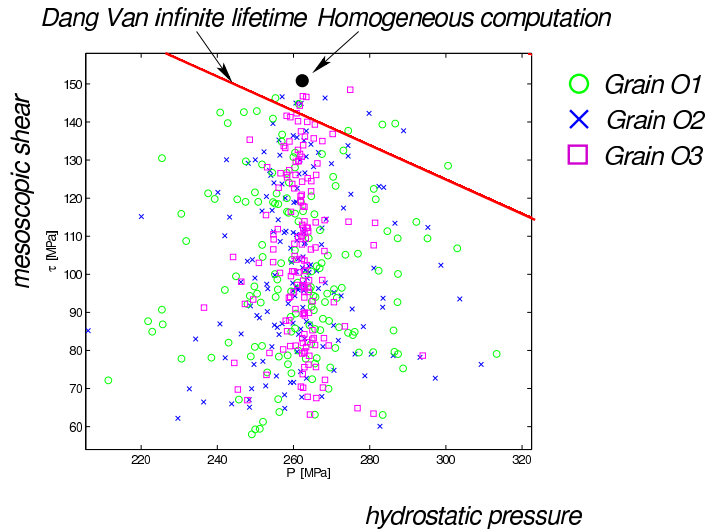


Figure 9. Three clouds of grain critical instances plotted in a Dang Van diagram for different lattice orientations (O_1 , O_2 , O_3) in the case of repeated tensile loading.

critical instances (see Figures 8 and 9) are just below the homogeneous critical instant and that the most critical grain approximately coincides with the homogeneous solution.

If the applied load is below the fatigue limit it can be seen that the clouds of critical instances are compact with a small range of mesoscopic hydrostatic stresses, corresponding to small pressures (see Figure 10, $\sigma_{\max} = S_2$ (shear loading) and Figure 11, $\tau_{\max} = T_2$ (tensile loading)). When loading is increased the clouds of grain critical instances closely follow the homogenous elastic solution point. Also the clouds spread out and the range of mesoscopic hydrostatic stress increases significantly (see Figure 10, $\sigma_{\max} = S_1$ (shear loading) and Figure 11, $\sigma_{\max} = T_1$ (tensile loading)). This is a direct consequence of increased grain plasticity with increasing applied load, as the hydrostatic stress range is directly related to the residual stress distribution in the individual grains. However when the residual stress average is computed across all the grains it is close to zero as one would expect. This confirms that macroscopically the homogeneous elastic solution can still be used as a reference.

The case of shear loading (see Figure 10) presents an interesting distribution of grain critical instances in the case where practically no plastic deformation has been observed. The zero mesoscopic hydrostatic stress observed in this case is coherent with and as expected for shear loading. As load increases however, the range of mesoscopic hydrostatic stresses increases and the grain critical instances spread out, as a result of the increase in grain plasticity.

5.2. Results and discussion: with stress gradient. Next we shall examine the behaviour of the grain inclusion when exposed to a stress gradient. This gradient was introduced by means of a notch with a stress concentration factor of approximately 5 (see Figure 3, configuration c)

One would expect the strain and stress distribution in the homogenous specimen to differ significantly from that in the specimen containing the granular inclusion. However, they show considerable similarity

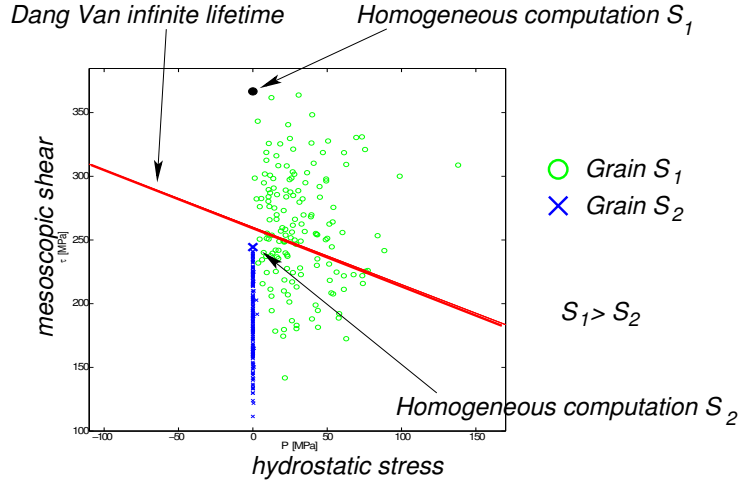


Figure 10. Two clouds of grain critical instances plotted in a Dang Van diagram for two different applied shear loads, $\tau_{\max} = S_1, S_2$.

since we primarily remain in the elastic regime which is the same for both models and only a few grains show small plastic deformations.

Figure 12 displays in a Dang Van diagram the grain critical instant clouds found in the case of repeated tensile loading of the notched specimen. For the specimen containing the granular inclusion two different clouds were computed:

- one by projecting the mesoscopic stresses onto the slip systems in each grain, and thus computing greatest mesoscopic shear and hydrostatic stress (post processing DV1), and
- one by direct application of the DVC to each grain (post processing DV2).

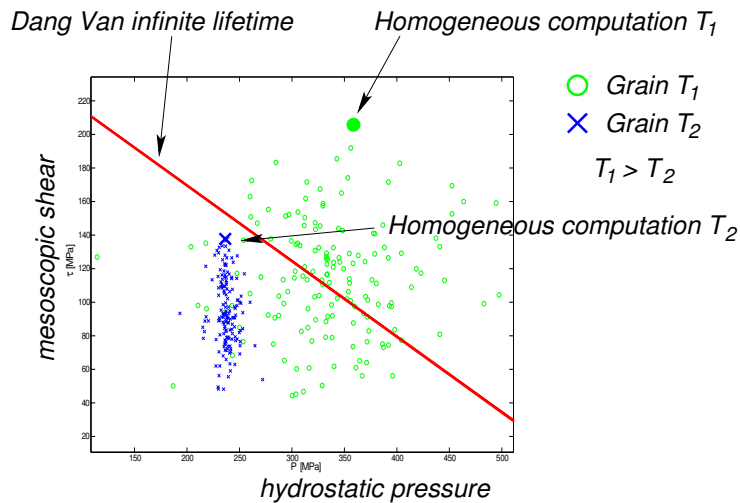


Figure 11. Two clouds of grain critical instances plotted in a Dang Van diagram for two different applied tensile loads, $\sigma_{\max} = T_1, T_2$.

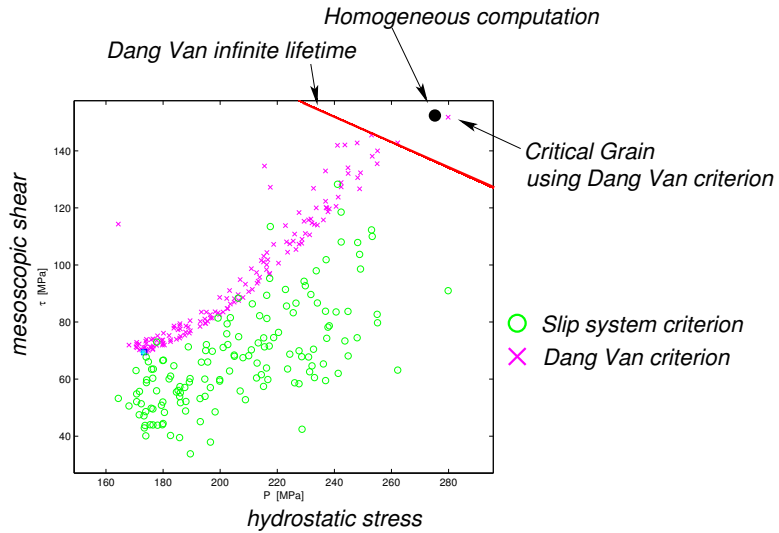


Figure 12. Two clouds of grain critical instances plotted in a Dang Van diagram for the case of repeated tensile loading for a notched specimen. The first cloud is computed by projecting the mesoscopic stresses onto the most critical slip system (post processing DV1), whilst the second cloud shows the application of the DVC to each grain (post processing DV2).

The most critical point in the cloud obtained using the grain DVC (post processing DV2) is close to the point obtained from the homogeneous solution (post processing DV3). This means that the most critical grain will experience the same damage as the notch tip, as they both experience a similar stress path.

Let us recall however that the DVC searches the mesoscopic shear on the most critical plane. In reality the likelihood that any slip system of the grain at the notch tip is aligned with the most critical plane predicted by Dang Van is small. Therefore the grain actually experiences a smaller mesoscopic shear amplitude on its active slip systems. Consequently we find that the second cloud obtained from the most critical slip system in each grain (post processing DV1) is far below the homogeneous computation point.

It is important to understand the role of the stress gradient in this setting. Because of the high stress gradient only a few grains located at the notch tip will experience the high stresses, reducing the chance of alignment of a slip system with the most critical plane. The bulk of the grains will experience a much lower stress approaching the far field stress applied to the boundaries of the notched specimen which is much lower than the stress at the notch tip.

In contrast, if the stress gradients are small, as in the box specimen, all grains experience very similar stresses and the likelihood that a slip system of a grain is aligned with the most critical plane predicted by Dang Van is high.

Figure 13 shows the distribution of grain critical instances for a notched specimen loaded in tension for a number of different randomly allocated sets of lattice orientations (O_1 , O_2 , O_3). As the clouds of grain critical instances occupy the same region of the Dang Van plot, the observed behaviour of the notched specimen can be seen as representative of the general case for any set of randomly distributed lattice orientations. A more quantitatively orientated investigation of the statistics of grain critical instant

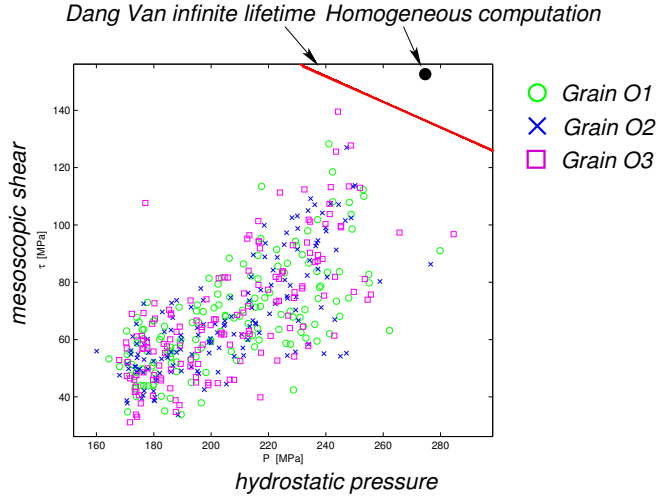


Figure 13. Three clouds of grain critical instances plotted in a Dang Van diagram for different lattice orientations (O_1 , O_2 , O_3) in the case of repeated tensile loading of a notched specimen.

distribution and its distance to the point of the homogeneous solution would allow the definition of an equivalent notch factor or a critical distance as usually employed in classical fatigue analysis.

When the applied load is increased (see Figure 14), the clouds of critical instances found from slip system projection in each grain (post processing DV1) follow the homogeneous solution as expected. However, we remark that the density of the cloud near the homogeneous solution is smaller than observed in the case of the box specimen. This is a direct consequence of the high stress gradient which places emphasis on a small number of grains independently of the bulk load.

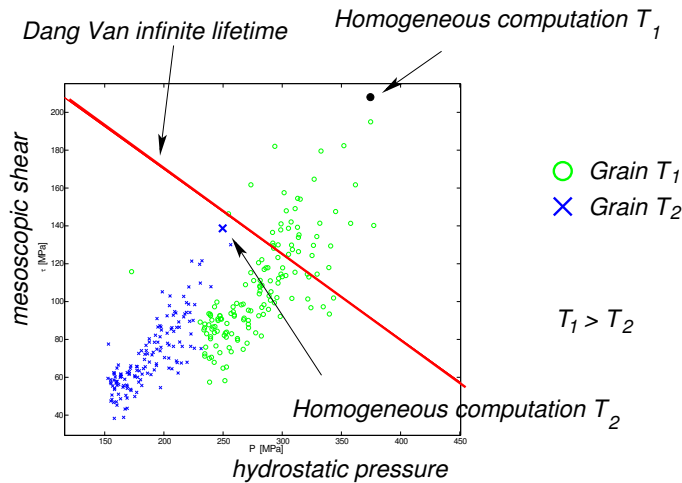


Figure 14. Two clouds of grain critical instances plotted in a Dang Van diagram for two different applied tensile loads (T_1 , T_2) in the case of a notched specimen.

This observed behaviour is directly due to the breakdown of the underlying homogenisation assumption of the Dang Van fatigue criterion. In the presence of a steep stress gradient with a characteristic length scale of the same order as the local microstructural length scale, the assumption that at each material point a uniform distribution of all lattice orientations is present fails. The local number of grains at each material point is too small to constitute a macroscopic representative volume element. This breakdown of the homogenisation assumption is phenomenologically captured in the notch factor or critical distance concept in fatigue analysis.

Although the discussion here was based on a two dimensional model of an austenitic steel with FCC crystal structure, this does not limit the generality of the presented results and conclusions. The breakdown of homogenisation assumption when macroscopic loading and microstructural length scales approach can be readily extended to three dimensions and other crystal lattice structures.

6. Conclusion

In this paper a number of numerical fatigue experiments using a simplified polycrystalline model at the mesoscopic/grain scale have been presented. The impact of the applied stress gradient on fatigue life prediction was shown considering the example of a notched specimen. It was shown that high localization of stresses causes failure of the homogenisation assumptions as only few grains are exposed to very high stresses. This failure implies a fundamental change in the application of fatigue criteria in general and more particularly of critical plane criteria such as Dang Van criterion in cases of high stress gradient. The results also provide a physical explanation for critical distances and notch factors encountered in engineering practice and underline the fact that whenever the length scales of macroscopic loading and local microstructure approach, difficulties with macroscopic homogenisation assumptions will be encountered.

A finer quantitative analysis both at the meso and macroscopic levels should allow the development of a better understanding and ability to predict fatigue for stress states involving high gradients such as those found in notched or cracked specimens and bending experiments.

References

- [Adib and Pluvinaige 2003] H. Adib and G. Pluvinaige, “Theoretical and numerical aspects of the volumetric approach for fatigue life prediction in notched components”, *Int. J. Fatigue* **25**:1 (2003), 67–76.
- [Asaro 1983] R. J. Asaro, “Micromechanics of crystals and polycrystals”, *Adv. Appl. Mech.* **23** (1983), 1–115.
- [Bennet and McDowell 2003] V. P. Bennet and D. L. McDowell, “Polycrystal orientation distribution effects on microslip in high cycle fatigue”, *Int. J. Fatigue* **25**:1 (2003), 27–39.
- [Bertolino et al. 2007] G. Bertolino, J. Crepin, and N. Bilger, “Modeling microstructures and microstructural effects on macroscopic and intragranular mechanical behavior”, *Comput. Mater. Sci.* **40**:3 (2007), 408–416.
- [Cano et al. 2004] F. Cano, A. Constantinescu, and H. Maitournam, “Critère de fatigue polycyclique pour des matériaux anisotropes: application aux monocristaux”, *C. R. Mécanique* **332**:2 (2004), 115–121.
- [Cast3M 2008] Cast3M, “Cast3M: an object oriented finite element toolbox”, 2008, Available at <http://www-cast3m.cea.fr/>.
- [Dang Van 1993] K. Dang Van, “Macro-micro approach in high-cycle multiaxial fatigue”, pp. 120–130 in *Advances in multiaxial fatigue* (San Diego, 1991), edited by D. L. McDowell and R. Ellis, STP **191**, ASTM, Philadelphia, 1993.
- [Dang Van and Papadopoulos 1999] K. Dang Van and I. V. Papadopoulos (editors), *High-cycle metal fatigue: theory to applications*, CISM Courses and Lectures **392**, Springer, Vienna, 1999.

- [Dunne et al. 2007] F. P. E. Dunne, A. J. Wilkinson, and R. Allen, “Experimental and computational studies of low cycle fatigue crack nucleation in a polycrystal”, *Int. J. Plast.* **23**:2 (2007), 273–295.
- [Kothari and Anand 1998] M. Kothari and L. Anand, “Elasto-viscoplastic constitutive equations for polycrystalline metals: application to tantalum”, *J. Mech. Phys. Solids* **46**:1 (1998), 51–67.
- [Kowalczyk and Gambin 2004] K. Kowalczyk and W. Gambin, “Model of plastic anisotropy evolution with texture-dependent yield surface”, *Int. J. Plast.* **20**:1 (2004), 19–54.
- [Manonkul and Dunne 2004] A. Manonkul and F. P. E. Dunne, “High- and low-cycle fatigue crack initiation using polycrystal plasticity”, *Proc. R. Soc. Lond. A* **460**:2047 (2004), 1881–1903.
- [Naik et al. 2005] R. A. Naik, D. B. Lanning, T. Nicholas, and A. R. Kallmeyer, “A critical plane gradient approach for the prediction of notched HCF life”, *Int. J. Fatigue* **27**:5 (2005), 481–492.
- [Orowan 1939] E. Orowan, “Theory of the fatigue of metals”, *Proc. R. Soc. Lond. A* **171**:944 (1939), 79–106.
- [Papadopoulos 1994] I. V. Papadopoulos, “A new criterion of fatigue strength for out-of phase bending and torsion of hard metals”, *Int. J. Fatigue* **16**:6 (1994), 377–384.
- [Papadopoulos 1995] I. V. Papadopoulos, “A high cycle fatigue criterion applied in biaxial and triaxial out-of-phase stress conditions”, *Fatigue Fract. Eng. Mater. Struct.* **18**:1 (1995), 79–91.
- [Saanouni and Abdul-Latif 1996] K. Saanouni and A. Abdul-Latif, “Micromechanical modeling of low cycle fatigue under complex loadings, I: Theoretical formulation”, *Int. J. Plast.* **12**:9 (1996), 1111–1121.
- [Taylor 1999] D. Taylor, “Geometrical effects in fatigue: a unified theoretical model”, *Int. J. Fatigue* **21**:5 (1999), 413–420.
- [Weng 1983] G. J. Weng, “A micromechanical theory of grain-size dependence in metal plasticity”, *J. Mech. Phys. Solids* **31**:3 (1983), 193–203.

Received 14 Dec 2007. Revised 14 Jun 2008. Accepted 17 Jul 2008.

FELIX HOFMANN: felix.hofmann@gmx.de

Department of Engineering Science, Oxford University, Parks Road, Oxford, OX1 3PJ, United Kingdom

GRATIELA BERTOLINO: bertolin@lms.polytechnique.fr

Centro Atómico Bariloche, 8400 San Carlos de Bariloche, Río Negro, Argentina

ANDREI CONSTANTINESCU: andrei.constantinescu@lms.polytechnique.fr

Laboratoire de Mécanique des Solides - CNRS, École Polytechnique, 91128 Palaiseau, France

MOHAMED FERJANI: ferjanimoh@yahoo.fr

Institut Français du Pétrole, 1-4 rue Bois-Préau, 92852 Rueil-Malmaison, France

CYCLIC APPROXIMATION OF THE HEAT EQUATION IN FINITE STRAINS FOR THE HEAT BUILD-UP PROBLEM OF RUBBER

YOHAN LE CHENADEC, IDA RAOULT, CLAUDE STOLZ AND MAC-LAN NGUYEN-TAJAN

It is well-known that rubber exhibits hysteretic mechanical behavior and has a low thermal conductivity. The main consequences are the heat generation and heat build-up phenomena which occur in a rubber component when subjected to repeated deformations. Estimating the heat build-up temperature implies the solution of a coupled thermomechanical problem. Due to the difference between the mechanical and the thermal diffusion characteristic times, a cyclic uncoupled approach is often used to solve the heat build-up problem.

In the uncoupled approach, the heat sources are first determined with a mechanical analysis, and the heat equation is then solved on a fixed geometry. At finite strains, the geometry of the body varies with the deformation but the foregoing method does not account for such changes in geometry. The exact solution would require describing the body deformation while solving the thermal problem, but this does not take advantage of the difference between the characteristic times of the thermal diffusion and the mechanical behaviour, respectively, and the exact numerical resolution is therefore unnecessarily time-consuming.

The purpose of the current work is to take into account kinematics in the thermal problem when using a cyclic uncoupled approach. The heat problem is written in the reference configuration. That implies that the problem is defined on a fixed domain: the initial configuration of the body. The changes in geometry in the reference heat equation are thus described by mechanical time-dependent variables. The cyclic assumption allows mean variables to be defined, for example the mean temperature. A time-integration method and an approximation of the heat equation are developed, leading to a simplified formulation with mechanical time-independent terms. This simplified heat problem is based on the mean variables.

1. Introduction

Heat build-up is the rise in temperature caused by heat generation in a component subjected to repeated deformation. It is characterized by a slow increase of the mean temperature until the steady-state temperature field is reached; see [Figure 1](#). A high frequency oscillation in temperature is superimposed on the mean temperature, as seen on the second half of the figure. This oscillation is caused by thermoelastic effects: [Gough \[1805\]](#) first reported that the temperature of rubber increased when it was stretched. Measurements have shown that the temperature first decreases until reaching a minimum and then increases [[Joule 1859](#); [James and Guth 1943](#)]. This thermoelastic inversion phenomenon has been studied and is due to the competition between the thermal expansion and the entropic aspect of rubber elasticity [[Chadwick 1974](#); [Treloar 1975](#)]. This thermoelastic oscillation is not relevant in the

Keywords: thermomechanical couplings, heat build-up, heat equation, finite strains, rubber.
This research was supported by PSA Peugeot Citroën.

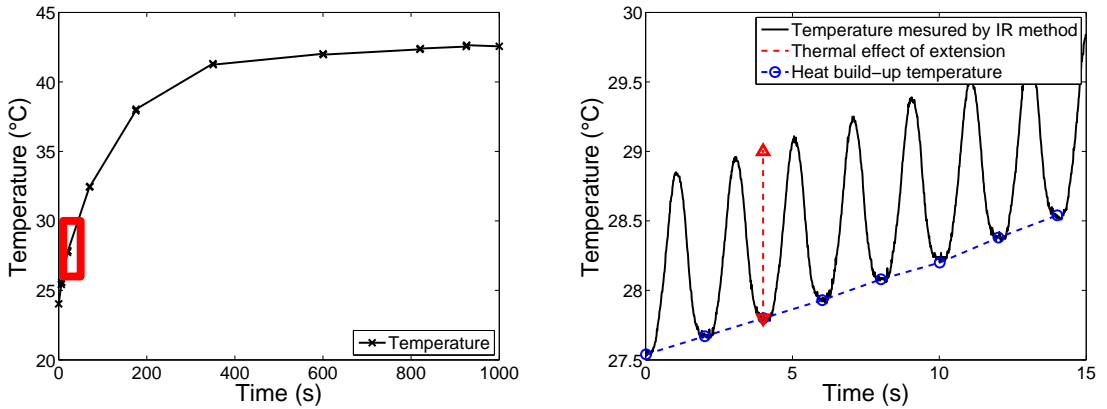


Figure 1. Temperature of a rubber component subjected to a cyclic deformation (0.5 Hz), measured with an infrared thermometer. The rectangle is enlarged on the right.

heat build-up problem; the heat build-up temperature is then defined as the actual temperature minus the thermoelastic effect.

Heat generation is the conversion of applied mechanical energy into heat due to the hysteretic behaviour of rubber. The causes and consequences of heat generation have been extensively explored in the literature [Medalia 1991; Meinecke 1991; Kar and Bhowmick 1997a; 1997b; Park et al. 2000]. It is generally attributed to molecular friction, but at large strains in natural rubber much of the hysteresis is thought to be associated with strain-induced crystallisation [Toki et al. 2000; Trabelsi et al. 2003]. The main consequence of hysteresis is to increase the temperature.

Estimating the heat build-up in a rubber component requires the solution of a coupled thermomechanical problem, defined both by mechanical and thermal equations. Coupled or uncoupled methods are available to solve the thermomechanical problem

- (1) Fully-coupled algorithm where both thermal and mechanical variables are calculated at the same time [Simo and Taylor 1991; Reese and Govindjee 1997].
- (2) Staggered (fractional-step) methods that split the coupled thermomechanical problem into two simpler problems [Armero and Simo 1992].
- (3) Uncoupling the thermal and mechanical problems under the assumption of cyclic loading [Whicker et al. 1981b; Sarkar et al. 1987; Ebbott et al. 1999]. Three analysis modules are needed: deformation, dissipation and thermal modules. First the mechanical problem is solved for one cycle. Then the dissipation is estimated and the thermal problem is solved for many cycles on one fixed geometry until the rise of temperature is significant, whereupon the mechanical problem is updated using the actual temperature.

The two first methods are exact and are indicated for solution of the thermoelastic problem over a limited number of cycles [Holzapfel and Simo 1996; Reese and Govindjee 1997]. The last method is often used in the literature for heat build-up problems since it offers a quicker solution.

The main difficulty of the thermomechanical problem lies in the coupling between the thermal and mechanical variables. Thermomechanical coupling can be classified into four categories:

- (1) The dependence of stress on temperature [Anthony et al. 1942].
- (2) Thermal effect of extension [James and Guth 1943; Treloar 1975].
- (3) The heat generation produced by the hysteretic behaviour of rubber.
- (4) The coupling between kinematics and the heat equation due to finite strains.

The dependence of stress on temperature is required for a very precise analysis of stress over a cycle but is not essential when the main issue is the determination of the heat build-up temperature. The contribution of the thermoelastic coupling to the dissipation over a stabilized cycle is actually zero [Lion 1997]. The consequence is that it is not necessary to represent thermoelastic effects to estimate the heat build-up temperature. The heat generation correspond to the heat sources: it is generally assumed that the loss of energy is nearly completely converted into heat. The last type of coupling arises from the solution of the thermal problem on a moving domain due to the cyclic deformation of the body. This coupling is neglected in uncoupled cyclic solutions since the heat equation is solved on one fixed geometry and not on the actual geometry. This coupling is negligible under the small strain condition, but it becomes essential to account for it at large strains. The purpose of our work is to account for this coupling within the cyclic uncoupled solving.

The paper is organized as follows: in Section 2 the cyclic uncoupled thermomechanical approach is presented. In Section 3 the heat equation is approximated over a cycle in order to take into account kinematics. Section 4 deals with numerical simulations of heat build-up of a specimen. Our conclusions are summarized in the final section.

2. Cyclic uncoupled thermomechanical approach

Let us assume a cyclic mechanical loading. Our analysis is based on the difference between mechanical and thermal diffusion characteristic times in heat build-up problems: the temperature increases slowly and many mechanical cycles are needed to reach the steady-state temperature. This difference between characteristic times cannot be emphasized by coupled thermomechanical algorithms since both thermal and mechanical variables are calculated at the same time. Only an uncoupled approach is able to take advantage of this difference in terms of time computation. The uncoupled cyclic algorithm is thus adopted [Whicker et al. 1981b; Sarkar et al. 1987].

This algorithm is the following (see Figure 2): the mechanical problem is first solved, then the loss of energy is estimated and put into the heat equation as heat sources. The thermal problem is solved on a fixed geometry and a loop is finally created. This loop allows us to update material data with the temperature. This algorithm uncouples the mechanical and the thermal solution in the sense that the heat equation is solved over a interval bigger than the mechanical characteristic time. The most interesting case is when the mechanical problem is solved for one cycle and the thermal problem is solved for many cycles. This feature quickens the solving of the heat build-up problem. The cyclic uncoupled thermomechanical algorithm is consistent with the definition of the heat build-up temperature given in the introduction that neglects the thermoelastic coupling.

The uncoupled approach is accurate for small or medium strains. The deformation of the body is small and the boundary value problem of heat equation can be solved on a fixed geometry and totally uncoupled from the mechanical analysis. But if the strain is large enough to modify the geometry of the component

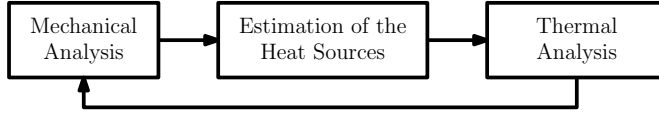


Figure 2. Simplified diagram of the thermomechanical uncoupled approach proposed in [Whicker et al. 1981a].

during the mechanical cycle, the solution of the thermal analysis will strongly depend on the choice of the fixed geometry, independently of the configuration. That implies that it would be necessary at large strains to follow the geometry deformation during the cycle to solve the thermomechanical problem. This coupled solution does not take advantage of the cyclic uncoupling algorithm in terms of time. In the next section a simplified heat equation, which accounts for the changes in geometry, is presented.

3. The simplified heat equation

In the case of periodic deformation, a method is developed to solve the heat equation with taking into account the cyclic geometry deformation. The scheme of the solving is the following: first write the heat equation in the reference configuration, then define mean variables such as temperature or thermal conductivity, and finally integrate with respect to time and approximate the heat equation and the boundary conditions.

3A. Kinematic description. A material point in actual (or Eulerian) configuration \vec{x} is linked to its position in the reference (or Lagrangian) configuration \vec{X} by the motion $\vec{x} = \Phi(\vec{X}, t)$. The deformation gradient \mathbf{F} and the right Cauchy-Green strain tensor \mathbf{C} are defined by

$$\mathbf{F} = \frac{\partial \Phi(\vec{X}, t)}{\partial \vec{X}} \quad \text{and} \quad \mathbf{C} = \mathbf{F}^T \mathbf{F}. \quad (3-1)$$

The transformation relations for the length, surface, and volume elements between the reference and the actual configuration are given by

$$d\vec{x} = \mathbf{F} d\vec{X}, \quad d\vec{a} = ds \vec{n} = J \mathbf{F}^{-T} d\vec{A}, \quad \text{and} \quad dv = J dV \quad \text{with} \quad d\vec{A} = dS \vec{N} \quad \text{and} \quad J = \det \mathbf{F}. \quad (3-2)$$

The density is obtained by $\rho_0(\vec{X}, t) = J \rho(\vec{x}, t)$.

3B. Thermal conductivity. The thermal conductivity of rubber is generally isotropic in the natural state, but it is sensible to think that it becomes anisotropic with the deformation. Previous models based on the network theory related the thermal conductivity to the deformation [van den Brule 1989; van den Brule and O'Brien 1990]. Measurements proved that the thermal conductivity of rubber actually changes with the deformation [Tautz 1959; Broerman et al. 1999]. The conductivity is higher in the stretched direction and lower in the compressed direction. The magnitude of this phenomenon depends on the material and varies from 10 to 100% at stretch $\lambda = 2$. In the following, the thermal conductivity tensor $\mathbf{k}(\mathbf{C}, T)$ is implicitly supposed to be dependent of the deformation.

3C. The heat equation in the reference configuration. The local equation of the energy balance in the actual reference and the boundary conditions are

$$\rho c \frac{\partial T}{\partial t} = \rho r - \operatorname{div} \vec{q} \quad \text{and} \quad \vec{q} \cdot d\vec{a} = h da (T - T_\infty) \quad (3-3)$$

where c is the heat capacity, $r(\vec{x}, t)$ represents the generation of heat supplied from outside, div is the divergence operator with respect to the coordinates \vec{x} in the actual configuration and $\vec{q}(\vec{x}, t)$ is the heat flux. The latter is related to the temperature by Fourier's law

$$\vec{q}(\vec{x}, t) = -\mathbf{k}(\mathbf{C}, T) \operatorname{grad} T \quad (3-4)$$

where $\mathbf{k}(\mathbf{C}, T)$ is the thermal conductivity tensor and grad is the gradient operator with respect to the actual reference. Note that the boundary conditions represent either the natural convection with the convection coefficient h and the far-field air temperature T_∞ , or the imposed temperature when $h \rightarrow \infty$. These equations are written in the reference configuration

$$\rho_0 c \frac{\partial T}{\partial t} = \rho_0 R - \operatorname{Div} \vec{Q}, \quad \vec{Q} \cdot d\vec{A} = \underbrace{\frac{da}{dA} h}_{H(\vec{X}, t)} dA (T - T_\infty), \quad (3-5)$$

where Div is the divergence operator with respect to the reference configuration, $R(\vec{X}, t) = r(\vec{x}, t)$ is the external heat production rate and \vec{Q} is the transported heat flux given by

$$\vec{Q}(\vec{X}, t) = J(\vec{X}, t) \mathbf{F}^{-1}(\vec{X}, t) \vec{q}(\vec{x}, t). \quad (3-6)$$

Fourier's law and the relationship $\operatorname{grad} T = \mathbf{F}^{-T} \vec{\nabla} T$ between the Eulerian and Lagrangian gradient lead to the Lagrangian expression of the heat flux

$$\vec{Q}(\vec{X}, t) = -J(\vec{X}, t) (\mathbf{F}^{-1} \mathbf{k}(\mathbf{C}, T) \mathbf{F}^{-T})(\vec{X}, t) \vec{\nabla} T(\vec{X}, t). \quad (3-7)$$

The reference thermal conductivity is given by

$$\mathbf{K}(\mathbf{C}, T) = J(\vec{X}, t) (\mathbf{F}^{-1} \mathbf{k}(\mathbf{C}, T) \mathbf{F}^{-T})(\vec{X}, t). \quad (3-8)$$

The convection coefficient H in the reference configuration depends on time and takes into account the variation of the surface along the deformation. Its expression is given by

$$H(\vec{X}, t) = h \frac{da}{dA} = h J \sqrt{\vec{N}^T \mathbf{C}^{-1} \vec{N}}. \quad (3-9)$$

3D. Approximation of the heat equation. Suppose that the deformation is periodic with period P . A mean temperature T_m function of time can be defined by

$$T_m(\vec{X}, t) = \frac{1}{P} \int_{-P/2}^{P/2} T(\vec{X}, t + \tau) d\tau \quad (3-10)$$

and the temperature oscillation is

$$T_a(\vec{X}, t, \tau) = T(\vec{X}, t + \tau) - T_m(\vec{X}, t) \quad (3-11)$$

By extension, the mean and oscillatory variables are described by subscripts m and a . For the sake of readability, the space variable \vec{X} will not be written anymore in the following. The mean conductivity tensor is given at a fixed temperature $T_m(t)$ by

$$\mathbf{K}_m(t) = \mathbf{K}_m(T_m(t)) = \frac{1}{P} \int_{-P/2}^{P/2} \mathbf{K}(\mathbf{C}(t + \tau), T_m(t)) d\tau. \quad (3-12)$$

The oscillatory conductivity tensor is thus

$$\mathbf{K}_a(t, \tau) = \mathbf{K}(\mathbf{C}, T)(t + \tau) - \mathbf{K}_m(t). \quad (3-13)$$

Given t , the heat equation becomes

$$\rho_0 c \frac{\partial T}{\partial \tau}(t + \tau) = \rho_0 R(t + \tau) + \text{Div}(\mathbf{K}_m(t) + \mathbf{K}_a(t, \tau))(\vec{\nabla} T_m(t) + \vec{\nabla} T_a(t, \tau)). \quad (3-14)$$

The heat equation for T_m is obtained by the integration of (3-14) over one period

$$\begin{aligned} \rho_0 c \frac{\partial T_m}{\partial t} = \rho_0 R_m + \text{Div}(\mathbf{K}_m(t) \vec{\nabla} T_m(t)) \\ + \underbrace{\text{Div}\left(\frac{1}{P} \int_{-P/2}^{P/2} \mathbf{K}_m(t) \vec{\nabla} T_a(t, \tau) d\tau\right)}_{=0} + \underbrace{\text{Div}\left(\frac{1}{P} \int_{-P/2}^{P/2} \mathbf{K}_a(t, \tau) \vec{\nabla} T(t + \tau) d\tau\right)}_{A(t)}. \end{aligned} \quad (3-15)$$

The left-hand term is developed as

$$\frac{\partial T_m}{\partial t} = \frac{\partial\left(\frac{1}{P} \int_{-P/2}^{P/2} T(t + \tau) d\tau\right)}{\partial t} = \frac{T(t + P/2) - T(t - P/2)}{P} = \frac{1}{P} \int_{-P/2}^{P/2} \frac{\partial T(t + \tau)}{\partial \tau} d\tau. \quad (3-16)$$

As regards the boundary conditions, the same considerations lead to

$$\begin{aligned} -\mathbf{K}_m(t) \vec{\nabla} T_m(t) \cdot \vec{dA} = H_m(t) dA (T_m(t) - T_\infty) + \underbrace{\frac{1}{P} \int_{-P/2}^{P/2} \mathbf{K}_m(t) \vec{\nabla} T_a(t, \tau) \cdot \vec{dA} d\tau}_{=0} \\ + \underbrace{\frac{1}{P} \int_{-P/2}^{P/2} (H_a(t, \tau) T_a(t, \tau) dA + \mathbf{K}_a(t, \tau) \vec{\nabla} T(t + \tau) \cdot \vec{dA}) d\tau}_{B(t)}. \end{aligned} \quad (3-17)$$

In the heat build-up problem of a rubber component, $A(t)$ and $B(t)$ are negligible. This approximation is justified by one assumption linking the two characteristic times: the mechanical characteristic time t_{mech} is small compared to the thermal diffusion time $t_{therm} = \rho c L^2 / k$ where L is the characteristic dimension.

Note that for rubber the thermal diffusivity is of the order of $k / \rho c \simeq 1.10^{-7}$ (Table 1). Typical mechanical time for heat build-up problem ranges between 0.1 s and 1 s. The characteristic thermal length corresponding to this mechanical time is then $\sqrt{k T_{mech} / (\rho c)} \simeq 0.3 \text{ mm} \ll L$. This means that a variation in the thermal problem during a mechanical cycle is limited at this scale. The conclusion to be drawn is that at the macroscopic scale (of the order of one millimeter), the temperature oscillation due

| Parameter | Value |
|------------------------|---|
| Heat generation | $2 \times 10^{-8} \text{ J m}^{-3}$ |
| Frequency | 10 Hz |
| Thermal conductivity | $0.23 \text{ W m}^{-1} \text{ K}^{-1}$ |
| Convection coefficient | $30 \text{ W m}^{-2} \text{ K}^{-1}$ |
| Temperature | 23 °C |
| Density | 1.1 |
| Heat capacity | $1.45 \times 10^3 \text{ J Kg}^{-1} \text{ K}^{-1} (23 \text{ °C})$ |
| Thermal diffusivity | $1.44 \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$ |

Table 1. Parameters used for the numerical simulation.

to the cyclic changes in the boundary conditions and in the conductivity is small compared to the mean temperature: $T_a \ll T_m$. The regularity of the heat equation leads to $\vec{\nabla} T_a \ll \vec{\nabla} T_m$. These two conditions imply that at the macroscopic scale, the terms $A(t)$ and $B(t)$ are negligible. The mean thermal problem is then properly defined on the reference configuration.

This approximation finally leads to this simplification of the heat equation

$$\rho_0 c \frac{\partial T_m}{\partial t} = \rho_0 R_m + \text{Div } \mathbf{K}_m(T_m) \vec{\nabla} T_m \quad (3-18)$$

with the boundary conditions

$$-\mathbf{K}_m(T_m) \vec{\nabla} T_m \cdot \vec{dA} = H_m dA (T_m - T_\infty), \quad (3-19)$$

where the heat flux is defined by $\vec{Q}_m = -\mathbf{K}_m(T_m) \vec{\nabla} T_m$.

This problem is defined on the reference configuration and the thermal parameters are constant. The thermal conductivity tensor is anisotropic and accounts for the mean path of the heat flux in the component over a cycle. The mean convection coefficient represents the mean of the exchange with ambient air. The resolution of this problem is also faster than the resolution of the coupled one, because the mechanical time does not appear anymore.

4. Numerical simulation

The simplified heat equation has been tested through FE modeling. The modeled specimen is a 5 mm radius cylinder. A sinusoidal deformation is imposed between stretches $\lambda = 1$ and $\lambda = 2$ (Figure 3). The heat source is homogeneous and temperature-independent. For the sake of simplicity, the thermal conductivity and the convection coefficient are supposed to be Eulerian constants. The value of the parameters are given in Table 1 and are typical for a carbon-black-filled rubber. In this example, the constitutive law has no importance. Four calculations have been undertaken:

- Fully-coupled calculation (50 increments per cycle).
- Uncoupled calculation on the maximal deformed geometry.
- Uncoupled calculation on the undeformed geometry.
- Uncoupled calculation with the simplified heat equation.

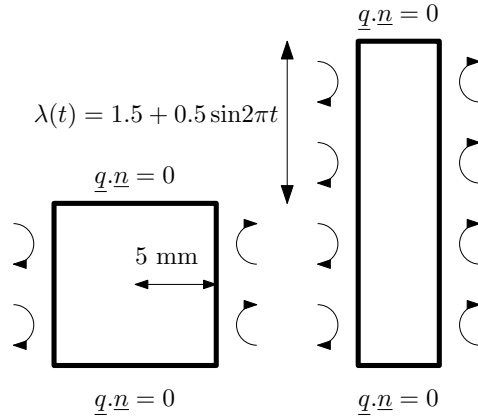


Figure 3. Undeformed and maximal deformed geometry of the 5 mm radius cylinder.

The results are presented in Figure 4. Whereas the uncoupled calculation on a fixed geometry implies an error of about 25% on the temperature compared to the fully-coupled solution, the error between the approximated solution and the coupled one is less than one percent and the two solutions can be considered to be equivalent. The time of calculation for the simplified problem is thus about three orders of magnitude smaller than for the fully-coupled problem for this simple specimen.

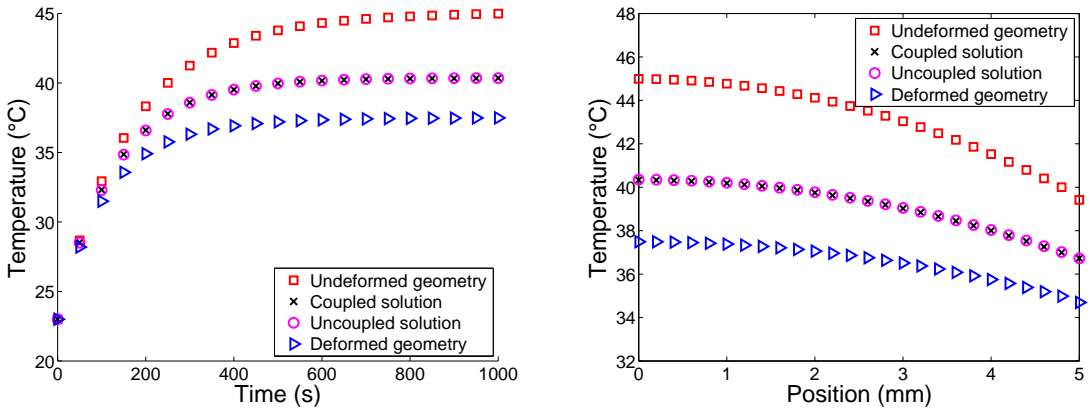


Figure 4. Left: Maximal temperature versus time for the four calculations. Right: Temperature versus radial position.

5. Conclusions

The heat build-up temperature has been defined as the actual temperature minus the temperature change due to thermoelastic effects. This heat build-up in rubber leads to a coupled thermomechanical problem. Due to the difference between the mechanical and thermal characteristic times, the quickest way to numerically solve this problem is to use a cyclic algorithm uncoupling the mechanical and the thermal problems. Unfortunately, the uncoupling approach does not account for the changes in geometry during the mechanical loading since the thermal problem is supposed to be solved on a fixed geometry. Under

the cyclic assumption, an approximation of the heat equation has been developed and allows the use of this algorithm at finite strains. This approximation is based on the integration of the heat equation and the boundary conditions over one cycle. These equations are written in the reference configuration so that the definition domain remains constant during one cycle. This approximation permits the use of the powerful uncoupling algorithm at finite strains: a numerical simulation illustrates the usefulness of this method.

Acknowledgements

The authors thank Vince Coveney from UWE in Bristol for helpful discussions.

References

- [Anthony et al. 1942] R. L. Anthony, R. H. Caston, and E. Guth, “Equations of state for natural and synthetic rubber-like materials, I: Unaccelerated natural soft rubber”, *J. Phys. Chem.* **46**:8 (1942), 826–840.
- [Armero and Simo 1992] F. Armero and J. C. Simo, “A new unconditionally stable fractional step method for non-linear coupled thermomechanical problems”, *Int. J. Numer. Methods Eng.* **35**:4 (1992), 737–766.
- [Broerman et al. 1999] A. W. Broerman, D. C. Venerus, and J. D. Schieber, “Evidence of the stress-thermal rule in an elastomer subjected to simple elongation”, *J. Chem. Phys.* **111**:15 (1999), 6965–6969.
- [van den Brule 1989] B. H. A. A. van den Brule, “A network theory for the thermal conductivity of an amorphous polymeric material”, *Rheol. Acta* **28**:4 (1989), 257–266.
- [van den Brule and O’Brien 1990] B. H. A. A. van den Brule and S. B. G. O’Brien, “Anisotropic conduction of heat in a flowing polymeric material”, *Rheol. Acta* **29**:6 (1990), 580–587.
- [Chadwick 1974] P. Chadwick, “Thermo-mechanics of rubberlike materials”, *Philos. Tr. R. Soc. S. A* **276**:1260 (1974), 371–403.
- [Ebbott et al. 1999] T. G. Ebbott, R. L. Hohman, J.-P. Jeusette, and V. Kerchman, “Tire temperature and rolling resistance prediction with finite element analysis”, *Tire Sci. Technol.* **27**:1 (1999), 2–21.
- [Gough 1805] J. Gough, “A description of a property of Caoutchouc, or Indian rubber”, *Mem. Lit. Phil. Soc. Manch. S. 2* **1** (1805), 288–295.
- [Holzapfel and Simo 1996] G. A. Holzapfel and J. C. Simo, “Entropy elasticity of isotropic rubber-like solids at finite strains”, *Comput. Methods Appl. Mech. Eng.* **132**:1–2 (1996), 17–44.
- [James and Guth 1943] H. M. James and E. Guth, “Theory of the elastic properties of rubber”, *J. Chem. Phys.* **11**:10 (1943), 455–481.
- [Joule 1859] J. P. Joule, “On some thermo-dynamic properties of solids”, *Philos. Tr. R. Soc.* **149** (1859), 91–131.
- [Kar and Bhowmick 1997a] K. K. Kar and A. K. Bhowmick, “High-strain hysteresis of rubber vulcanizates over a range of compositions, rates, and temperatures”, *J. Appl. Polym. Sci.* **65**:7 (1997), 1429–1439.
- [Kar and Bhowmick 1997b] K. K. Kar and A. K. Bhowmick, “Hysteresis loss in filled rubber vulcanizates and its relationship with heat generation”, *J. Appl. Polym. Sci.* **64**:8 (1997), 1541–1555.
- [Lion 1997] A. Lion, “A physically based method to represent the thermo-mechanical behaviour of elastomers”, *Acta Mech.* **123** (1997), 1–25.
- [Medalia 1991] A. I. Medalia, “Heat generation in elastomer compounds: causes and effects”, *Rubber Chem. Tech.* **64**:3 (1991), 481–492.
- [Meinecke 1991] E. Meinecke, “Effect of carbon-black loading and crosslink density on the heat build-up in elastomers”, *Rubber Chem. Tech.* **64**:2 (1991), 269–284.
- [Park et al. 2000] D. M. Park, W. H. Hong, S. G. Kim, and H. J. Kim, “Heat generation of filled rubber vulcanizates and its relationship with vulcanizate network structures”, *Eur. Polym. J.* **36**:11 (2000), 2429–2436.

- [Reese and Govindjee 1997] S. Reese and S. Govindjee, “Theoretical and numerical aspects in the thermo-viscoelastic material behaviour of rubber-like polymers”, *Mech. Time-Depend. Mat.* **1**:4 (1997), 357–396.
- [Sarkar et al. 1987] K. Sarkar, Y. D. Kwon, and D. C. Prevorsek, “A new approach for the thermomechanical analysis of tires by the finite element method”, *Tire Sci. Technol.* **15**:4 (1987), 261–275.
- [Simo and Taylor 1991] J. C. Simo and R. L. Taylor, “Quasi-incompressible finite elasticity in principal stretches: continuum basis and numerical algorithms.”, *Comput. Methods Appl. Mech. Eng.* **85**:3 (1991), 273–310.
- [Tautz 1959] H. Tautz, “Bestimmung der Wärmeleitfähigkeit von Kautschukvulkanisaten in Abhängigkeit von der Dehnung”, *Exp. Tech. Phys.* **7**:1 (1959), 1–14.
- [Toki et al. 2000] S. Toki, T. Fujimaki, and M. Okuyama, “Strain-induced crystallization of natural rubber as detected real-time by wide-angle X-ray diffraction technique”, *Polymer* **41**:14 (2000), 5423–5429.
- [Trabelsi et al. 2003] S. Trabelsi, P.-A. Albouy, and J. Rault, “Crystallization and melting processes in vulcanized stretch natural rubber”, *Macromolecules* **36**:20 (2003), 7624–7639.
- [Treloar 1975] L. R. G. Treloar, *The physics of rubber elasticity*, 3rd ed., Clarendon, Oxford, 1975.
- [Whicker et al. 1981a] D. Whicker, A. L. Browne, and D. J. Segalman, “The structure and use of the GMR combined thermo-mechanical tire power loss model”, *SAE Tech. Paper Series* **810164** (1981).
- [Whicker et al. 1981b] D. Whicker, A. L. Browne, D. J. Segalman, and L. E. Wickliffe, “A thermomechanical approach to tire power loss modeling”, *Tire Sci. Technol.* **9**:1 (1981), 3–18.

Received 19 Dec 2007. Revised 17 Dec 2008. Accepted 17 Dec 2008.

YOHAN LE CHENADEC: yohan.le-chenadec@polytechnique.edu

Laboratoire de Mécanique des Solides, CNRS UMR 7649, École Polytechnique, 91128 Palaiseau Cedex, France

IDA RAOULT: ida.raoult@mpsa.com

PSA Peugeot Citroën, Route de Gisy, 78943 Vélizy-Villacoublay Cedex, France

CLAUDE STOLZ: stolz@lms.polytechnique.fr

Laboratoire de Mécanique des Solides, CNRS UMR 7649, École Polytechnique, 91128 Palaiseau Cedex, France

MAC-LAN NGUYEN-TAJAN: maclan.nguyen@mpsa.com

PSA Peugeot Citroën, Route de Gisy, 78943 Vélizy-Villacoublay Cedex, France

INELASTIC HEAT FRACTION EVALUATION FOR ENGINEERING PROBLEMS INVOLVING DYNAMIC PLASTIC LOCALIZATION PHENOMENA

PATRICE LONGÈRE AND ANDRÉ DRAGON

The evaluation of the temperature produced during adiabatic dissipative processes for a large class of engineering materials (metals and some polymers) remains a major subject of interest, notably in the fields of high-speed machining and impact dynamics. The hypothesis consisting in considering the proportion of plastic work dissipated as heat (quantified by the inelastic heat fraction β) as independent on the loading path is now recognized as highly simplistic. Experimental investigations have shown indeed the dependence of the inelastic heat fraction on strain, strain rate and the temperature itself. The theoretical studies available nowadays are not entirely conclusive on various features regarding the history dependence and the evolution of β . The present work attempts to provide a systematic approach to the temperature rise and the inelastic heat fraction evolution for a general loading within the framework of thermoelastic/viscoplastic standard modelling including a number of quantitative variants regarding strain hardening/thermal softening and thermomechanical coupling description. The theoretical results thus obtained are confronted with experimental data from the literature. An analysis of the effects of various model simplifications on the evaluation of temperature growth with regard to conditions for dynamic plastic localization occurrence is also carried out. It is shown that the value of critical shear strain at localization incipience is strongly dependent on the level of simplification admitted.

1. Introduction

As a consequence of dynamic loading under (quasi) adiabatic conditions, many materials, and especially high strength metallic ones, are subject to plastic deformation localization in the form of narrow bands. This phenomenon, known as adiabatic shear banding (ASB), intervenes as thermal softening overcomes strain hardening and is often the precursor of structural failure. Many experimental as well as theoretical studies have been devoted to this deterioration process induced by thermal instability in terms of onset conditions, collective band behaviour and postcritical response [Zener and Hollomon 1944; Recht 1964; Marchand and Duffy 1988; Bai and Bodd 1992; Longère et al. 2005]. They all point out the crucial role of temperature. An accurate knowledge of the heat \dot{Q} generated by the plastic work rate \dot{W}^p during adiabatic dissipative evolution is thus needed. The quantity characterizing the proportion of plastic work dissipated as heat is commonly denoted β , such that $\dot{Q} = \beta \dot{W}^p$, and is usually called the *inelastic heat fraction Taylor–Quinney coefficient* [Taylor and Quinney 1934]. The engineering viewpoint consists in supposing that the inelastic heat fraction coefficient β remains constant during any process. When simulating numerically dynamic adiabatic problems using engineering computation codes, this assumption — which supposes that the coefficient β may be postulated a priori — is useful in the sense that it allows for accounting for thermomechanical couplings without solving the heat equation. In the literature, the value

Keywords: thermomechanics, viscoplasticity, inelastic heat fraction, nonlinear modelling, dynamics, localization.

of β typically ranges from 80 to 100% [Campagne et al. 2005; Guo et al. 2005] — in other words, between 80% or 100% of the plastic work is supposed to be converted into heat. However many experimental investigations have shown that the proportion of plastic work dissipated as heat is strongly dependent on temperature, strain and strain rate. Starting from this crude fact, this paper aims at proposing a more rigorous assessment of the temperature rate \dot{T} and the related inelastic heat fraction coefficient β (as an evolving quantity) for any three-dimensional loading path and thermoelastic/viscoplastic standard materials by using irreversible thermodynamics concepts in the context of adiabatic processes. As an application of this work, the consequences of various levels of simplification for evaluating both of the aforementioned quantities are studied in the context of ASB dynamic plastic localization onset conditions.

According to the extensive review by Bever et al. [1973] concerning the experimental investigations of the ratio a of the stored energy \dot{W}^s to the mechanical energy \dot{W} ($a = \dot{W}^s/\dot{W}$) during various plastic deformation processes, this ratio may be as expected strongly dependent on the loading path, on the material and on the deformation magnitude. In the present work we are rather interested in the energy adiabatically transformed into heat \dot{Q} as a linear function of the plastic deformation energy \dot{W}^p via the so-called inelastic heat fraction β ($\beta = \dot{Q}/\dot{W}^p$). Particularly useful in the numerical simulation of transient problems (crash, shock impact, etc.) for taking into account the thermomechanical coupling under adiabatic conditions without solving the thermal problem, the inelastic heat fraction remains an important subject of investigation, the accuracy of temperature measurement increasing thanks to the improvement of the relevant devices. In this context, some experimental determination of heating during plastic deformation at various strain rates tends to prove the dependence of the inelastic heat fraction on strain, strain rate and temperature [Chrysochoos et al. 1989; Kapoor and Nemat-Nasser 1998; Oliferuk et al. 2004]. Theoretical analyses [Aravas et al. 1990; Zehnder 1991] devoted to the subject agree with the tendency mentioned above but qualitative results appear contradictory [Mason et al. 1994] depending on the basic concepts used. The present work aims to propose an evaluation of the proportion of plastic work dissipated as heat from a broad spectrum of thermoelastic/viscoplastic models within the internal variables framework employing well-established thermodynamic concepts. With this aim in view, and starting from the assumption of the existence of free energy and dissipation potential and applying the first and second principles of thermodynamics, a class of characteristic constitutive three-dimensional thermoelastic/viscoplastic models is investigated with respect to their capacity to reproduce the evolution of β as a function of strain, strain rate and temperature. As the works by Rosakis et al. [2000] for one-dimensional constitutive modelling and Clayton [2005] for crystalline plasticity based modelling, this analysis strives to provide some complementary insight regarding heat generation versus dissipative mechanisms involving strain, strain rate and temperature effects in dynamic plasticity. Comparison of experimental and theoretical results concerning inelastic heat fraction evolution during some simple loading is discussed showing striking effects and paradoxes.

A further item treated in the present work is the analysis of the effects of model simplifications in the evaluation of temperature growth regarding conditions for dynamic plastic localization occurrence. These conditions are obtained from a simplified analysis based on the theory of linear perturbation [Bai 1982; Clifton et al. 1984] in the form of a criterion relating the resolved shear stress to strain hardening, thermal softening and viscosity parameters [Batra and Chen 2001; Longère et al. 2003] in the context of adiabatic shear banding localization. The purpose of the present study is not to discuss or extend the theory but to apply the method in order to obtain a practical criterion reproducing experimental

observation—the reader can refer to [Molinari 1985; Fressengeas and Molinari 1985; Anand et al. 1987] for extensive methodological background. Various levels of simplification of the heat equation formulation are consequently considered yielding corresponding criteria for ASB onset.

In Section 2 selected thermomechanical ingredients are recalled leading to the expression of the heat equation in the case of thermoelastic/viscoplastic modelling. The context considered in this work concerns loading at high strain rates for which adiabatic conditions can be assumed. Dissipative material behaviour is thus supposed to be (visco)plastic and pressure insensitive. The strain rates range considered is typically comprised between 10^2 and 10^4 s⁻¹, and maximum value of temperature remains much lower than the melting point.

In Section 3 several typical material behaviour models are studied including the combined effects of isotropic strain hardening/softening, thermal softening and plastic viscosity. The influence of strain rate and initial temperature as well as the influence of the functions used to describe the aforementioned effects on the temperature rate and on the inelastic heat fraction are shown. A confrontation of the theoretical results with experimental observations available in literature is also discussed.

In Section 4 three levels of simplification of heat equation are studied: first, due to its weak contribution to temperature change the thermoelastic coupling is neglected ; second, thermodissipative couplings are neglected as well ; finally the temperature rise is supposed to be directly linked to the plastic work rate via the inelastic heat fraction assumed to have a constant value. The consequences of these simplifications are examined in the context of thermal instability-induced dynamic plastic localization. With this aim in view, the linear perturbation method is applied to the case of the simple shearing of a volume element consisting of a thermo/viscoplastic material. The further parametric analysis shows how the value of the critical shear strain at localization incipience is strongly dependent on the temperature rise evaluation method.

2. Thermodynamic context and constitutive framework

The irreversible thermodynamics framework is used here to describe the thermoelastic-inelastic response of a material [Perzyna 1966; Bataille and Kestin 1975]. For the first time, expressions of dissipated energy and heat equation are given within the internal variable framework. The different levels of simplification of the heat equation are then introduced in connection with the further study of heat evaluation at the dynamic localization onset.

The instantaneous state of the material is described by a thermodynamic potential, namely the Helmholtz free energy per unit mass $\psi(T, z)$ where T represents absolute temperature—which is supposed to be a well defined entity within the thermodynamic framework employed—and z , a set of normal variables. Gibbs' relation takes the form:

$$\rho_0 \dot{\psi} = -\rho_0 s \dot{T} + Z \dot{z}, \quad s = -\left. \frac{\partial \psi}{\partial T} \right|_z, \quad Z = \rho_0 \left. \frac{\partial \psi}{\partial z} \right|_T, \quad (1)$$

where ρ_0 represents the mass density in the initial configuration, s the entropy and Z the set of thermodynamic (conjugate) forces associated with the state variables z .

According to the second law of thermodynamics, the mechanical part of dissipated energy is written as

$$D = \boldsymbol{\tau} : \boldsymbol{d} - Z \dot{z} \geq 0, \quad (2)$$

where \mathbf{d} represents the rate of deformation tensor and $\boldsymbol{\tau}$ the Kirchhoff stress tensor defined as the Cauchy stress tensor $\boldsymbol{\sigma}$ multiplied by the Jacobian determinant J of the deformation gradient \mathbf{F} ($J = \det \mathbf{F}$), that is, $\boldsymbol{\tau} = J\boldsymbol{\sigma}$.

Combination of the first law of thermodynamics and of Gibbs relation gives the local form of the heat equation as

$$\rho_0 c_z \dot{T} + J \operatorname{div} \mathbf{q} - r = \boldsymbol{\tau} : \mathbf{d} - \left(Z - T \frac{\partial Z}{\partial T} \right) \dot{z}, \quad (3)$$

where \mathbf{q} is the heat flux vector per unit area, r the heat supply per unit volume, and

$$c_z = -T \left. \frac{\partial^2 \psi}{\partial T^2} \right|_z$$

is the heat capacity at given z .

Equation (3) above relates thermal terms on the left side, including temperature rate \dot{T} , heat conduction $\operatorname{div} \mathbf{q}$, and heat supply r , to mechanical terms on the right side, including mechanical work rate $\boldsymbol{\tau} : \mathbf{d}$ and other work rates $Z\dot{z}$, while the last term $T\partial Z/\partial T\dot{z}$ is referred explicitly to thermomechanical couplings.

The study is now reduced to thermoelastic/viscoplastic behaviour in the case of isotropic strain hardening. The set of state variables is assumed in the form $z \equiv (\mathbf{e}^e, p)$, where \mathbf{e}^e represents a measure of moderate elastic strain, ($\mathbf{e}^e = \ln \mathbf{V}^e$; $\mathbf{F}^e = \mathbf{V}^e \mathbf{Q}$ where \mathbf{F}^e proceeds from the multiplicative split of the deformation gradient $\mathbf{F} = \mathbf{F}^e \mathbf{F}^p$ and where \mathbf{V}^e and \mathbf{Q} represent the pure elastic stretching and the frame rotation tensors respectively), and

$$p = \int_t \sqrt{\frac{2}{3} \mathbf{d}^p : \mathbf{d}^p} dt$$

the isotropic strain hardening variable. The set of conjugate forces is thus given in the form $Z \equiv (\boldsymbol{\tau}, R)$, where R represents the isotropic hardening force (affinity).

The rate of deformation (symmetric part of the velocity gradient) $\mathbf{d} = [\partial \mathbf{v} / \partial \mathbf{x}]^s$ is furthermore decomposed into a reversible ('elastic') part

$$\mathbf{d}^e = \overset{\nabla}{\mathbf{e}}^e = \dot{\mathbf{e}}^e - \boldsymbol{\omega} \mathbf{e}^e + \mathbf{e}^e \boldsymbol{\omega},$$

(where ∇ designates the objective Jaumann derivative of a 2nd order tensor) and an irreversible part \mathbf{d}^p such that $\mathbf{d} = \mathbf{d}^e + \mathbf{d}^p$ [Sidoroff and Dogui 2001].

The mechanical dissipation in Equation (2) becomes

$$D = \boldsymbol{\tau} : \mathbf{d}^p - R\dot{p} \geq 0. \quad (4)$$

The constitutive laws are thus expressed as

$$s = - \left. \frac{\partial \psi}{\partial T} \right|_{\mathbf{e}^e, p}, \quad \boldsymbol{\tau} = \rho_0 \left. \frac{\partial \psi}{\partial \mathbf{e}^e} \right|_{T, p}, \quad R = \rho_0 \left. \frac{\partial \psi}{\partial p} \right|_{T, \mathbf{e}^e}. \quad (5)$$

The free energy density $\psi(T, \mathbf{e}^e; p)$ is expressed in the form

$$\psi(T, \mathbf{e}^e; p) = \psi^e(T, \mathbf{e}^e) + \psi^{\text{th}}(T) + \psi^b(T; p), \quad (6)$$

where $\psi^e(T, \mathbf{e}^e)$ is the recoverable energy including the isotropic linear thermoelasticity, $\psi^{\text{th}}(T)$ is the purely thermal energy, and $\psi^b(T; p)$ is the stored energy, which reflects the competition in the material

between isotropic strain hardening and thermal softening. These contributions are written as

$$\begin{aligned}\rho_0 \psi^e(T, \mathbf{e}^e) &= \frac{\lambda}{2} (\text{Tr } \mathbf{e}^e)^2 + \mu \mathbf{e}^e : \mathbf{e}^e - \alpha K (\text{Tr } \mathbf{e}^e) \vartheta, \\ \rho_0 \psi^{\text{th}}(T) &= -\rho_0 c_0 \left(T \ln \frac{T}{T_0} - \vartheta \right) - h(0) f(T), \\ \rho_0 \psi^b(T, p) &= h(p) f(T) - h(0) f(0),\end{aligned}\quad (7)$$

where λ and μ represent the Lamé elasticity constants, $K = \lambda + \frac{2}{3}\mu$ the bulk modulus, α the thermal dilatation coefficient, c_0 a thermal constant relative to heat capacity, $\vartheta = T - T_0$ the temperature rise, $h(p)$ the stored energy of cold work and $f(T)$ the thermal softening function. For simplification, the thermoelasticity coefficients and c_0 are supposed to be temperature independent.

From now on the elastic deformations will be considered as small (infinitesimal).

After partial derivation of Equation (6) with respect to (T, \mathbf{e}^e, p) , thermodynamic forces $(-s, \boldsymbol{\tau}, R)$ are written as

$$-\rho_0 s = -\alpha K \text{Tr } \mathbf{e}^e - \rho_0 c_0 \ln \frac{T}{T_0} - [h(0) - h(p)] f'(T), \quad (8)$$

$$\boldsymbol{\tau} = (\lambda \text{Tr } \mathbf{e}^e - \alpha K \vartheta) \boldsymbol{\delta} + 2\mu \mathbf{e}^e, \quad (9)$$

$$R = h'(p) f(T). \quad (10)$$

On the other hand, the heat capacity

$$c_z = -T \left. \frac{\partial^2 \psi}{\partial T^2} \right|_z = T \left. \frac{\partial s}{\partial T} \right|_z$$

takes the form

$$c_z = c_0 + T \frac{[h(0) - h(p)]}{\rho_0} f''(T).$$

Assuming a negligible contribution of the thermoplastic coupling term of this expression (this term is always zero since $f(T)$ is an affine function), the heat capacity is supposed to be constant, that is, $c_z = c_0$.

By rewriting the mechanical dissipation equation (4) in the form $D = \bar{\mathbf{Z}} \dot{\bar{\mathbf{z}}} \geq 0$, $\bar{\mathbf{Z}} = (\boldsymbol{\tau}, R)$ represents the set of thermodynamic forces associated to the set of fluxes $\dot{\bar{\mathbf{z}}} = (\mathbf{d}^p, -\dot{p})$. Considering now time-dependent plasticity and applying the normality rule with respect to the dual $\Psi(\bar{\mathbf{Z}})$ of the dissipation potential, depending on $\bar{\mathbf{Z}}$ via the loading (yield) function $F(\bar{\mathbf{Z}})$ (that is, standard rule for $\dot{\bar{\mathbf{z}}}$), internal state variable evolution laws are deduced from:

$$\dot{\bar{\mathbf{z}}} = \Lambda \frac{\partial F(\bar{\mathbf{Z}})}{\partial \bar{\mathbf{Z}}}, \quad \Lambda = \frac{\partial \Psi(F)}{\partial F} \geq 0, \quad (11)$$

where Λ represents the viscous multiplier governing dissipative mechanisms of plasticity.

Applying the normality rule to the present case yields

$$\mathbf{d}^p = \Lambda \frac{\partial F}{\partial \boldsymbol{\tau}}; \quad -\dot{p} = \Lambda \frac{\partial F}{\partial R}. \quad (12)$$

The yield function $F(\bar{Z})$ in Equation (11) is assumed in the form of a Huber–von Mises criterion:

$$F(\boldsymbol{\tau}, R; T) = J_2(\boldsymbol{\tau}) - g(R, T), \quad (13)$$

where $J_2(\boldsymbol{\tau}) = \sqrt{\frac{3}{2} \boldsymbol{s} : \boldsymbol{s}}$ and $\boldsymbol{s} = \boldsymbol{\tau} - \frac{\text{Tr} \boldsymbol{\tau}}{3} \boldsymbol{\delta}$ represents the stress deviator tensor.

The strain hardening function $g(R, T)$ in Equation (13) representing the Huber–von Mises surface radius is expressed by

$$g(R, T) = R_0 f(T) + R(T, p). \quad (14)$$

Evolution laws in Equation (12) are thus detailed as

$$\boldsymbol{d}^p = \frac{3}{2} \Lambda \frac{\boldsymbol{s}}{J_2}; \quad \dot{p} = \Lambda. \quad (15)$$

The model is completed by the expression of the force potential in (11):

$$\Psi(F) = \frac{Y}{m+1} \left\langle \frac{F}{Y} \right\rangle^{m+1}, \quad \Lambda = \left\langle \frac{F}{Y} \right\rangle^m = H(F). \quad (16)$$

Inverting Equation (16)₂ and using (13), (14) and (15) yield

$$J_2 = R_0 f(T) + R(T, p) + \Phi(T, p, \dot{p}) \quad (17)$$

with $\Phi = H^{-1}$. According to Equation (15)_{1,2}, the rate of plastic work $\boldsymbol{\tau} : \boldsymbol{d}^p$ in (4) is thus given by:

$$\boldsymbol{\tau} : \boldsymbol{d}^p = J_2 \dot{p} = [R_0 f(T) + R(T, p) + \Phi(T, p, \dot{p})] \dot{p}. \quad (18)$$

Furthermore Equation (3) becomes

$$\rho_0 c_0 \dot{T} + J \text{div} \boldsymbol{q} - r = \boldsymbol{\tau} : \boldsymbol{d}^p + T \frac{\partial \boldsymbol{\tau}}{\partial T} : \boldsymbol{d}^e - \left(R - T \frac{\partial R}{\partial T} \right) \dot{p}. \quad (19)$$

The context considered here concerns loading at high strain rate excluding heat supply and for which conditions can be assumed as adiabatic (dynamic loading). Relation (19) is thus reduced to

$$\rho_0 c_0 \dot{T} = \boldsymbol{\tau} : \boldsymbol{d}^p + T \frac{\partial \boldsymbol{\tau}}{\partial T} : \boldsymbol{d}^e - \left(R - T \frac{\partial R}{\partial T} \right) \dot{p}. \quad (20)$$

In Equation (20), $\boldsymbol{\tau} : \boldsymbol{d}^p$ represents the plastic part of the mechanical work rate (plastic work rate), $R \dot{p}$ the stored energy rate (the difference $\boldsymbol{\tau} : \boldsymbol{d}^p - R \dot{p}$ represents the unrecoverable energy rate dissipated by heating; see Equation (4)), $T (\partial \boldsymbol{\tau} / \partial T) : \boldsymbol{d}^e$ the thermoelastic coupling contribution which describes cooling during a tensile loading and heating during a compressive one, and $T (\partial R / \partial T) \dot{p}$ the thermo-plastic coupling contribution which expresses the stored energy release rate during the temperature rise (see [Clayton 2005] for a similar approach).

In the following, the effects of strain hardening and thermal softening (acting both on isotropic hardening force) on temperature rise are studied. Thermoelastic coupling contribution to temperature rise is actually particularly significant in problems involving very high velocity impact and/or high pressure shock loading. In the context of this work, velocity and pressure are considered moderate and thermoelastic coupling is neglected.

Accounting for Equation (18), relation (20) is thus reduced to

$$\dot{T} = \frac{1}{\rho_0 c_0} \left[J_2 - \left(R - T \frac{\partial R}{\partial T} \right) \right] \dot{p}. \quad (21)$$

Injecting the expression (17) of the stress invariant J_2 and the expression (10) for R yields

$$\dot{T} = \frac{R_0 f(T) + T h'(p) f'(T) + \Phi(T, p, \dot{p})}{\rho_0 c_0} \dot{p}. \quad (22)$$

Remark 1. Comparing the dissipation form in Equations (4) and (22) yields

$$\rho_0 c_0 \dot{T} = D + T h'(p) f'(T) \dot{p}. \quad (23)$$

The dissipation is positive while the second term on the right side is negative (f' being negative for thermal softening material). The temperature increases if the dissipation is high enough to exceed the thermomechanical coupling term which proceeds from $T (\partial R / \partial T) \dot{p}$ in Equation (21). The condition for temperature growth is written from (23) as

$$R_0 f(T) + \Phi(T, p, \dot{p}) > T h'(p) f'(T). \quad (24)$$

For $R_0 = 0$, there is thus a competition between viscous and thermomechanical effects regarding temperature changes. Neglecting this last contribution leads to the common relation

$$\rho_0 c_0 \dot{T} = D. \quad (25)$$

Furthermore, starting from the definition of the inelastic heat fraction β as $\beta = \rho_0 c_y \dot{T} / \boldsymbol{\tau} : \mathbf{d}^p$ and under the aforementioned assumptions, the following expression is deduced from (21):

$$\beta = 1 - \frac{R - T \frac{\partial R}{\partial T}}{J_2}. \quad (26)$$

Injecting the complete expressions of J_2 and R from (17) and (10) gives

$$\beta = 1 - \frac{h'(p) [f(T) - T f'(T)]}{[R_0 + H'(p)] f(T) + \Phi(T, p, \dot{p})}. \quad (27)$$

Consequently the inelastic heat fraction β is explicitly a function of temperature, strain and strain rate. Starting from relation (27), a series of comments can be made.

Remark 2. For cumulated plastic strain $p_2 > p_1$ close enough to consider that $T_1 \approx T_2 \approx T$, one can write from (27):

$$\beta(T, p_2, \dot{p}) - \beta(T, p_1, \dot{p}) \approx - [h'(p_2) - h'(p_1)] \frac{f(T) - T f'(T)}{[R_0 + h'(p)] f(T) + \Phi(T, p, \dot{p})}. \quad (28)$$

Using the notation

$$\chi = \frac{f(T) - T f'(T)}{[R_0 + h'(p)] f(T) + \Phi(T, p, \dot{p})}, \quad (29)$$

with $\chi > 0$ as long as $f(T) - Tf'(T) > 0$, which is the case for most (thermal softening) materials, relation (28) is reduced to

$$\frac{\partial\beta}{\partial p} \approx -\chi h''(p). \quad (30)$$

According to (30), it is possible to conclude that for a material exhibiting strain hardening ($h''(p) > 0$), the inelastic heat fraction is decreasing for increasing strain, that is, $\partial\beta/\partial p < 0$. Oppositely, for material exhibiting strain softening ($h''(p) < 0$), the inelastic heat fraction is increasing for increasing strain, that is, $\partial\beta/\partial p > 0$.

Remark 3. According to (27), $\beta(T, p, \dot{p})$ is equal to unity when $h'(p)[f(T) - Tf'(T)] = 0$ which is satisfied for a perfectly plastic material ($h'(p) = 0$) or for a material exhibiting a linear dependence of strain hardening on temperature ($f(T) = AT$). Finally, $\beta = 1$ for any loading path allowing for $f(T) = Tf'(T)$.

Remark 4. Suppose now that $f(T_0) = \varsigma$ ($0 < \varsigma < 1$), $f'(T_0) = -\zeta$ ($\zeta > 0$), and let us note that $h'(0) = R_1 > 0$.

In this case, according to (27),

$$\beta_0 = \beta(T_0, 0, \dot{p}) = 1 - \frac{R_1(\varsigma + \zeta T_0)}{(R_0 + R_1)\varsigma + \Phi(T_0, 0, \dot{p})},$$

and for any $R_1 \neq 0$, one obtains $\beta_0 < 1$.

Remark 5. Suppose finally that $f(T) = 1 - AT$ ($A > 0$), $R_0 = 0$ and $\Phi(T, p, \dot{p}) = h'(p)f(T)z(\dot{p})$.

In this case

$$\beta(T, p, \dot{p}) = 1 - \frac{1}{f(T)[1 + z(\dot{p})]} = \beta(T, \dot{p}).$$

Thus β does not depend on the plastic strain p . On the other hand, for a given strain rate \dot{p} , as $f(T)$ decreases with increasing temperature T , then β decreases with increasing temperature, and for a given temperature T , as $z(\dot{p})$ increases with increasing strain rate \dot{p} , then β increases with increasing strain rate \dot{p} .

By employing the inelastic heat fraction considered as a constant, heat equation is reduced to

$$\dot{T} = \frac{\beta}{\rho_0 c_0} J_2 \dot{p}. \quad (31)$$

Injecting (17) into (31) yields

$$\dot{T} = \beta \frac{[R_0 + h'(p)]f(T) + \Phi(T, p, \dot{p})}{\rho_0 c_0} \dot{p}. \quad (32)$$

Expression (32) above has to be compared with expression (22). In particular, the temperature rate here is always positive (see Remark 1).

3. Application to usual models

Following conventional experimental procedures, material constants for most models are identified from a series of tests carried out at various strain rates for a given initial temperature and/or at different temperatures for a given nominal (generally low) strain rate. In this context, the respective contributions

of separated effects of temperature, strain and strain rate can be determined during the material constant numerical identification. The latter needs, for high strain rate (adiabatic) tests, the evaluation of heating produced by the plastic work, that is, an expression for the inelastic heat fraction.

In this section, various thermodynamic and/or engineering models are studied in terms of stress invariant, temperature and inelastic heat fraction evolutions. Independently of the genuine temperature evolution in the material, the influence of the choice of the model is pointed out. As usually assumed, adiabatic conditions are supposed to prevail at strain rates higher than 100 s^{-1} .

According to the methodology detailed previously, a series of mathematical functions available in literature has been selected and reported in the following tables.

Tables 1–3 give the usual functions describing the effects of strain hardening/softening, thermal softening and strain rate, respectively. Corresponding abbreviations used later in the analysis are also given.

The Ramberg–Osgood (HRO) model in Table 1 for strain hardening has the form of a power law. For $n = 0$, stored energy is linear (HL with $\text{HL} = \text{HRO}_{n=0}$) with respect to cumulated plastic strain while the material is perfectly plastic; for $n = 1$, stored energy evolves as a quadratic function (HQ with

| | Ramberg–Osgood HRO | Voce HV | Strain softening HS |
|----------|-------------------------|--|---------------------------------|
| $h(p)$ | $k \frac{p^{n+1}}{n+1}$ | $R_\infty [p + \frac{1}{k} \exp(-kp)]$ | $-\frac{R_\infty}{k} \exp(-kp)$ |
| $h'(p)$ | kp^n | $R_\infty [1 - \exp(-kp)]$ | $R_\infty \exp(-kp)$ |
| $h''(p)$ | nkp^{n-1} | $kR_\infty \exp(-kp)$ | $-kR_\infty \exp(-kp)$ |

Table 1. Usual strain hardening/softening functions $h(p)$.

| | Linear FL | Power FP | Exponential FE |
|----------|-----------|---|---------------------------------|
| $f(T)$ | $1 - AT$ | $1 - \left(\frac{T - T_r}{T_m - T_r}\right)^m$ | $\exp(-\gamma T)$ |
| $f'(T)$ | $A > 0$ | $-\frac{m}{T_m - T_r} \left(\frac{T - T_r}{T_m - T_r}\right)^{m-1}$ | $-\gamma \cdot \exp(-\gamma T)$ |
| $f''(T)$ | 0 | $\frac{m(1-m)}{(T_m - T_r)^2} \left(\frac{T - T_r}{T_m - T_r}\right)^{m-2}$ | $\gamma^2 \exp(-\gamma T)$ |

Table 2. Usual thermal softening functions $f(T)$.

| | additive type ADD | multiplicative type MUL | Exponential type EXP |
|---------------------------------|------------------------------------|----------------------------------|---------------------------------------|
| $\dot{p} = H \langle F \rangle$ | $1/\tau_0 \langle F/R_0 \rangle^m$ | $1/\tau_0 \langle F/R \rangle^m$ | $\dot{p}_0 \exp \langle F/CR \rangle$ |
| τ_0 | $\langle Y/R_0 \rangle^m$ | τ_0 | $1/\dot{p}_0$ |
| $\Phi = H^{-1}$ | $Y \dot{p}^{1/m}$ | $R \tau_0^{1/m} \dot{p}^{1/m}$ | $CR \ln(\dot{p}/\dot{p}_0)$ |

Table 3. Usual strain rate functions $H \langle F \rangle$.

HQ = HRO_{n=1}) of cumulated plastic strain and hardening is linear with respect to cumulated plastic strain; for most metals the value of n is comprised between 0 and 1. The Voce (HV) model describes strain hardening with saturation stress, the latter being represented by R_∞ . Finally, the strain softening model (HS) is retained in this study as an exponential decreasing function. On the other hand, the combination of several strain hardening/softening functions is designated by linking corresponding abbreviations. HMLRO designates, for example, hardening/softening behaviour (H) described using multiple (M) functions including linear hardening (L) and Ramberg–Osgood (RO). HMSQ designates hardening/softening behaviour (H) described using multiple (M) functions including softening (S) and the quadratic hardening Ramberg–Osgood (Q) function.

The linear function (FL) in Table 2 for thermal softening can be considered as an approximation, for low temperature changes, of a more complex dependence of material behaviour on temperature. The power law (FP) accounts for melting point T_m and room temperature T_r in a power law expression as proposed notably by Johnson and Cook [1983]. Finally a third thermal model (FE) is proposed in the form of an exponential decreasing function which is valid in a wider range of temperature variations than linear function (FL).

The additive type (ADD) and multiplicative type (MUL) functions in Table 3 are used to describe phenomenologically the effects of plastic viscosity in bcc/hcp and fcc metals respectively. The origin of the exponential type function (EXP) proceeds from physical considerations including the concept of thermally activated plastic deformation mechanism.

In the multiplicative type (MUL) in Table 3, the strain rate function

$$\Phi(T, p, \dot{p}) = R \tau_0^{1/m} \dot{p}^{1/m}$$

can be rewritten in the form

$$\Phi(T, p, \dot{p}) = h'(p) f(T) g(\dot{p}),$$

where $g(\dot{p}) = \tau_0^{1/m} \dot{p}^{1/m}$. In the same way, in the exponential type (EXP) in Table 3, the strain rate function

$$\Phi(T, p, \dot{p}) = C R \ln(\dot{p}/\dot{p}_0)$$

can be rewritten in the form

$$\Phi(T, p, \dot{p}) = h'(p) f(T) g(\dot{p}),$$

where $g(\dot{p}) = C \ln(\dot{p}/\dot{p}_0)$. These expressions may illustrate Remark 5.

Table 4 gives the abbreviations corresponding to models with or without initial yield stress, that is, an initial (or no) nonzero value for the radius of the Huber–von Mises surface, see (14). Note that this concept of initial yield stress may be entirely contained in the state potential (free energy) through a contribution attributed to residual stresses induced by thermomechanical processes and described via a quadratic strain hardening (HRO type model with $n=1$; see Table 1). As shown previously, the retained viewpoint (nonzero value for the initial radius of the Huber–von Mises surface or residual stresses) is not without consequence for heat evaluation (see (22) and Remark 1).

Remark 6. Consider the combination HL-FE-WYS (see Tables 1, 2 and 4 with HL = HRO_{n=0}) such that $\rho_0 \psi^b = R_1 p \exp(-\gamma T)$. The yield function is written as

$$F = J_2(\sigma) - [R_0 \exp(-\gamma T) + R].$$

| | |
|-------------------------------|----------------------------------|
| With initial yield stress WYS | Without initial yield stress NYS |
| $R_0 f(T)$ | 0 |

Table 4. Existence of initial yield stress.

The strain hardening force and its derivative with respect to temperature are, respectively,

$$R = \rho_0 \partial \psi^b / \partial p = R_1 \exp(-\gamma T), \quad T \partial R / \partial T = -\gamma T R.$$

In this case, $f(T_0) = \exp(-\gamma T_0) > 0$ and $f'(T_0) = -\gamma \exp(-\gamma T_0)$. As mentioned in [Remark 4](#), the initial inelastic heat fraction β_0 is lower than unity.

In order to analyse the three-dimensional constitutive models built from the combination of strain hardening and thermal softening functions in [Tables 1–4](#), the loading path considered here is the simple shearing. Designate by $\mathbf{l} = \partial \mathbf{v} / \partial \mathbf{x}$ the velocity gradient tensor and l_{ij} its components. We are considering here the simple shear loading such that $v_1 = \dot{\Gamma} x_2$. In this case, the components of the velocity gradient tensor \mathbf{l} are zero except $l_{12} = \partial v_1 / \partial x_2 = \dot{\Gamma} \neq 0$.

A set of thermoelastic/viscoplastic models has been tested including various forms for strain hardening, strain softening and their combinations. They all use the physical material constants reported in [Table 5](#).

Evolution of the second invariant of the stress-deviator J_2 , temperature T and inelastic heat fraction β is given versus shear strain $e_{12} = \gamma_{12}/2$ (the strain tensor \mathbf{e} is obtained by time integration of the nonobjective strain rate tensor $\dot{\mathbf{e}}$, with $\dot{\mathbf{e}} = \mathbf{d} + \boldsymbol{\omega} \mathbf{e} - \mathbf{e} \boldsymbol{\omega}$) at various strain rates $\dot{\Gamma}$ and initial temperatures T_0 in [Figures 1–5](#) for several models including strain hardening/softening, thermal softening and viscous effects. Adiabatic conditions are assumed for strain rates higher than 100 s^{-1} .

The physical feature of the numerical results obtained is further discussed in [Section 3.1](#) by considering experimental data from literature.

3.1. Strain hardening models. Consider the combination HV-FL-ADD-WYS (see [Tables 1–4](#)) summarized in [Table 6](#). As mentioned before, this combination may correspond to a bcc metal with saturating strain hardening and linear approximation of temperature dependence ([Figure 1](#)), such as a high strength martensitic steel.

[Figure 1a](#) shows the thermal softening induced under adiabatic conditions for simulations at different plastic strain rates higher than 100 s^{-1} . [Figure 1b](#) shows that the temperature rate is higher for higher strain rate, while [Figure 1e](#) shows that the temperature rate is lower for higher initial temperature. At 1000 s^{-1} and for $T_0 = 300 \text{ K}$, the maximal value of temperature increase is close to 60 K .

According to [Figure 1c–f](#), the initial value of β is equal to 1 whatever the strain rate and the initial temperature. At large strain β converges to a value which depends on strain rate and initial temperature with a rate (negative according to [Remark 3](#)) whose absolute value increases with decreasing strain

| E (GPa) | ν | ρ_0 (kg/m ³) | c_0 (J/kg.K) | α (K ⁻¹) |
|---------|-------|-------------------------------|----------------|-----------------------------|
| 200 | 0,33 | 7800 | 420 | 10^{-6} |

Table 5. Physical steel like material constants.

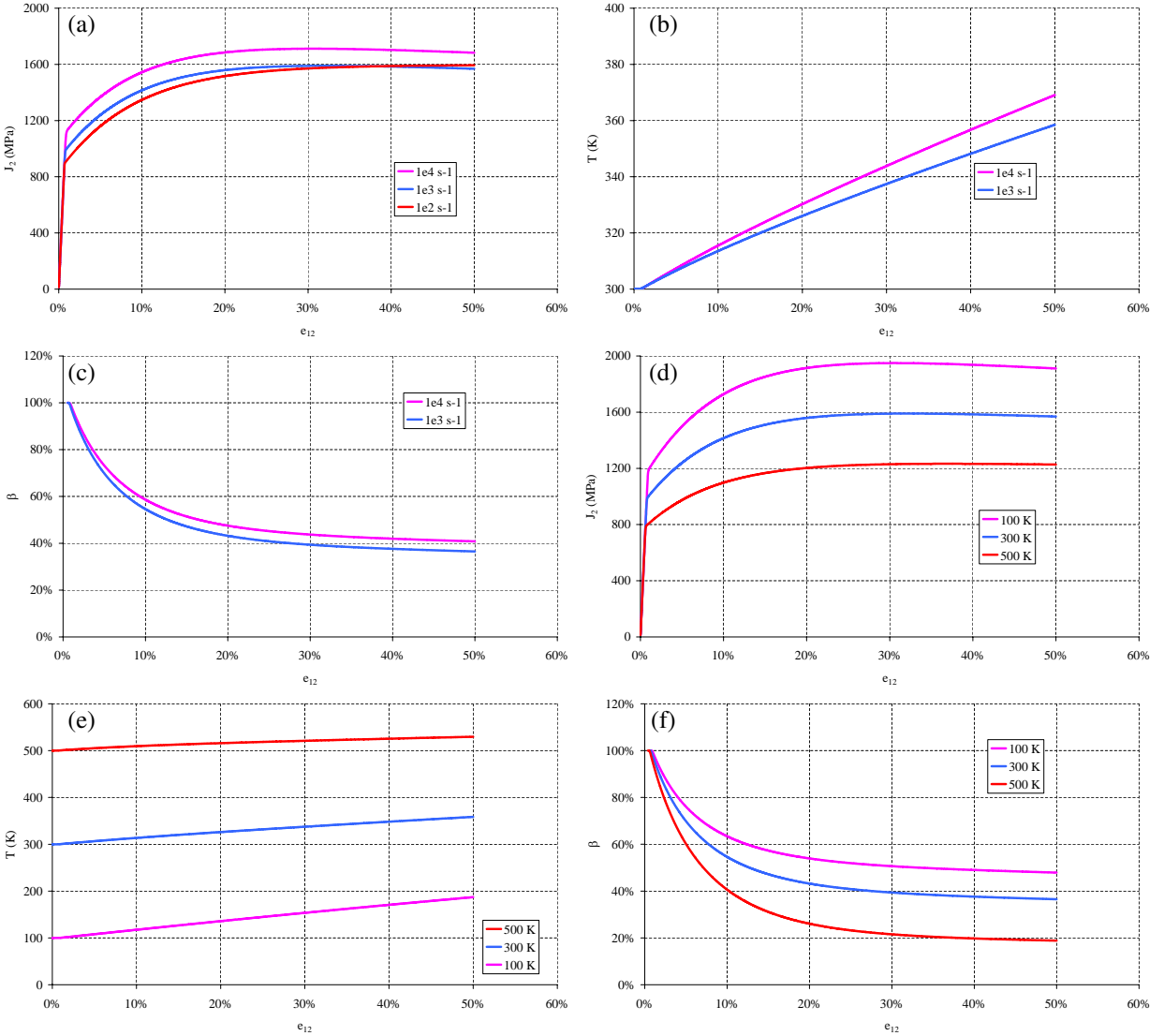


Figure 1. Influence of shear strain rate $\dot{\Gamma}$ and initial temperatures T_0 on stress invariant, heating and inelastic heat fraction for the HV-FL-ADD-WYS hardening model: (a) stress invariant J_2 versus strain e_{12} , with $T_0 = 300$ K; (b) temperature T versus strain e_{12} , with $T_0 = 300$ K; (c) inelastic heat fraction β versus strain e_{12} , with $T_0 = 300$ K; (d) stress invariant J_2 versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (e) temperature T versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (f) inelastic heat fraction β versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$.

rate and increasing initial temperature. In other words, β depends explicitly on strain, strain rate and temperature.

Consider next the combination HMLRO-FE-MUL-NYS (see Tables 1–4) summarized in Table 7. This combination may correspond to a fcc metal with a power law strain hardening ($n = 0.5$) and exponential dependence on temperature (Figure 2), such as an austenitic steel or a pure copper.

| | | | | | |
|---|-----|------------------|-----------------------|-----------------------------|-----|
| $h(p)$ | | $f(T)$ | $\Phi(T, p, \dot{p})$ | R_0 | |
| $R_\infty \left[p + \frac{1}{k} \exp(-kp) \right]$ | | $1 - AT$ | $Y \dot{p}^{1/m}$ | $R_2(1 - AT)$ | |
| R_∞ (MPa) | k | $A(K^{-1})$ | R_2 (MPa) | Y (MPa s ^{1/m}) | m |
| 1000 | 20 | 10 ⁻³ | 1000 | 100 | 6 |

Table 6. Thermoviscoplastic material constants for HV-FL-ADD-WYS model.

In contrast to the previous case, the thermal softening induced under adiabatic conditions for simulations at plastic strain rates greater than 100 s⁻¹ does not seem to be really significant (see Figure 2a). In fact, adiabatic thermal effects are present but they are here not able to compensate for strain and strain rate hardening. Figure 2b shows that the temperature rate is higher for higher strain rate while Figure 2e shows that the temperature rate is lower for higher initial temperature. At 1000 s⁻¹ and for $T_0 = 300$ K, the maximal value of temperature increase is close to 28 K.

According to Figure 2c–f, the initial value of β is much lower than 1, increasing with increasing strain rate and with decreasing initial temperature. As expected (see Remark 5), the rate of β is close to zero— β does not depend explicitly on strain.

Consider finally the Johnson–Cook engineering model [Johnson and Cook 1983] as the combination HMLRO-FP-EXP-NYS (see Tables 1–4) summarized in Table 8. This model, supposed to describe any metal behaviour, is mainly employed in numerical simulations involving very high strain rates as encountered notably in plate impact and shock.

In Figure 3a the thermal softening induced under adiabatic conditions for simulations at plastic strain rates higher than 100 s⁻¹ is not really significant. The temperature rate is seen to be higher for higher

| | | | | | | | | | |
|---------------------------------|-------------------|--------------------------------|-------|-------------|------|-----|-----------------------------|------------------|-----|
| $h(p)$ | $f(T)$ | $\Phi(T, p, \dot{p})$ | R_0 | R_1 (MPa) | k | n | γ (K ⁻¹) | τ_0 (s) | m |
| $R_1 p + k \frac{p^{n+1}}{n+1}$ | $\exp(-\gamma T)$ | $R \tau_0^{1/m} \dot{p}^{1/m}$ | 0 | 500 | 1500 | 0.5 | 10 ⁻³ | 10 ⁻⁴ | 10 |

Table 7. Thermoviscoplastic material constants for HMLRO-FE-MUL-NYS model.

| | | | | | | | |
|----------------------------|-----------|--|-------------------------------------|-----------|-----|--------------------------------|------|
| $h(p)$ | | $f(T)$ | $\Phi(T, p, \dot{p})$ | R_0 | | | |
| $R_1 p + k(p^{n+1}/n + 1)$ | | $1 - \left(\frac{T - T_r}{T_f - T_r} \right)^m$ | $C R \ln \frac{\dot{p}}{\dot{p}_0}$ | 0 | | | |
| R_1 (MPa) | k (MPa) | n | T_f (K) | T_r (K) | m | \dot{p}_0 (s ⁻¹) | C |
| 600 | 1000 | 0.3 | 1800 | 100 | 1 | 1 | 0,05 |

Table 8. Thermoviscoplastic material constants for Johnson–Cook model.

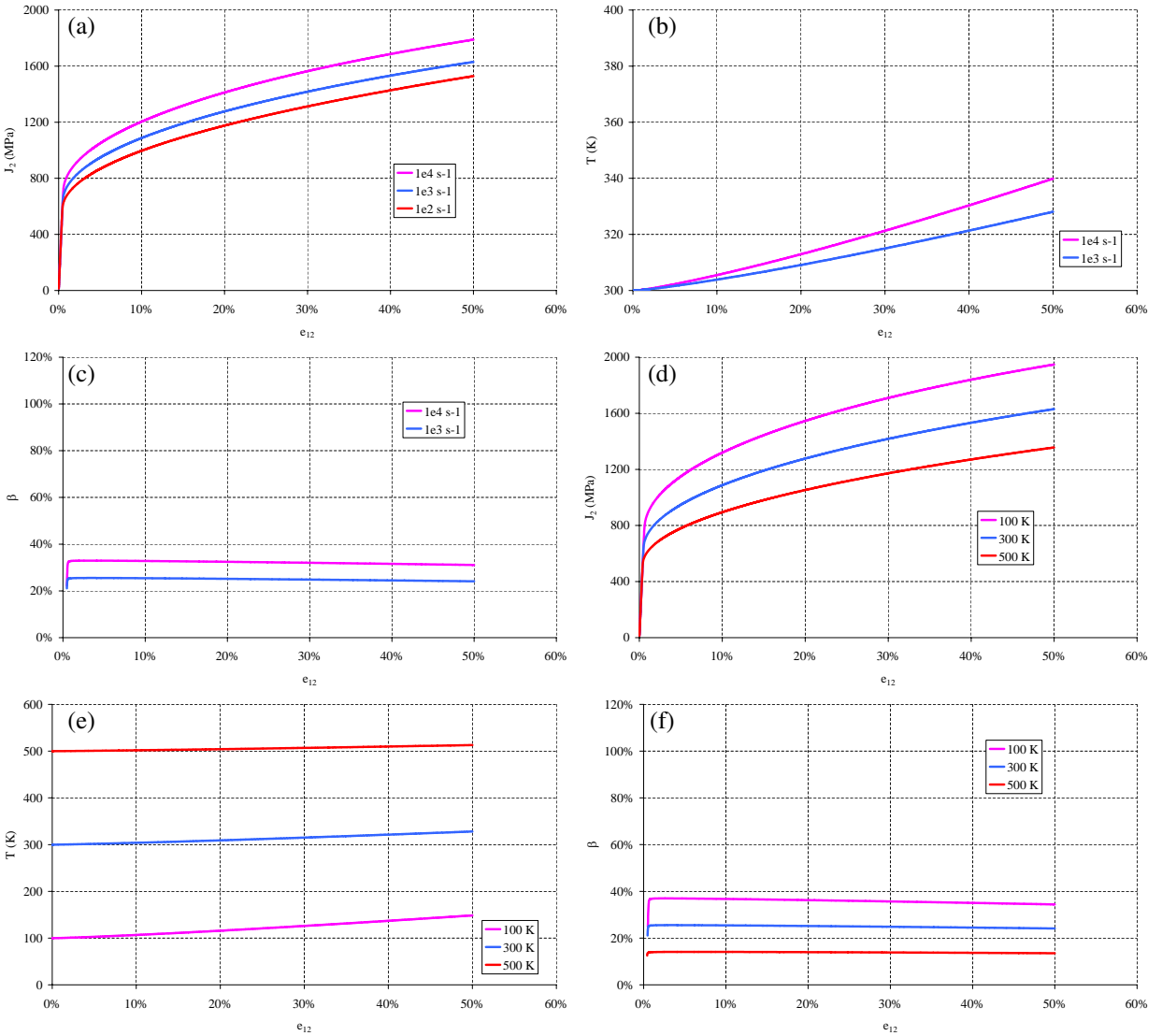


Figure 2. Influence of shear strain rate $\dot{\Gamma}$ and initial temperatures T_0 on stress invariant, heating and inelastic heat fraction for the HMLRO-FE-MUL-NYS hardening model: (a) stress invariant J_2 versus strain e_{12} , with $T_0 = 300$ K; (b) temperature T versus strain e_{12} , with $T_0 = 300$ K; (c) inelastic heat fraction β versus strain e_{12} , with $T_0 = 300$ K; (d) stress invariant J_2 versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (e) temperature T versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (f) inelastic heat fraction β versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$.

strain rate (Figure 3b) and lower for higher initial temperature (Figure 3e). At 1000 s $^{-1}$ and for $T_0 = 300$ K, the maximum value of temperature increase is close to 10 K, which is not significant.

According to Figure 3c–f, the initial value of β is much lower than 1, increasing with increasing strain rate and with decreasing initial temperature. As in the previous case, the rate of β is close to zero (see also Remark 6).

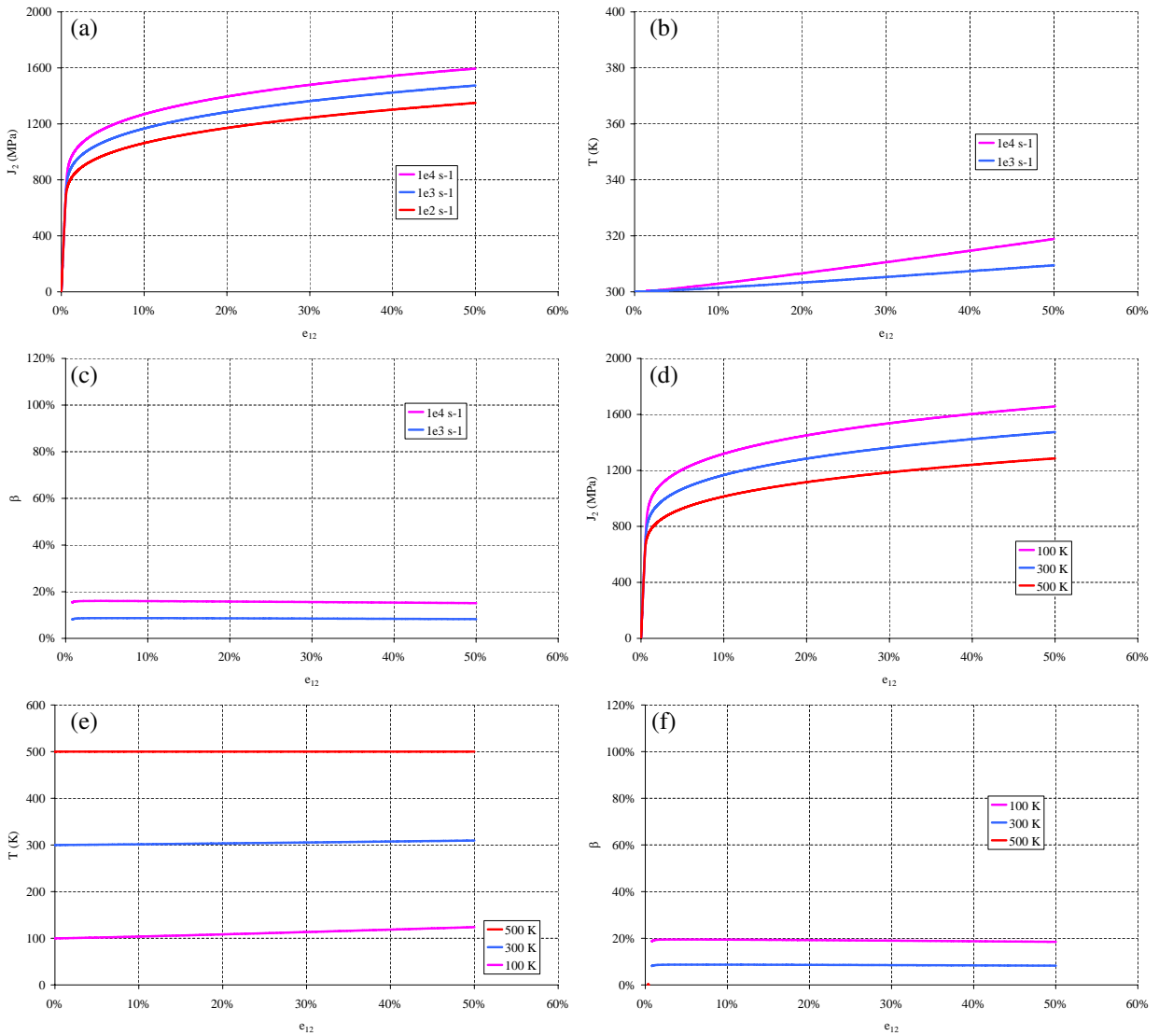


Figure 3. Influence of shear strain rate $\dot{\Gamma}$ and initial temperatures T_0 on stress invariant, heating and inelastic heat fraction for the Johnson–Cook hardening model: (a) stress invariant J_2 versus strain e_{12} , with $T_0 = 300$ K; (b) temperature T versus strain e_{12} , with $T_0 = 300$ K; (c) inelastic heat fraction β versus strain e_{12} , with $T_0 = 300$ K; (d) stress invariant J_2 versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (e) temperature T versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (f) inelastic heat fraction β versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$.

3.2. Strain softening models. Consider now the combination HS-FL-ADD-WYS (see Tables 1-4) summarized in Table 9. This combination may correspond to a very high strength (strongly hardened) bcc metal with a linear approximation of temperature dependence (see Figure 4), such as a very high strength martensitic steel or a pure tungsten.

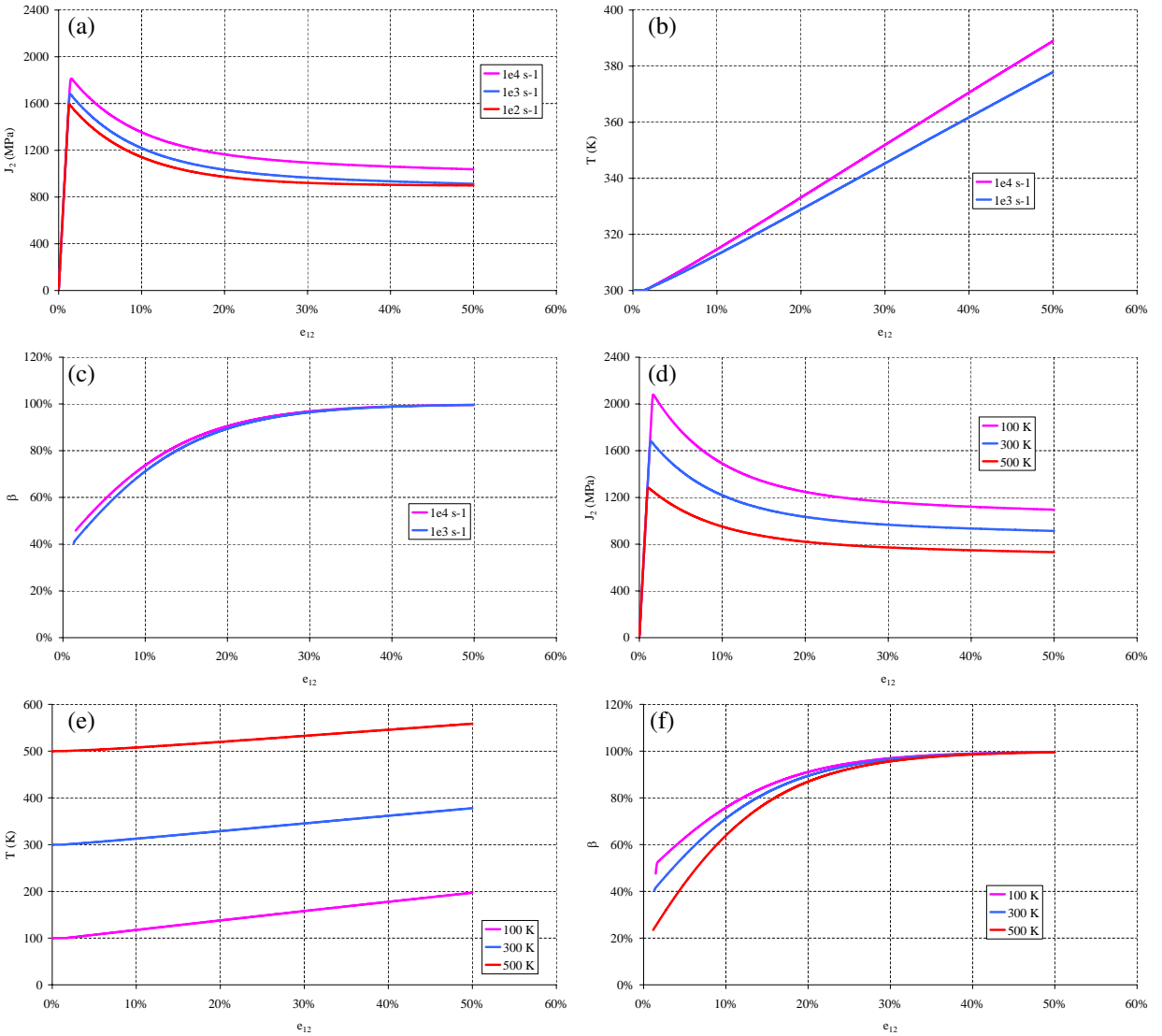


Figure 4. Influence of shear strain rate $\dot{\Gamma}$ and initial temperatures T_0 on stress invariant, heating and inelastic heat fraction for the HS-FL-ADD-WYS softening model: (a) stress invariant J_2 versus strain e_{12} , with $T_0 = 300$ K; (b) temperature T versus strain e_{12} , with $T_0 = 300$ K; (c) inelastic heat fraction β versus strain e_{12} , with $T_0 = 300$ K; (d) stress invariant J_2 versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (e) temperature T versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (f) inelastic heat fraction β versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$.

Figure 4a shows the thermal softening induced under adiabatic conditions for simulations at plastic strain rates higher than 100 s $^{-1}$. According to Figure 4b, the temperature rate is higher for higher strain rate while Figure 4e shows that the temperature rate is lower for higher initial temperature. At 1000 s $^{-1}$ and for $T_0 = 300$ K, the maximal value of temperature increase is close to 80 K.

| | | | | | | |
|---------------------------------|-----|------------------|-----------------------|-----------------------------|-----|--|
| $h(p)$ | | $f(T)$ | $\Phi(T, p, \dot{p})$ | R_0 | | |
| $-\frac{R_\infty}{k} \exp(-kp)$ | | $1 - AT$ | $Y \dot{p}^{1/m}$ | $R_2(1 - AT)$ | | |
| R_∞ (MPa) | k | $A(K^{-1})$ | R_2 (MPa) | Y (MPa s ^{1/m}) | m | |
| 1000 | 20 | 10 ⁻³ | 1000 | 100 | 6 | |

Table 9. Thermoviscoplastic material constants for HS-FL-ADD-WYS model.

According to Figure 4c–f, the initial value of β is lower than 1, increasing with increasing strain rate and with decreasing initial temperature. The rate of β is positive (according to Remark 3) and its absolute value which is high for small strain tends to 0 for large strain. β tends to unity (100%) for large strain.

Consider finally the combination HMSQ-FL-ADD-WYS (see Tables 1-4) summarized in Table 10. This combination may correspond to a polymer type material (for which the thermomechanical framework used in Section 2 may be applied) with a linear dependence on temperature (see Figure 5).

Figure 5a shows the thermal softening induced under adiabatic conditions for simulations at plastic strain rates higher than 100 s⁻¹. Figure 5b shows that the temperature rate is higher for higher strain rate while Figure 5e shows that the temperature rate is lower for higher initial temperature. At 1000 s⁻¹ and for $T_0 = 300$ K, the maximal value of temperature increase is close to 32 K.

According to Figure 5c–f, the initial value of β is lower than 1, increasing with increasing strain rate and with decreasing initial temperature. The rate of β is initially positive, becomes zero for the minimum of the stress-strain curve (see Figure 5a–d) and becomes negative afterwards (these different stages agree with Remark 3). Its absolute value does not depend on strain rate and on initial temperature at the initial stage (growing β) but depends on initial temperature afterwards.

3.3. Summary and complementary analysis. In the previous subsection a series of constitutive models involving strain hardening/softening and thermal softening in the context of rate dependent plasticity has been analysed in terms of evolution of stress invariant, temperature and inelastic heat fraction determined from the heat equation under adiabatic constraint.

The first comment concerns the tendency deduced from Remark 2: for a strain hardening model, the inelastic heat fraction decreases with increasing plastic deformation. This tendency has been notably observed experimentally by Lerch et al. [2003] and Jovic et al. [2006] on aluminium alloy and stainless

| | | | | | | | |
|--|-------|-------|------------------------|-----------------------|-----------------------------|-----|--|
| $h(p)$ | | | $f(T)$ | $\Phi(T, p, \dot{p})$ | R_0 | | |
| $-\frac{R_\infty}{k_1} \exp(-k_1 p) + k_2 \frac{p^2}{2}$ | | | $1 - AT$ | $Y \dot{p}^{1/m}$ | $R_2(1 - AT)$ | | |
| R_∞ (MPa) | k_1 | k_2 | A (K ⁻¹) | R_2 (MPa) | Y (MPa s ^{1/m}) | m | |
| 800 | 60 | 5000 | 10 ⁻³ | 500 | 100 | 6 | |

Table 10. Thermoviscoplastic material constants for HMSQ-FL-ADD-WYS model.

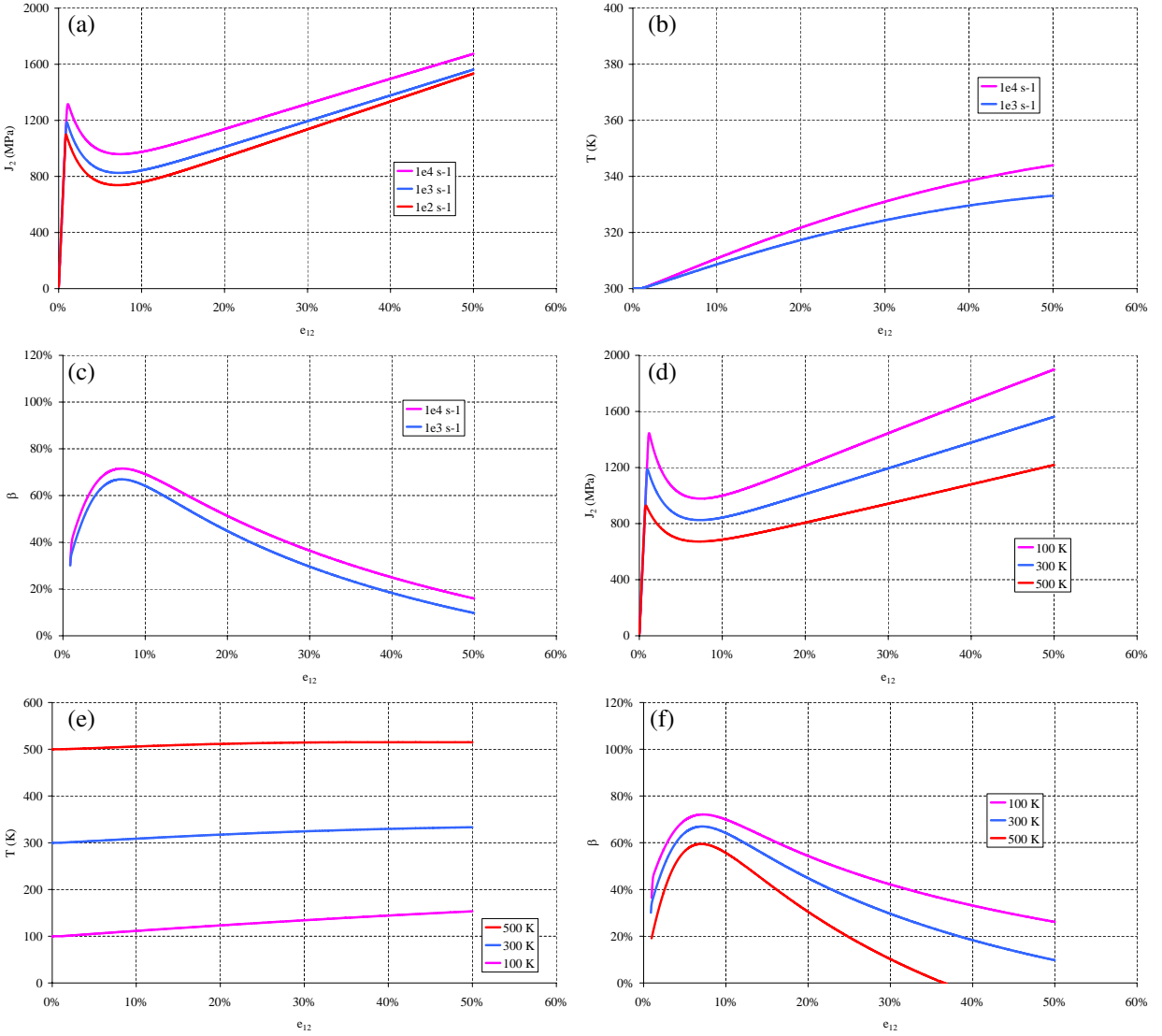


Figure 5. Influence of shear strain rate $\dot{\Gamma}$ and initial temperatures T_0 on stress invariant, heating and inelastic heat fraction for the HMSQ-FL-ADD-WYS softening model. (a) stress invariant J_2 versus strain e_{12} , with $T_0 = 300$ K; (b) temperature T versus strain e_{12} , with $T_0 = 300$ K; (c) inelastic heat fraction β versus strain e_{12} , with $T_0 = 300$ K; (d) stress invariant J_2 versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (e) temperature T versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$; (f) inelastic heat fraction β versus strain e_{12} , with $\dot{\Gamma} = 10^3$ s $^{-1}$.

steel respectively. The inelastic heat fraction evolution they observe is qualitatively identical to what is obtained numerically herein within the HV-FL-ADD-WYS model (see [Figure 1c](#)). Oppositely, according to [Remark 2](#), for a strain softening model the inelastic heat fraction increases and converges to unity. This result (see [Figure 4c](#) for HS-FL-ADD-WYS model) can be compared with experimental ones obtained on tungsten by [Subhash et al. \[1994\]](#). A model combining both strain softening and strain hardening (HMSQ)

allows for an increasing then decreasing inelastic heat fraction. This behaviour is typical of polymeric materials, and the evolution of the inelastic heat fraction obtained numerically for HMSQ-FL-ADD-WYS model (see Figure 5c) reproduces (qualitatively) very well experimental results for polycarbonate [Lerch et al. 2003] and for glassy polymer [Rittel 1999].

The second comment concerns the particular case studied in Remark 5: for a multiplicative rate dependent model (MUL and EXP) combined with the absence of initial yield stress (NYS), the inelastic heat fraction is quasiconstant and its value is very low (see Figure 2b and Figure 3b), leading to weak (and negligible) temperature increase (see Figure 2c and Figure 3c).

On the other hand, for strain hardening models with very similar behaviour, including HV-FL-ADD-WYS, HMLRO-FE-MUL-NYS and Johnson–Cook ones, the maximum value of temperature increase is observed for the former while the minimum value of temperature increase is observed for the latter. Indeed, the corresponding conditions regarding thermomechanical behaviour let in the plastic dissipation induced thermal softening in the first model. On the contrary, this can be hardly reproduced for the Johnson–Cook model. A complete thermodynamics-based approach cannot be applied in this case. This agrees with the experiment-based remark of Chrysochoos et al. [1989] concerning constitutive modelling: some engineering models are not able to reproduce observed phenomena. In order to palliate this deficiency, simplifications are usually done to evaluate temperature growth under adiabatic plasticity conditions. As mentioned previously, they consist (often implicitly), first, in neglecting thermodissipative couplings (see Remark 2) and, second, in considering the inelastic heat fraction as a constant (see Equation (50)). The consequences of these simplifications are now studied.

The heat equation (19) can be decomposed into various contributions. By designating $\dot{Q} = \rho_0 c_0 \dot{T}$, $D = \boldsymbol{\tau} : \mathbf{d}^P - R\dot{p} \geq 0$, $\dot{W}^\tau = T \frac{\partial \boldsymbol{\tau}}{\partial T} : \mathbf{d}^e$ and $\dot{W}^R = T \frac{\partial R}{\partial T} \dot{p}$, Equation (19) can be rewritten as

$$\dot{Q} = \dot{W}, \quad \text{with } \dot{W} = D + \dot{W}^\tau + \dot{W}^R. \quad (33)$$

Three cases are now distinguished depending on the terms retained on the right side of (33)_{1,2}.

In the first case, called *quasicomplete heat evaluation*, thermoelastic coupling effects are neglected due to their weak contribution to temperature change: $\dot{W}^\tau = 0$. Relation (33)₂ is thus reduced to

$$\dot{W} = D + \dot{W}^R = \left[J_2 - \left(R - T \frac{\partial R}{\partial T} \right) \right] \dot{p}. \quad (34)$$

In the second case, the heat equation (20) is further simplified by neglecting both thermoelastic and thermoplastic couplings contributions (see [Voyiadjis and Abed 2006] for a similar simplification). The resulting equation is here called a *simplified heat evaluation*. This means that $\dot{W}^\tau = \dot{W}^R = 0$ and (33)₂ becomes

$$\dot{W} = D = [J_2 - R] \dot{p}. \quad (35)$$

The last case is equivalent to the representation involving a constant inelastic heat fraction β — also known as the Taylor–Quinney coefficient [Mason et al. 1994] — as is frequently done in engineering problems. The resulting equation is here called a *basic heat evaluation*. The expression of work rate \dot{W} in (33)₂ in this case reduces to

$$\dot{W} = \beta J_2 \dot{p}. \quad (36)$$

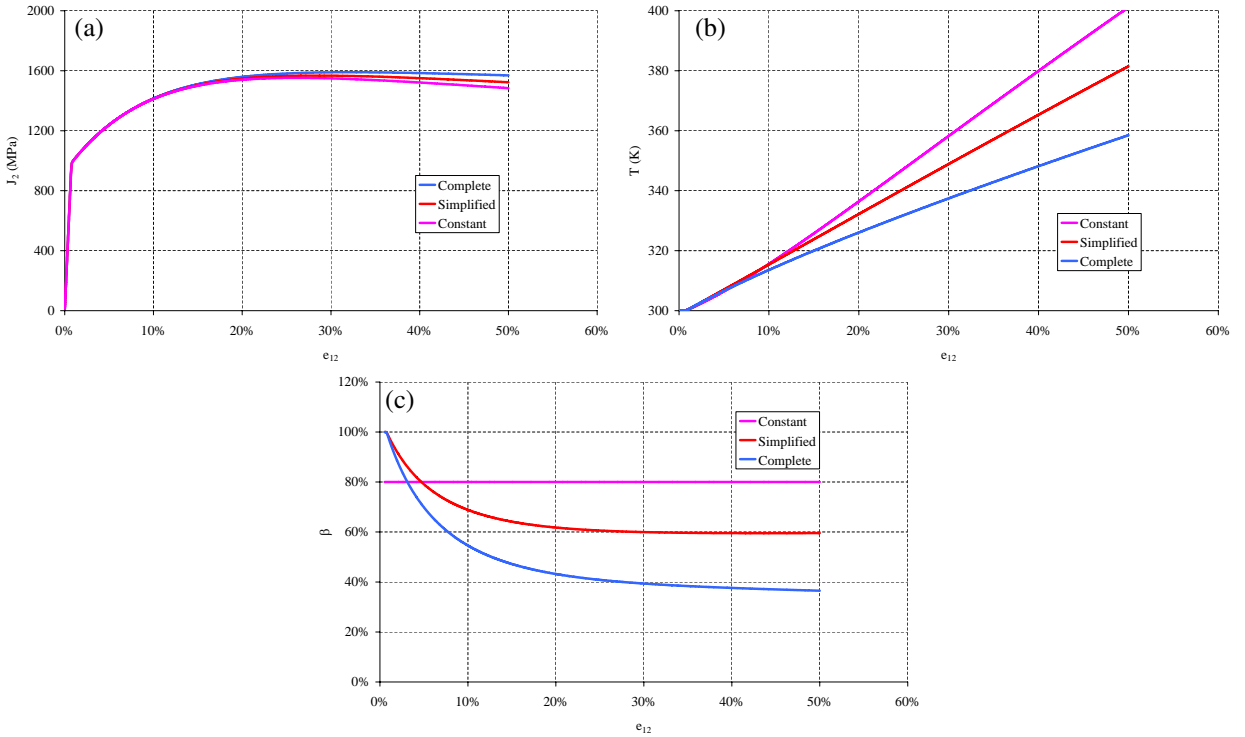


Figure 6. Influence of the simplifications made in the heat equation on stress invariant, heating and inelastic heat fraction considering HV-FL-ADD-WYS hardening model (see Table 6 and Figure 1 for comparison). (a) Stress invariant J_2 versus strain e_{12} ; (b) temperature T versus strain e_{12} ; (c) inelastic heat fraction β versus strain e_{12} . In each case, $T_0 = 300$ K and $\dot{\Gamma} = 10^3$ s $^{-1}$.

Figures 6–10 show the evolution of the stress invariant, temperature and inelastic heat fraction for a *complete* description, for a *simplified* one and finally, assuming a constant value of the inelastic heat fraction, for the combinations used previously, namely the HV-FL-ADD-WYS type model (Figure 6), the HMLRO-FE-MUL-NYS-type model (Figure 7), the Johnson–Cook model (Figure 8), the HS-FL-ADD-WYS-type model (Figure 9) and the HMSQ-FL-ADD-WYS-type model (Figure 10).

Considering strain hardening models (see Figures 6, 7 and 8), the use of a constant value for β (part c) leads to higher values for temperature (part b) causing a more significant thermal softening (part a). This remark also holds concerning strain hardening/softening model in Figure 10. On the other hand the different methods give very similar results in terms of temperature rates and material behaviour for strain softening model.

4. Effects of heat evaluation method on dynamic plastic localization occurrence

This section aims at showing the influence of the choice of method for evaluating the plastic work-induced heating on the determination of the conditions for dynamic plastic localization onset. The localization phenomenon at stake is adiabatic shear banding (ASB) which constitutes a precursor of

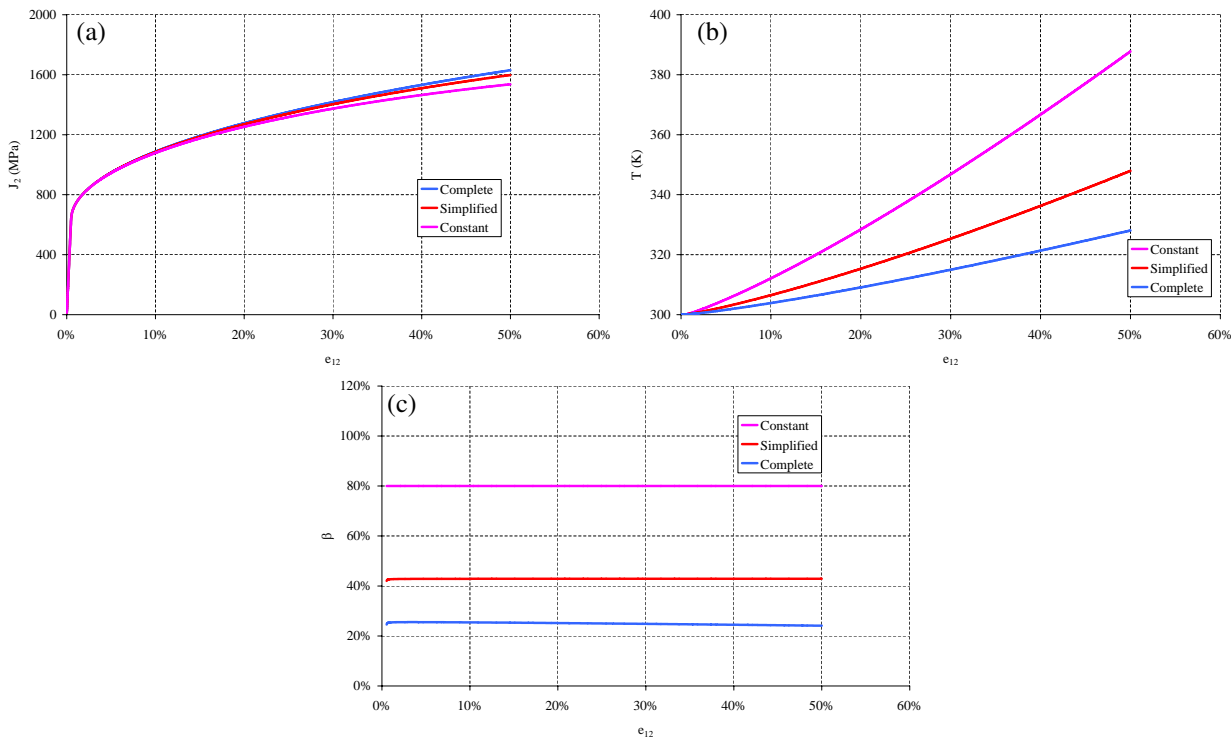


Figure 7. Influence of the simplifications made in the heat equation on stress invariant, heating and inelastic heat fraction considering HMLRO-FE-MUL-NYS hardening model (see Table 7 and Figure 2 for comparison). (a) Stress invariant J_2 versus strain e_{12} ; (b) temperature T versus strain e_{12} ; (c) inelastic heat fraction β versus strain e_{12} . In each case, $T_0 = 300$ K and $\dot{\Gamma} = 10^3$ s $^{-1}$.

failure of structures submitted to dynamic loading and which intervenes as thermal softening overcomes strain hardening [Marchand and Duffy 1988; Bai and Bodd 1992; Longère et al. 2003].

Conditions for ASB occurrence are commonly obtained from the linear perturbation method which is in general applied in the case of simple shear under constant velocity boundary conditions. Assuming negligible elastic effects, laminar viscoplastic flow and adiabatic conditions, the problem can be reduced to a one-dimensional formulation [Bai 1982; Clifton et al. 1984; Molinari 1985; Batra and Wei 2006] (see e.g. [Anand et al. 1987]) when three-dimensional generalization is presented). Admitting analytical solutions, the linear perturbation method provides in this case a criterion of instability onset, which is interpreted as the incipience of the adiabatic shear banding process, providing in fact the necessary condition for the onset of formation of bands (possibility of a shear band type instability). In the case of ASB phenomena, the extension of the one-dimensional loading criterion to a complex three-dimensional loading can be performed today [Longère et al. 2003]. Simplifications are assumed in order to reproduce qualitatively the different stages of deformation localization (weak and strong) as observed by Marchand and Duffy [1988].

Starting from the governing equations (momentum balance, energy balance, kinematics and constitutive law) in the case of simple shear in a (1, 2)-plane under adiabatic conditions and neglecting elasticity,

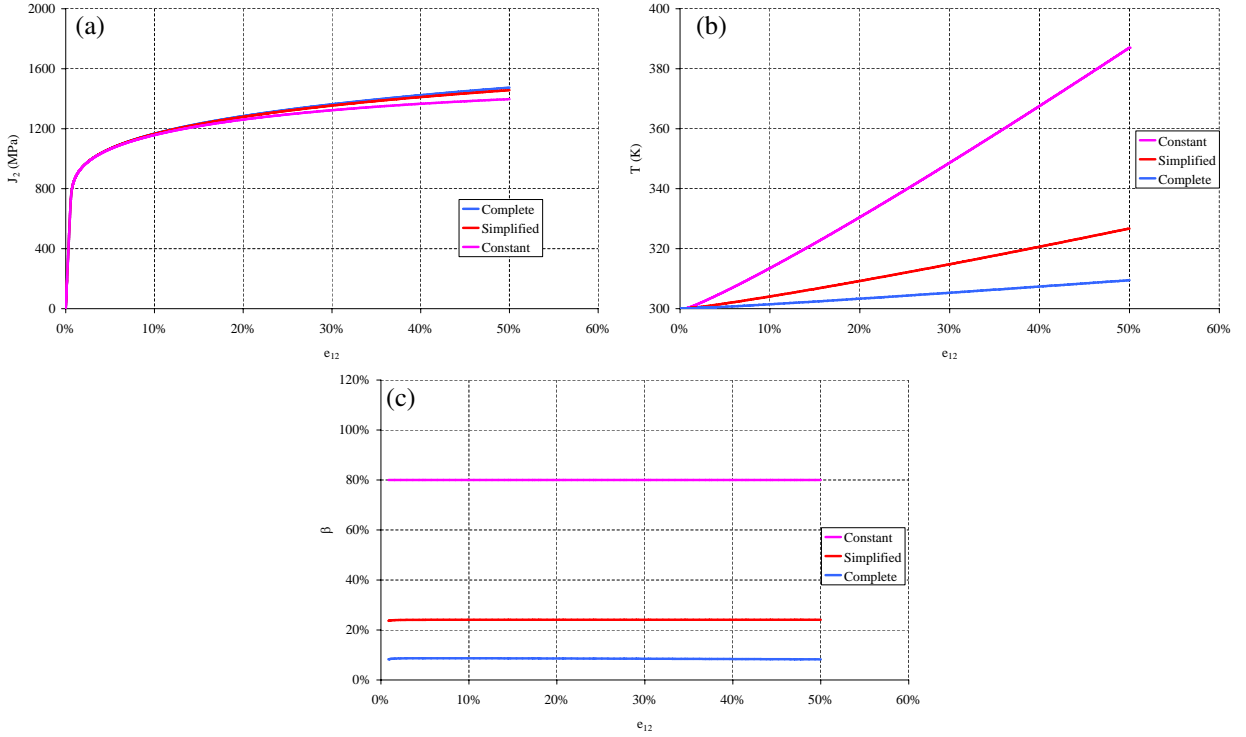


Figure 8. Influence of the simplifications made in the heat equation on stress invariant, heating and inelastic heat fraction considering Johnson–Cook hardening model (see Table 8 and Figure 3 for comparison). (a) Stress invariant J_2 versus strain e_{12} ; (b) temperature T versus strain e_{12} ; (c) inelastic heat fraction β versus strain e_{12} . In each case, $T_0 = 300$ K and $\dot{\Gamma} = 10^3$ s $^{-1}$.

one obtains the following system:

$$\begin{cases} s_{12,2} = \rho_0 \dot{v}_1, \\ \rho_0 c_0 \dot{T} = \dot{W}, \\ d_{12} = \frac{1}{2} (v_{1,2} + v_{2,1}) \approx d_{12}^p = \frac{3}{2} \dot{p} \frac{s_{12}}{J_2}, \\ \dot{p} = \Lambda (s_{12}, p, T). \end{cases} \quad (37)$$

With $J_2 = \sqrt{3}s_{12}$, $v_1 = \dot{\Gamma}x_2$, $v_2 = 0$ and $\Gamma = \dot{\Gamma}t$, the system (37) is reduced to

$$\begin{cases} s_{12,2} - \rho_0 \dot{v}_1 = 0, \\ \rho_0 c_0 \dot{T} - \dot{W} = 0, \\ v_{1,2} - \sqrt{3} \dot{p} = 0, \\ \dot{p} - \Lambda (s_{12}, p, T) = 0. \end{cases} \quad (38)$$

A small perturbation $\delta U = (\delta v_1, \delta s_{12}, \delta p, \delta T)$ is now superimposed on the set of homogeneous solutions $U = (v_1, s_{12}, p, T)$: $U \Rightarrow U + \delta U$ with $\delta U \ll U$. Let the perturbation have a wave-like form:

$$\delta U = \bar{U} \exp(\varpi t + ikx_2) = \bar{U} \exp[\varpi_R t] \exp[ik(ct + x_2)], \quad (39)$$

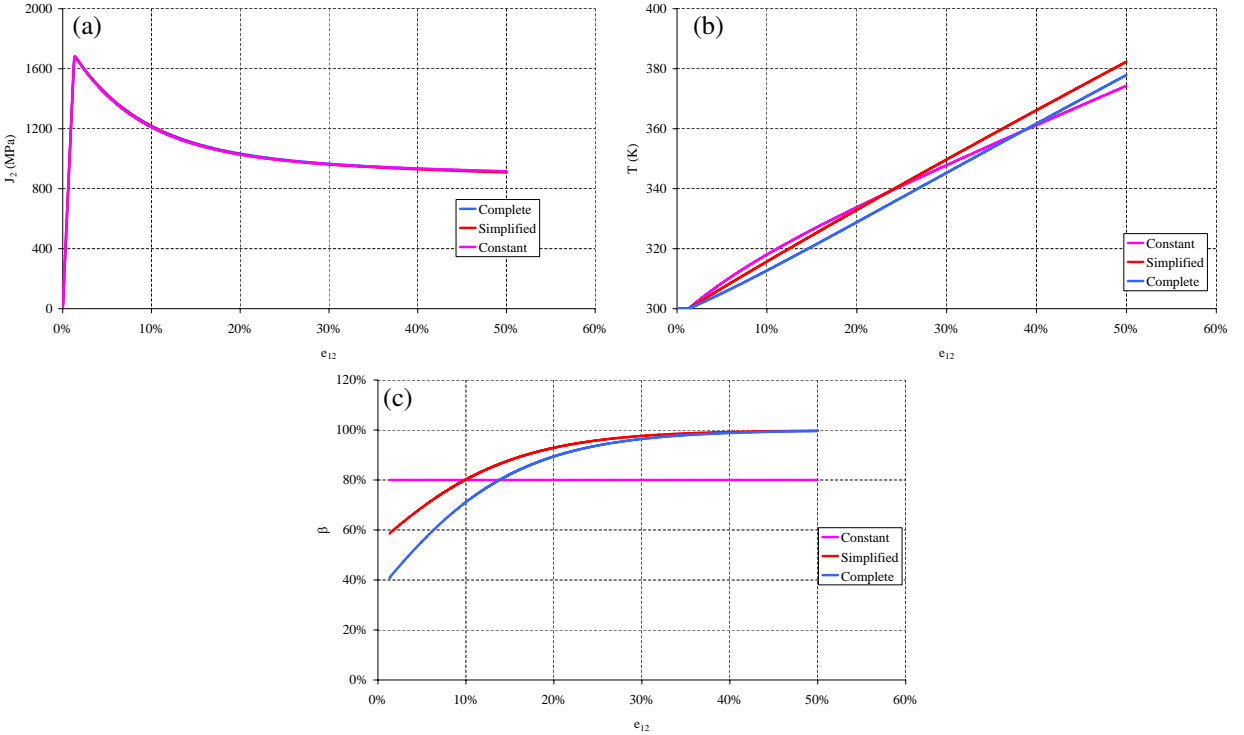


Figure 9. Influence of the simplifications made in the heat equation on stress invariant, heating and inelastic heat fraction considering HS-FL-ADD-WYS hardening model (see Table 9 and Figure 4 for comparison). (a) Stress invariant J_2 versus strain e_{12} ; (b) temperature T versus strain e_{12} ; (c) inelastic heat fraction β versus strain e_{12} . In each case, $T_0 = 300$ K and $\dot{\Gamma} = 10^3$ s $^{-1}$.

where \bar{U} is the perturbation magnitude, ϖ the wave pulsation, k the wave number, x_2 the wave plane normal, ϖ_R and ϖ_I the real and imaginary parts of the wave pulsation ϖ , and $c = \varpi_I/k$ the wave velocity.

According to the right side of (39), the case $\varpi_R = 0$ points the transition between the stable and unstable states:

- (i) if $\varpi_R > 0$, the perturbation may grow with time and the instability mentioned is possible;
- (ii) if $\varpi_R < 0$, the perturbation decreases with time.

The objective consists thus in looking for the conditions of the transition from the stable state to the possible unstable state by studying the sign of ϖ_R .

After linearization and using the notations in Table 11 (see next page), the system (38) may be rewritten as follows (see [Longère and Dragon 2007] for further details):

$$\begin{bmatrix} -\rho_0 \varpi & ik & 0 & 0 \\ 0 & -w_s & -(w_p + w_{\dot{p}} \varpi) & (\rho_0 c_0 \varpi - w_T) \\ ik & 0 & -\sqrt{3} \varpi & 0 \\ 0 & -2P_{12} & (\varpi - B) & -E \end{bmatrix} \begin{bmatrix} \bar{v}_1 \\ \bar{s}_{12} \\ \bar{p} \\ \bar{T} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (40)$$

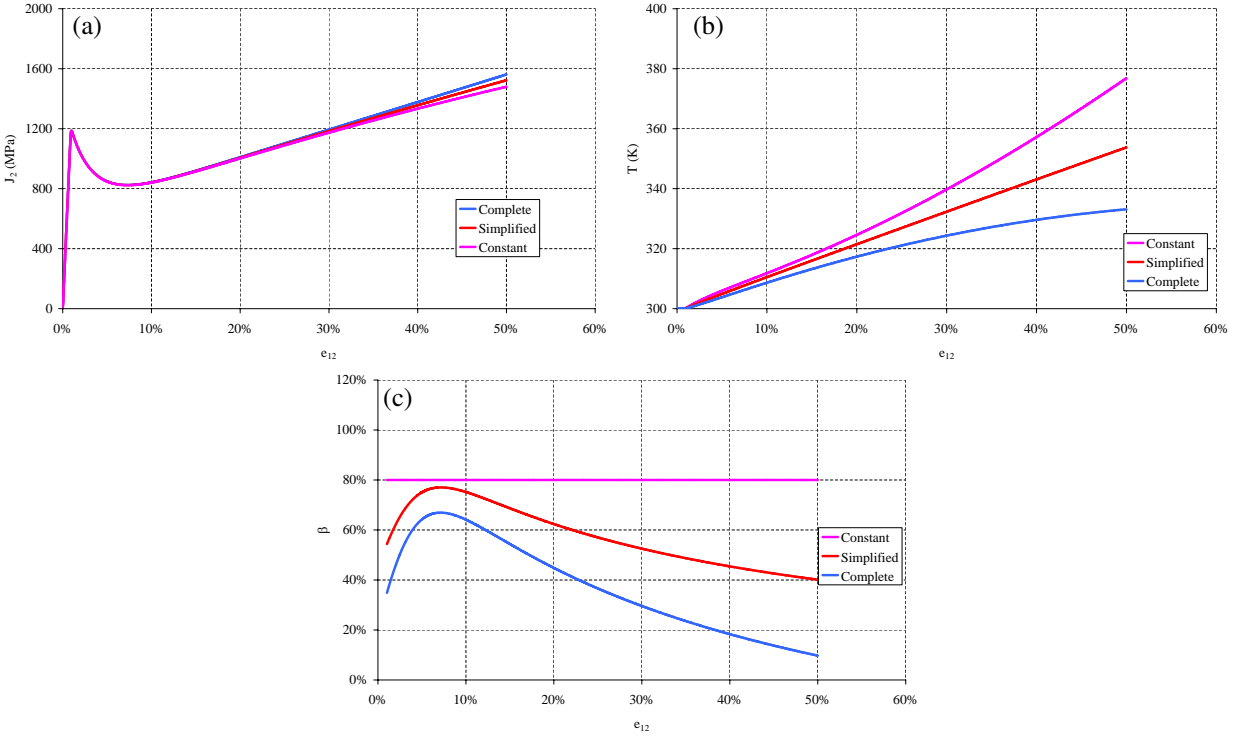


Figure 10. Influence of the simplifications made in the heat equation on stress invariant, heating and inelastic heat fraction considering HMSQ-FL-ADD-WYS hardening model (see Table 10 and Figure 5 for comparison). (a) Stress invariant J_2 versus strain e_{12} ; (b) temperature T versus strain e_{12} ; (c) inelastic heat fraction β versus strain e_{12} . In each case, $T_0 = 300$ K and $\dot{\Gamma} = 10^3$ s $^{-1}$.

or otherwise:

$$[A] \{\bar{U}\} = \{0\}. \quad (41)$$

The determinant of the matrix $[A]$ in (41), whose components are denoted a_{ij} , is simply

$$\det [A] = a_{11}a_{22}a_{33}a_{44} - a_{11}a_{42}a_{33}a_{24} + a_{12}a_{31} (a_{23}a_{44} - a_{43}a_{24}). \quad (42)$$

| | Basic | Simplified | Quasicomplete |
|------------|------------------------|----------------------|-----------------------------|
| ws | $\sqrt{3}\beta\dot{p}$ | $\sqrt{3}\dot{p}$ | $\sqrt{3}\dot{p}$ |
| wp | 0 | $-Q\dot{p}$ | $-(Q - TY)\dot{p}$ |
| wT | 0 | $-S\dot{p}$ | $TZ\dot{p}$ |
| $w\dot{p}$ | $\sqrt{3}\beta s_{12}$ | $\sqrt{3}s_{12} - R$ | $\sqrt{3}s_{12} - (R - TS)$ |

Table 11. Expressions of w_i .

The spectral equation deduced from (41) and (42) is a third-degree polynomial in ϖ :

$$\det [A] = a_3 \varpi^3 + a_2 \varpi^2 + a_1 \varpi + a_0, \quad (43)$$

with

$$\begin{cases} a_3 = \sqrt{3} \rho_0^2 c_0 2P_{12}, \\ a_2 = \rho_0 [k^2 c_0 + \sqrt{3} (E w_s - 2P_{12} w_T)], \\ a_1 = -k^2 (E w_{\dot{p}} + w_T + \rho_0 c_0 B), \\ a_0 = k^2 (B w_T - E w_p), \end{cases} \quad (44)$$

where

$$P_{ij} = \frac{3}{2} \bar{\alpha} \frac{s_{ij}}{J_2}, \quad B = -\bar{\alpha} Q, \quad E = -\bar{\alpha} [R_0 f'(T) + S], \quad (45)$$

$$Q = \frac{\partial R}{\partial p} = h''(p) f(T), \quad S = \frac{\partial R}{\partial T} = h'(p) f'(T), \quad \bar{\alpha} = \frac{\partial \Lambda}{\partial F}. \quad (46)$$

As observed by Marchand and Duffy [1988] and demonstrated by Molinari [1985], instability does not imply localization rigorously. Thus the use of the linear perturbation method provides a necessary condition only, leading to a possible *lower* bound for the effective localization incipience. The idea is here to *delay* the instability onset, that is, push it towards an *upper* bound in the sense of approaching the strong localization incipience. In other words we distinguish the *instability point* characterizing locally the equilibrium between strain hardening and thermal softening (maximum of the shear stress-shear strain curve in most cases) and the *localization point* beyond which shear stress drops strongly. The linear perturbation method provides the instability point and we are looking, via a pragmatic engineering evaluation approach, for the localization point which eventually succeeds the instability point.

Adiabatic shear banding occurs as thermal softening overcomes strain hardening. Before the instability point, strain hardening is predominant and the material is necessarily stable, while past the instability point, thermal softening becomes predominant and the material may become unstable. In the linear perturbation method, attenuating thermal softening allows consequently for pushing forward the instability point and approaching the localization point (see [Longère et al. 2003; 2005; Longère and Dragon 2007, for practical applications]). Detection of plastic localization onset is essential to control a process of softening behaviour for three-dimensional modelling of thermoelastic/viscoplastic materials incorporating adiabatic shear banding formation and growth.

It is notable that the approximate instability criteria in the sense given above can be applied for any constitutive equations based on irreversible thermodynamics.

As mentioned previously, delaying the strong localization onset with respect to the supposed instability onset is favoured by making $f'(T)$ vanish in the linear perturbation method (concept of 'upper' bound approximation). This approach is consequently used in the following.

Employing Table 11, when considering the basic evaluation, the coefficients a_0 – a_3 in (44) become

$$\begin{cases} a_3 = \sqrt{3} \rho_0^2 c_0 2P_{12}, \\ a_2 = \rho_0 [k^2 c_0 + 3E\beta \dot{p}], \\ a_1 = -k^2 (E \sqrt{3} \beta s_{12} + \rho_0 c_0 B), \\ a_0 = 0. \end{cases} \quad (47)$$

The spectral equation is, in this case, indeed reduced to a second-degree polynomial in ϖ and $\varpi > 0$ if and only if $a_1 a_3 < 0$. Accordingly, the condition for possible perturbation growth (instability occurrence) is

$$\sqrt{3}s_{12} > \frac{\rho_0 c_0}{\beta} \frac{(\partial R / \partial p)}{-(\partial R / \partial T)}. \quad (48)$$

According to [Table 11](#), the coefficients a_0 to a_3 in (44) in the simplified evaluation become

$$\begin{cases} a_3 = \sqrt{3}\rho_0^2 c_0 2P_{12}, \\ a_2 = \rho_0[k^2 c_0 + 3\dot{p}(E + \frac{2}{\sqrt{3}}P_{12}S)], \\ a_1 = -k^2[E(\sqrt{3}s_{12} - R) - S\dot{p} + \rho_0 c_0 B], \\ a_0 = k^2\dot{p}(EQ - BS), \end{cases} \quad (49)$$

with $E + \frac{2}{\sqrt{3}}P_{12}S = 0$ and $EQ - BS = 0$.

The corresponding components in (49) are thus expressed by

$$\begin{cases} a_3 = \sqrt{3}\rho_0^2 c_0 2P_{12}, \\ a_2 = \rho_0 k^2 c_0, \\ a_1 = -k^2[E(\sqrt{3}s_{12} - R) - S\dot{p} + \rho_0 c_0 B], \\ a_0 = 0. \end{cases} \quad (50)$$

The spectral equation is once more reduced to a second-degree polynomial in ϖ and the condition for perturbation growth is given by

$$\sqrt{3}s_{12} > R - \frac{\dot{p}}{\bar{\alpha}} + \rho_0 c_0 \frac{(\partial R / \partial p)}{-(\partial R / \partial T)}. \quad (51)$$

Following the previous procedure for quasicomplete evaluation gives

$$\begin{cases} a_3 = \sqrt{3}\rho_0^2 c_0 2P_{12}, \\ a_2 = \rho_0[k^2 c_0 + 3\dot{p}(E - \frac{2}{\sqrt{3}}P_{12}TZ)], \\ a_1 = -k^2(E[\sqrt{3}s_{12} - (R - TS)] + TZ\dot{p} + \rho_0 c_0 B), \\ a_0 = k^2\dot{p}[EQ - T(EY - BZ)]. \end{cases} \quad (52)$$

The spectral equation remains a third-degree polynomial in ϖ for which the condition of instability onset is not trivial. In order to obtain a practical criterion in the sense of [\[Bai 1982\]](#), the condition for instability onset is again deduced from the sign of the product $a_1 a_3$.

The condition for perturbation growth in this case is given by

$$\sqrt{3}s_{12} > R - T \left(\frac{\partial R}{\partial T} - \frac{(\partial^2 R / \partial T^2) \dot{p}}{(\partial R / \partial T) \bar{\alpha}} \right) + \rho_0 c_0 \frac{(\partial R / \partial p)}{-(\partial R / \partial T)}.$$

The conditions for dynamic plastic localization occurrence for the basic heat evaluation method, the simplified heat evaluation method and the quasicomplete heat evaluation method, as defined above, are

$$\sqrt{3}s_{12} > \frac{\rho_0 c_0}{\beta} \frac{(\partial R / \partial p)}{-(\partial R / \partial T)}, \quad \sqrt{3}s_{12} > R - \frac{\dot{p}}{\bar{\alpha}} + \rho_0 c_0 \frac{(\partial R / \partial p)}{-(\partial R / \partial T)},$$

$$\sqrt{3}s_{12} > R - T \left(\frac{\partial R}{\partial T} - \frac{(\partial^2 R / \partial T^2) \dot{p}}{(\partial R / \partial T) \bar{a}} \right) + \rho_0 c_0 \frac{(\partial R / \partial p)}{-(\partial R / \partial T)}.$$

(see [Longère and Dragon 2007]). The description of the postlocalization behaviour is not the purpose of the present paper. But in order to visualize the drop in stress induced by adiabatic shear banding, the ASB deterioration model developed by Longère et al. [2003] has been used. The details regarding numerical calculations and corresponding algorithmic procedure can be found in [Longère et al. 2005].

Figures 11 and 12 give, respectively, the material response to shearing and the temperature evolution for the three cases of heat evaluation (two values for β are given). According to Figure 11, the value of shear strain at localization (critical shear strain) onset is close to 30% for basic evaluation with $\beta = 1.0$, close to 32% for basic evaluation with $\beta = 0.8$, close to 35% for simplified evaluation, and close to 38% for complete evaluation. In parallel, the influence of heat evaluation is shown in Figure 12, which gives the evolution of temperature. According to the basic evaluation temperature increases until a value close to 380 K while according to the quasicomplete evaluation, the maximum of temperature remains under 350 K. This difference of 30 K is enough to provoke a difference of 8% for the estimation of the critical shear strain in the conditions prescribed here.

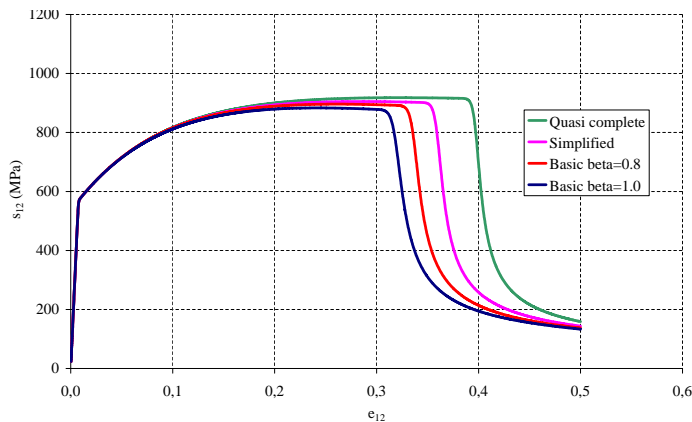


Figure 11. Shear stress versus shear strain: $T_0 = 300 \text{ K}$; $\dot{\Gamma} = 10^3 \text{ s}^{-1}$.

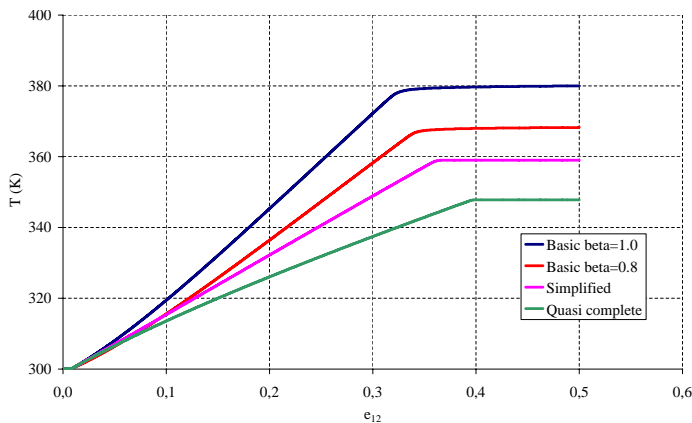


Figure 12. Temperature versus shear strain: $T_0 = 300 \text{ K}$; $\dot{\Gamma} = 10^3 \text{ s}^{-1}$.

| Evaluation type | β | R_0 (MPa) | A (K^{-1}) |
|-----------------|---------|-------------|-------------------------|
| Basic | 1.0 | 1000 | $1. \cdot 10^{-3}$ |
| | 0.8 | 1015 | $1.05 \cdot 10^{-3}$ |
| Simplified | X | 1050 | $1.1 \cdot 10^{-3}$ |
| Quasicomplete | X | 1080 | $1.15 \cdot 10^{-3}$ |

Table 12. Material constants.

In this example, the shear stress-shear strain curves begin to diverge from each other when shear strain is close to 10% (see Figure 11). As a consequence, the value of shear stress at the maximum of each curve is different (the highest is obtained for the quasicomplete evaluation while the lowest is obtained for the basic evaluation for $\beta = 1$). So, we are considering here next shear stress vs. shear strain curves with the same maximum value of shear stress (see Table 12 for the new sets of material constants). Numerical results including instability criterion are given in Figures 13 and 14. In these figures, the values of critical

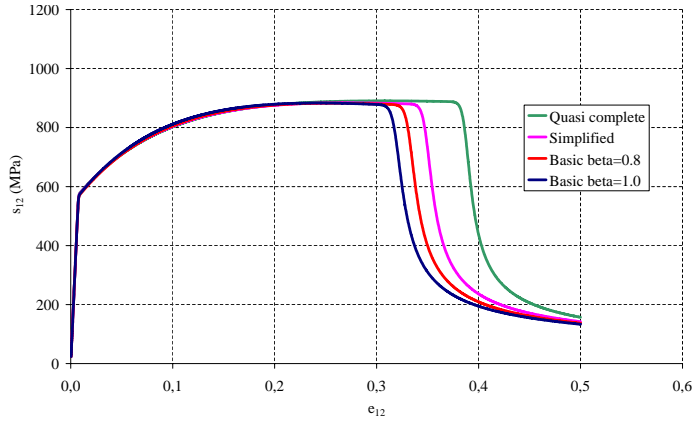


Figure 13. Shear stress versus shear strain: $T_0 = 300 \text{ K}$; $\dot{\Gamma} = 10^3 \text{ s}^{-1}$.

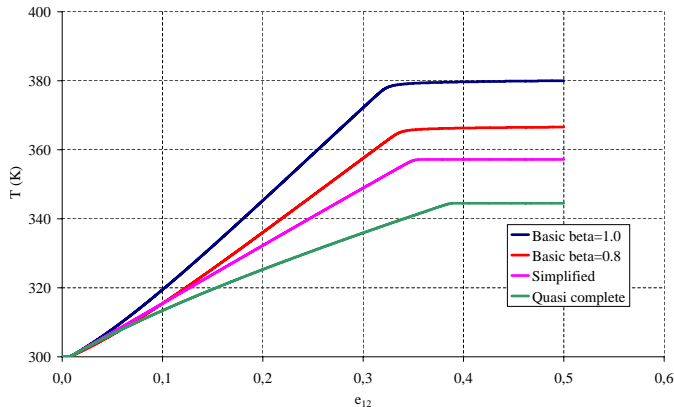


Figure 14. Temperature versus shear strain: $T_0 = 300 \text{ K}$; $\dot{\Gamma} = 10^3 \text{ s}^{-1}$.

shear strain are lower than the previous ones (see [Figure 11](#)) but the tendency is preserved: critical shear strain is greater for the quasicomplete evaluation.

5. Concluding remarks

Following an internal variable approach including strain, strain rate and temperature dependence in the context of viscoplastic behaviour embracing a wide range of features involved for a large class of engineering materials, a three-dimensional rather general expression of the evolving inelastic heat fraction has been obtained. This relation allows us to conclude that for a strain hardening model, the inelastic heat fraction decreases for increasing strain, and for a strain softening model the inelastic heat fraction increases and converges to unity. These theoretical results are in agreement with experimental observations. However, depending on the formulation of the constitutive equations, these effects can be well reproduced or totally erased. The choice of the constitutive laws in terms of strain hardening/softening, thermal softening and strain rate dependence is thus crucial. Consequently some models are intrinsically able to reproduce observed phenomena, notably the temperature rise induced by plastic deformation under adiabatic conditions, while others are not. The latter fact sometimes invites selection of an arbitrary, albeit conservative, empirical value for the (constant) inelastic heat fraction. But the use of a constant value for the inelastic heat fraction constitutes a coarse and mostly conservative simplification which can sometimes lead to erroneous results.

The quantitative investigation presented, which employs a spectrum of existing thermoviscoplastic modelling variants at finite strain, aims to show some history effects on the evolution of the inelastic heat fraction β in dynamic plasticity. It points out some limitations regarding commonly employed hypotheses and simplifications concerning the variation of β in engineering applications. It indicates the need for more involved experimental research and for enhanced description of thermomechanical couplings. The multiscale aspects were deliberately set apart in this text, which, in some aspects, provides a critical review of the field while remaining subject to the limits of phenomenological modelling. However, the thermodynamical and physical status of absolute temperature with respect to microscopic heterogeneity of, say, polycrystalline metals and other engineering materials, represents in itself a further challenge and needs more clarification.

References

- [Anand et al. 1987] L. Anand, K. H. Kim, and T. G. Shawki, “Onset of shear localization in viscoplastic solids”, *J. Mech. Phys. Solids* **35**:4 (1987), 407–429.
- [Aravas et al. 1990] N. Aravas, K.-S. Kim, and F. A. Leckie, “On the calculations of the stored energy of cold work”, *J. Eng. Mater. Technol. (ASME)* **112**:4 (1990), 465–470.
- [Bai 1982] Y. L. Bai, “Thermo-plastic instability in simple shear”, *J. Mech. Phys. Solids* **30**:4 (1982), 195–207.
- [Bai and Bodd 1992] Y. L. Bai and B. Bodd, *Adiabatic shear localization: occurrence, theories, and applications*, Pergamon Press, Oxford, 1992.
- [Bataille and Kestin 1975] J. Bataille and J. Kestin, “L’interprétation physique de la thermodynamique rationnelle”, *J. Mécanique* **14**:2 (1975), 365–384.
- [Batra and Chen 2001] R. C. Batra and L. Chen, “Effect of viscoplastic relations on the instability strain, shear band initiation strain, the strain corresponding to the minimum shear band spacing, and the band width in a thermoviscoplastic material”, *Int. J. Plast.* **17**:11 (2001), 1465–1489.

- [Batra and Wei 2006] R. C. Batra and Z. G. Wei, “Shear band spacing in thermoviscoplastic materials”, *Int. J. Impact Eng.* **32**:6 (2006), 947–967.
- [Bever et al. 1973] M. B. Bever, D. L. Holt, and A. L. Titchener, “The stored energy of cold work”, *Prog. Mater. Sci.* **17** (1973), 5–177.
- [Campagne et al. 2005] L. Campagne, L. Daridon, and S. Ahzi, “A physically based model for dynamic failure in ductile metals”, *Mech. Mater.* **37**:8 (2005), 869–886.
- [Chrysochoos et al. 1989] A. Chrysochoos, O. Maisonneuve, G. Martin, H. Caumon, and J. C. Chezeaux, “Plastic and dissipated work and stored energy”, *Nucl. Eng. Des.* **114**:3 (1989), 323–333.
- [Clayton 2005] J. D. Clayton, “Dynamic plasticity and fracture in high density polycrystals: constitutive modelling and numerical simulation”, *J. Mech. Phys. Solids* **53**:2 (2005), 261–301.
- [Clifton et al. 1984] R. J. Clifton, J. Duffy, K. A. Hartley, and T. G. Shawki, “On critical conditions for shear band formation at high strain rates”, *Scr. Metall.* **18**:5 (1984), 443–448.
- [Fressengeas and Molinari 1985] C. Fressengeas and A. Molinari, “Inertia and thermal effects on the localization of plastic flow”, *Acta Metall. Mater.* **33**:3 (1985), 387–396.
- [Guo et al. 2005] Y. B. Guo, Q. Wen, and M. F. Horstemeyer, “An internal state variable plasticity-based approach to determine dynamic loading history effects on material property in manufacturing processes”, *Int. J. Mech. Sci.* **47**:9 (2005), 1423–1441.
- [Johnson and Cook 1983] G. R. Johnson and W. H. Cook, “A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures”, pp. 541–547 in *Proceedings of the Seventh International Symposium on Ballistics* (The Hague, 1983), 1983.
- [Jovic et al. 2006] C. Jovic, D. Wagner, P. Hervé, G. Gary, and L. Lazzarotto, “Mechanical behaviour and temperature measurement during dynamic deformation on split Hopkinson bar of 304L stainless steel and 5754 aluminium alloy”, *J. Phys. (France) IV* **134** (2006), 1279–1285.
- [Kapoor and Nemat-Nasser 1998] R. Kapoor and S. Nemat-Nasser, “Determination of temperature rise during high strain rate deformation”, *Mech. Mater.* **27**:1 (1998), 1–12.
- [Lerch et al. 2003] V. Lerch, G. Gary, and P. Hervé, “Thermomechanical properties of polycarbonate under dynamic loading”, *J. Phys. (France) IV* **110** (2003), 159–164.
- [Longère and Dragon 2007] P. Longère and A. Dragon, “Adiabatic heat evaluation for dynamic plastic localization”, *J. Theor. Appl. Mech.* **45**:2 (2007), 203–223.
- [Longère et al. 2003] P. Longère, A. Dragon, H. Trumel, T. de Resseguier, X. Deprince, and E. Petitpas, “Modelling adiabatic shear banding via damage mechanics approach”, *Arch. Mech.* **55**:1 (2003), 3–38.
- [Longère et al. 2005] P. Longère, A. Dragon, H. Trumel, and X. Deprince, “Adiabatic shear banding-induced degradation in a thermo-elastic/viscoplastic material under dynamic loading”, *Int. J. Impact Eng.* **32**:1-4 (2005), 285–320.
- [Marchand and Duffy 1988] A. Marchand and J. Duffy, “An experimental study of the formation process of adiabatic shear bands in a structural steel”, *J. Mech. Phys. Solids* **36**:3 (1988), 251–283.
- [Mason et al. 1994] J. J. Mason, A. J. Rosakis, and G. Ravichandran, “On the strain and strain rate dependence of the fraction of plastic work converted to heat: an experimental study using high speed infrared detectors and the Kolsky bar”, *Mech. Mater.* **17**:2-3 (1994), 135–145.
- [Molinari 1985] A. Molinari, “Instabilité thermoviscoplastique en cisaillement simple”, *J. Mec. Theor. Appl.* **4** (1985), 659–684.
- [Oliferuk et al. 2004] W. Oliferuk, M. Maj, and B. Raniecki, “Experimental analysis of energy storage rate components during tensile deformation of polycrystals”, *Mater. Sci. Eng. A* **374**:1-2 (2004), 77–81.
- [Perzyna 1966] P. Perzyna, “Fundamental problems in viscoplasticity”, *Adv. Appl. Mech.* **9** (1966), 243–377.
- [Recht 1964] R. F. Recht, “Catastrophic thermoplastic shear”, *J. Appl. Mech. (Trans. ASME)* **31E** (1964), 189–193.
- [Rittel 1999] D. Rittel, “On the conversion of plastic work to heat during high strain rate deformation of glassy polymers”, *Mech. Mater.* **31**:2 (1999), 131–139.
- [Rosakis et al. 2000] P. Rosakis, A. J. Rosakis, G. Ravichandran, and J. Hodowany, “A thermodynamic internal variable model for the partition of plastic work into heat and stored energy in metals”, *J. Mech. Phys. Solids* **48**:3 (2000), 581–607.

- [Sidoroff and Dogui 2001] F. Sidoroff and A. Dogui, “Some issues about anisotropic elastic-plastic models at finite strain”, *Int. J. Solids Struct.* **38**:52 (2001), 9569–9578.
- [Subhash et al. 1994] G. Subhash, Y. J. Lee, and G. Ravichandran, “Plastic deformation of cvd textured tungsten, I: constitutive response”, *Acta Metall. Mater.* **42**:1 (1994), 319–330.
- [Taylor and Quinney 1934] G. I. Taylor and H. Quinney, “The latent energy remaining in a metal after cold working”, *Proc. R. Soc. A* **143**:849 (1934), 307–326.
- [Voyiadjis and Abed 2006] G. Z. Voyiadjis and F. H. Abed, “A coupled temperature and strain rate dependent yield function for dynamic deformations of bcc metals”, *Int. J. Plast.* **22**:8 (2006), 1398–1431.
- [Zehnder 1991] A. T. Zehnder, “A model for the heating due to plastic work”, *Mech. Res. Commun.* **18**:1 (1991), 23–28.
- [Zener and Hollomon 1944] C. Zener and J. H. Hollomon, “Effect of strain rate upon plastic flow steel”, *J. Appl. Phys.* **15**:1 (1944), 22–32.

Received 17 Dec 2007. Revised 29 May 2008. Accepted 19 Jun 2008.

PATRICE LONGÈRE: patrice.longere@univ-ubs.fr

Université Européenne de Bretagne, UBS-LIMATB (EA 4250), Rue de Saint Maudé, BP 92116, 56321 Lorient, France

ANDRÉ DRAGON: andre.dragon@lmpm.ensma.fr

ENSMA-LMPM (UMR CNRS 6617), Téléport 2, 1 avenue Clément Ader, BP 40109, 86961 Futuroscope-Chasseneuil, France

A CONTRIBUTION TO A CRITICAL REVIEW OF FRICTION STIR WELDING NUMERICAL SIMULATION

OLIVIER LORRAIN, JÉRÔME SERRI, VÉRONIQUE FAVIER,
HAMID ZAHROUNI AND MOURAD EL HADROUZ

Various modelling approaches using commercial numerical software have been proposed in the friction stir welding literature. We initiate a comparative analysis of such modellings, involving aspects such as the constitutive laws, the representation of the continuum medium and the contact between medium and tool, and the definition of the heat sources. Numerical problems are also considered: contact definition, mass scaling, and the use of the arbitrary Lagrangian Eulerian in ABAQUS/Explicit. Finally, we propose numerical tests to explore the ability of a Lagrangian code with an ALE option to simulate the process.

1. Introduction

Friction stir welding (FSW) is a process developed by The Welding Institute in the early 1990s. Its principle is simple: a rotating tool is plunged into the weld joint and is forced to translate along the joint line between two pieces of plate material which are butted together. The tool is made of two parts: a shoulder, which heats the sheet by friction (allowing the material to soften) and prevents the outflow of material due to the compressive effort, and a pin, which stirs the material to avoid holes and makes a compacts joint. The process is thus a combination of extruding, forging, and stirring. Figure 1 illustrates a weld in a 7020-T6 aluminium alloy.

The specific strengths of the process lie in the joining of the material without melting and the consequent ability to join hard-to-weld metals such as aluminium alloys, steel-aluminium, or copper-aluminium couples. FSW is used for applications where the original metal characteristics must remain unchanged as far as possible. Another advantage is that the weld can be made in all positions because no welding pool is needed. However, the process requires a very rigid clamping on the backing bar to prevent the abutting joint faces from being forced apart. In addition, the keyhole at the end of the weld can also be considered a drawback.

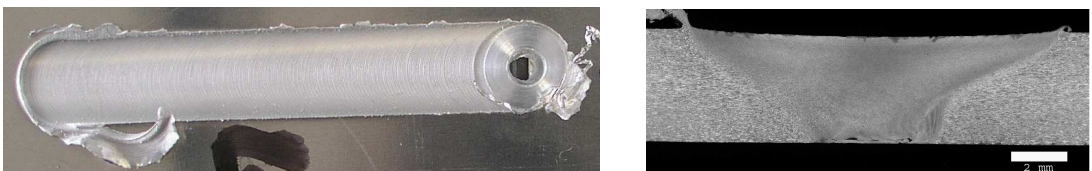


Figure 1. Top view of the weld (left) and cut orthogonal to the welding direction (right).

Keywords: friction stir welding, numerical simulation, finite element.

The authors acknowledge financial and experimental support from the Institut de Soudure and financial support from the Conseil Régional de Lorraine.

Objectives of the study. The thermomechanical conditions of FSW are very hard to determine experimentally. The profile of the pin influences the process directly, as do weld parameters such as the forging force applied to the tool and the rotational and feed speeds [Mishra and Ma 2005].

In order to optimize welding conditions for a given material and plate thickness, a lot of tests and metallurgical and mechanical characterisations of the welded joint have to be performed. Finite element method (FEM) simulations are widely used to obtain temperature and mechanical fields and to optimize the process, lowering development costs. They provide a better understanding of the flow and heating mechanisms of FSW.

A small but increasing number of papers dealing with FSW simulation have appeared in the literature. Such simulations require the modelling of friction, mechanical and thermal behaviour, and kinematics; they must also solve all the field equations. Though the mathematical laws are generally spelled out, their implementation and the numerical treatment of the differential equations are often not clearly explained.

Figure 2 offers a schematic view of inputs, outputs, and defining constraints involved in a numerical model. The inputs, on the left, consist of geometry, process parameters, and material grade. The outputs, on the right, are the mechanical and temperature fields. The choices required for the model are shown above the box.

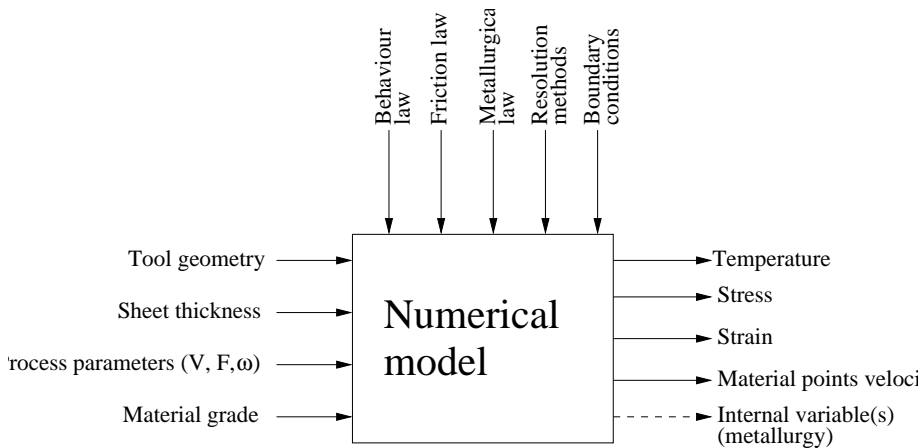


Figure 2. Schematic view of a numerical model.

This article highlights the points that require attention in FSW numerical simulations: mechanical formulation, handling of contact, boundary conditions, and constitutive laws. It also illustrates the influence of numerical parameters on the solution of the contact and field equations.

For the numerical simulation, we consider the finite element code ABAQUS/Explicit, which uses the Lagrangian formulation and offers adaptive meshing. The explicit scheme is well adapted to the building of process simulations.

Section 2 presents a state of the art modelling of the FSW process, consisting of the mechanical formulation, the heat source model, and the physical model. Section 3 discusses numerical issues concerning contact management, mass scaling, and adaptive meshing. A numerical analysis is proposed concerning FSW simulation and accounting for some aspects of the previous discussion.

Problems to be solved. Mechanically and numerically, two different but coupled problems have to be solved in order to provide a realistic computational simulation of the process. The first is the heating of the sheet, which softens the material, making it easy to stir. This heating is governed by the constitutive equation, the friction law, the contact management, thermomechanical coupling, and the boundary conditions (for example, heat exchange between the plates and the surrounding air and the anvil). The second problem is the representation of the kinematics of the flow. The flow undergoes very large deformations and controls the stirring and quality of the welded joint. Its description is therefore a fundamental clue. The challenge in modelling is to provide a numerical tool able to predict the mixing of the materials and the joint defaults.

Both features — large deformations and material stirring — are very difficult to solve numerically using FEM. Computation time is also an obstacle in FSW simulation, especially for industrial uses, and must be kept in mind.

2. Modelling the process

Compared to other joining processes, there are few studies of FSW in the literature, though in the last five years a symposium has been established to bring together different teams and record their work. Most of the papers published are devoted to experimental characterisations of the material created by FSW; only a few deal with computational simulations. This is undoubtedly due to the complexity of the process, which involves mechanical, thermal, and metallurgical phenomena all coupled together.

Mechanical formulations. Finite element simulations require discretization of the continuum medium. The equations of a thermomechanical problem can be written in two different classical formulations: Lagrangian and Eulerian. In the Lagrangian representation, the mesh is attached to material points (also called particles), whereas in the Eulerian representation, the mesh is attached to spatial points. Consequently, during the material flow, in the Lagrangian representation the mesh moves and follows the material points, whereas in the Eulerian representation the mesh is fixed. A mixed representation called the arbitrary Lagrangian Eulerian (ALE), developed to combine the advantages of the two previous ones, allows the mesh to have a different velocity than the material flow.

The Lagrangian representation is the most often used in solid mechanics, where material flow is limited. It consists in describing the movement of the continuum medium compared to a reference configuration. For each particle, which is completely defined by its position at the initial time, the classical equations for the continuum medium provide its current position and associated mechanical fields. Thus the thermomechanical history of each particles is known.

For FSW simulation, this is relevant because the final properties of the joint depend on its metallurgical state, which is strongly affected by the thermomechanical path of the particles [Chen and Kovacevic 2003a; 2003b; Gallais et al. 2004; Buffa et al. 2006b; 2006a; 2007; Fratini et al. 2007]. In addition, deformations, residual stresses, and forces applied on tools can be determined [Chao and Qi 1999; Ulysse 2002; Zhang et al. 2005; Bastier 2006]. However, in FSW, the material is so deformed and distorted that the quality of the mesh decreases during simulations, leading to numerical problems; therefore adaptive meshing (as in ABAQUS [Song and Kovacevic 2003b; 2003a; Schmidt and Hattel 2004; 2005; Zhang et al. 2005; 2007]) or remeshing (as in FORGE2005 [Fourment et al. 2004; Guerdoux and Fourment 2005], for example) is crucial in FSW simulations. In FORGE2005, to limit the size, the mesh is fine

only in a region near the tool. This mesh constraint is attached to the tool. As the tool goes forward, a remeshing is applied after some time increment which allows one to reduce distortions and to follow the tool. An alternative solution is the use of smooth particle hydrodynamics (SPH) [Tartakovsky et al. 2006], which has a Lagrangian particle nature. The material is then represented as particles interacting with each other. Other techniques based on natural element methods (NEM) have also been proposed to model the material flow of the FSW process [Alfaro et al. 2008; 2007].

The Eulerian representation is mainly used in fluid mechanics and consists of observing particles that pass successively through a spatial point. One determines the properties of the particle in terms of time and spatial coordinates. For simulations, the mesh is fixed and the material flows through it. Consequently, this representation does not capture deformations and free surfaces, and additional algorithms are required to introduce free surfaces. In the case of FSW simulations, the Eulerian representation is more appropriate for describing flow because it avoids mesh distortions, especially near the tool [Colegrove et al. 2000; Colegrove et al. 2003; Seidel and Reynolds 2003; Colegrove and Shercliff 2004; Colegrove and Shercliff 2005; 2006; Bastier 2006]. For this Eulerian approach the following commercial software has been used: CASTEM [Bastier 2006], FLUENT [Colegrove et al. 2000; Colegrove et al. 2003; Seidel and Reynolds 2003; Colegrove and Shercliff 2004; Colegrove and Shercliff 2005; 2006], and SYSWELD [Feulvarch et al. 2005a; 2005b; 2007]. However one cannot obtain sheet distortions and residual stresses because the strain history of the material is unknown.

For these reasons, the Eulerian representation is typically used for tool design analysing the material blend as well as the force applied to the tool [Colegrove et al. 2000; Colegrove et al. 2003; Colegrove and Shercliff 2004; Colegrove and Shercliff 2005; 2006], whereas the Lagrangian representation is used when the final microstructure and residual stress are desired. Some authors have used first the Eulerian formulation to obtain the temperature and flow field and then the Lagrangian formulation to compute residual stresses [Bastier 2006].

Another difficulty is determining the zone under study, which has to represent the boundary between “liquid” and “solid” material. The boundary conditions are not easy to estimate, in particular those concerning heat flux between the two parts of the sheet (the part modelled and the part not modelled).

The hybrid ALE method enjoys the advantages of both representations, while avoiding their drawbacks. In this formulation, the mesh changes during the flow but its velocity is different from that of the material, and convective terms appear in the equilibrium equations depending on the difference between the two velocities [Askes and Sluys 2000; Aymone et al. 2001; Gadala et al. 2002; Gadala 2004]. This representation has been used in the last four years to simulate the FSW process, and it seems to provide the most complete modelling. Indeed, the material flow [Xu and Deng 2002; 2003; Fourment et al. 2004; Guerdoux and Fourment 2005], the forces applied on the tool [Schmidt and Hattel 2004; 2005], and the microstructure [Xu and Deng 2003] all seem to be predictable using the ALE formulation.

ABAQUS/Explicit offers this possibility [Schmidt and Hattel 2004; 2005] via the repositioning of nodes in order to decrease mesh distortions. This is a procedure of remeshing of the structure without adding nodes which allows one to obtain elements with acceptable shape. But, even with the use of this formulation, the mixing of the materials is not captured in simulations.

Figure 3, left, presents the year-by-year evolution in the number of scientific papers using one of these three representations since the beginning of FSW modelling. Figure 3, right, shows the overall percentage of use of these representations. The Lagrangian formulation was used earliest. The Eulerian formulation

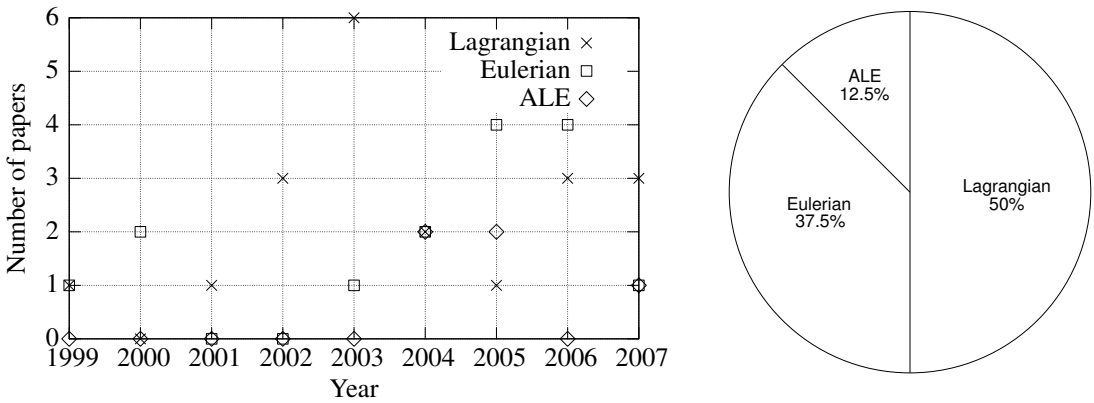


Figure 3. Year-by-year evolution of the mechanical formulation (left) and distribution over the last ten years (right).

was then introduced, perhaps to avoid the problem of excessive mesh distortion and because material flow is a crucial problem that needs to be understood if one is to get good information for industrial applications. The ALE formulation is under development and is the most promising for providing a complete numerical simulation.

To shorten computation times and reduce the difficulties linked to remeshing, one can work with separate models for two distinct regions. The first deals with thermal diffusion in the sheets and needs only a coarse spatial resolution. The second model deals with the region near the tool, which must describe correctly contact conditions, large deformation of materials, and stirring. This model can be based on FEM or on more original techniques such as SPH [Tartakovsky et al. 2006] or NEM [Alfaro et al. 2008; 2007]. To couple the two models one can use the Arlequin method originally proposed in Ben Dia 1998 (see also [Rateau 2003; Ben Dia and Rateau 2005]), or a similar technique called bridging [Xiao and Belytschko 2004]. These coupling techniques can be considered adequate for FSW modelling.

Heat sources. In FSW, temperature plays a key role because it controls the mechanical resistance of the material. During the process, heating comes from two sources: friction and plastic work. The first occurs at the interface between the tool and the sheets, whereas the second occurs in the bulk of the material. Temperature is the result of a thermal equilibrium written as

$$\rho C_p \frac{DT}{Dt} = \text{div}(\lambda \text{grad } T) + \dot{Q}, \tag{2-1}$$

where ρ is the material density, C_p the specific heat, T the temperature, D/Dt the particle derivative (material derivative), λ the thermal conductivity, and \dot{Q} the volumetric heat source strength, which we can write as $\dot{Q} = \beta \sigma : D$, where D is the deformation velocity gradient tensor, σ is the stress tensor, and $\beta = 0.9$.

Q_{surf} , the surface heat source due to friction, appears as a boundary condition in the interface between the tool and the sheet:

$$Q_{\text{surf}} = \lambda \text{grad}(T) \cdot n, \tag{2-2}$$

where n is the normal to the surface.

To determine the temperature, three different approaches in the literature can be distinguished, according to the desired objective (heating, material flow, microstructure, residual state, and so on). The differences between these three approaches also come from the level of integration of the process parameters and tool geometry and the calculation time needed. It is always necessary to find a compromise between completeness and computation time.

The first approach consists in using experimental measurements to determine the heat flux by an inverse method. [Chao and Qi \[1999\]](#) and [Dickerson et al. \[2003\]](#) measured temperature using thermocouples to determine the heat flux. Several teams have used torque measurements to determine the power involved in the process and so the heat flux [[Dickerson et al. 2003](#); [Shi et al. 2003](#); [Khandkar et al. 2006](#); [Simar et al. 2004](#); [Gallais et al. 2004](#)].

The second approach is based on an analytical formulation of the heat flux

$$Q_{\text{surf}} = \mu F \omega \alpha, \quad (2-3)$$

where Q_{surf} is the heat flux, μ is the friction coefficient between tool and sheet, F is the plunge force, ω is the rotational velocity, and α is a parameter related to the tool geometry.

This equation is then used either in analytical modelling using the Rosenthal equation [[Rosenthal and Schmerber 1938](#); [Rosenthal 1941](#); [Feng et al. 1998](#); [Mandal and Williamson 2006](#)], or in finite element simulations using moving sources [[Song and Kovacevic 2003b](#); [2003a](#); [Chen and Kovacevic 2003a](#)].

The last approach tends to calculate the heat flux directly by modelling the mechanical role of the tool and so is clearly more predictive than the first approach. The geometry of the tool and its mechanical interaction with the sheet is then determined by means of finite element simulations [[Ulysse 2002](#); [Fourment et al. 2004](#); [Schmidt and Hattel 2004](#); [2005](#); [Guerdoux and Fourment 2005](#); [Buffa et al. 2006b](#); [2006a](#); [2007](#); [Fratini et al. 2007](#)]. The heat generated by friction between tool and sheets is modelled using the classical friction equation

$$Q_{\text{surf}} = \tau v_s, \quad (2-4)$$

relating the tangential stress τ to the sliding velocity v_s through the friction parameters.

Clearly the contact conditions are critical for an accurate and realistic determination of the heat flux. If the tangential stress and/or the sliding velocity are not well estimated, the heat flux is not well estimated. The contact surface also plays an important role, because the total flux is given by the integration of (2-4) over the whole surface.

Physical models. In this section, the constitutive equation and friction law commonly used for FSW simulation are discussed.

Constitutive law. During the FSW process, the material is submitted to changes of temperature of a few hundred degrees and to widely varying strain rates. Thus the constitutive equation has to capture the strain-rate- and temperature-dependent behaviour over a large range of these parameters. In addition, the behaviour of the material changes with deformation related to strain hardening and microstructure evolution. In the literature, one can find both very simple and more sophisticated modellings involving microstructural changes, all of them based on a viscoplastic formalism, whether or not using a plastic threshold. For example, the model of [[Myhr and Grong 1991a](#); [1991b](#)] is used in [[Bastier et al. 2006](#); [2008](#); [Feulvarch et al. 2005a](#); [2005b](#); [2007](#); [Feulvarch 2006](#)]; it takes into account the dissolution of precipitates due to the temperature evolution in aluminium alloys.

In the viscoplastic formalism, two constitutive equations are more currently used. The first is the Johnson–Cook relation, used for example by Schmidt and Hattel [2004; 2005], which phenomenologically accounts for the strain, strain rate, and temperature dependence of the material behaviour in an uncoupled manner. It is written

$$\sigma_{em} = (A + B[\epsilon_{eq}^{pl}]^n) \left(1 + C \ln \frac{\dot{\epsilon}_{eq}^{pl}}{\dot{\epsilon}_0}\right) \left(1 - \left[\frac{T - T_{ref}}{T_{melt} - T_{ref}}\right]^m\right) \tag{2-5}$$

where σ_{eq} , ϵ_{eq}^{pl} , $\dot{\epsilon}_{eq}^{pl}$, and $\dot{\epsilon}_0$ (1 s^{-1}) are respectively the equivalent von Mises stress, the equivalent plastic strain, the equivalent plastic strain rate, and the normalising strain rate, while A , B , C , n , T_{melt} , T_{ref} , and m are material constants. The relation captures a plastic threshold and consequently accepts the presence of residual stresses when the strain rate is reduced to zero.

Power law-type constitutive equations such as the Norton–Hoff model are also used [Fourment et al. 2004; Guerdoux and Fourment 2005]. They are currently combined with the Eulerian representation

$$\sigma_{eq} = 3K (\sqrt{3}\dot{\epsilon}_{eq})^{m-1} \dot{\epsilon}_{eq}, \tag{2-6}$$

where $K = K_0(\epsilon + \epsilon_0)^n \exp(\beta/T)$ is the consistency, $\dot{\epsilon}_{eq}$ is the equivalent strain rate, and m , K_0 , ϵ_0 , n , and β are material constants. Contrary to the Johnson–Cook equation, when the strain rate is reduced to zero, no more stress exists within the material.

Contact law. For friction modelling, only two laws are found in the literature. The first, used mainly in solid mechanics [Xu and Deng 2001; Xu and Deng 2002; 2003; Schmidt and Hattel 2004; 2005], is a modification of Coulomb friction law

$$\tau = \mu p, \tag{2-7}$$

which expresses the tangential stress as a function of the contact pressure p and the friction coefficient μ . It is modified in that the tangential stress is limited to τ_{max} , a value related to the ultimate tensile strength σ_u by $\tau_{max} = \sigma_u/\sqrt{3}$; see Figure 4.

The second friction law is the Norton law, currently used in fluid mechanics [Fourment et al. 2004; Guerdoux and Fourment 2005; Feulvarch et al. 2005a; 2005b; 2007; Feulvarch 2006]. It expresses the tangential stress in terms of the friction coefficient μ , the consistency K , and the differential velocity V :

$$\tau = -\mu K |V|^{q-1} V, \tag{2-8}$$

where q is a material constant.

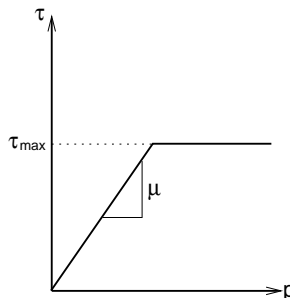


Figure 4. Modified Coulomb friction law.

But the heat partition between sheet and tool is not really justified by the different authors. How much heat is dissipated in the sheet? The only physical consideration used to determine this is the difference in material effusivity of each part. This problem needs to be highlighted because it governs the temperature evolution in the sheet, and so the final result. Moreover, the definition of friction must vary with the velocity and temperature in the transient phase. For the steady state, even if the conditions are known, the friction coefficient must be correctly evaluated.

3. Numerical issues

In addition to the thermomechanical modelling, algorithms have to be chosen to deal with the numerical resolution of the thermal and mechanical equations of the problem. This section considers three numerical issues in ABAQUS/Explicit: contact definition, mass scaling, and ALE use. These aspects are unfortunately not analysed in the literature, although they strongly affect the numerical results.

3.1. Contact definition. We turn first to the management of contact between sheet and tool, and in particular the contact between shoulder and sheet. Contact problems are among the most delicate modelled in the literature [Kikuchi 1982; Alart and Curnier 1991; Fortin and Glowinski 1983; Perić and Owen 1992; Simo and Laursen 1992; Wriggers 1995; Papadopoulos and Solberg 1998; Aggoune et al. 2006].

ABAQUS offers two main approaches for unilateral contact. The first is based on a hard contact definition and uses a Lagrange multiplier, representing the contact pressure, to enforce the contact constraint. The second is based on a softened contact definition, postulating a regular relation between the contact pressure and the overclosure between the master and slave surfaces.

Here we focus on the second approach, and ask whether the friction coefficient imposed in the input file of ABAQUS/Explicit is respected. Only the interaction between the shoulder and the sheets is considered; the interface between the pin and the sheet is not modelled, to uncouple the roles of each part of the tool. The master surface is the tool, which is considered undeformable, and the slave surface is the sheet.

For the slave surface definition, two choices are offered by ABAQUS/Explicit: node-based, where the contact forces are considered directly at the nodes, or element-based, where the reactions are calculated on the surface of the elements. ABAQUS/Explicit warns that “a node-based surface should be used with caution or not at all if accurate contact stresses are needed or if heat will be exchanged between the two surfaces” [ABAQUS 2006]. To explore this recommendation more precisely, numerical computations were performed using these two choices, as follows. The friction law is considered without the limitation of the equivalent friction stress. The dwelling phase, where the tool is in contact with the sheets and rotates around its own axis, is simulated. The friction coefficient is then postcalculated from the computed stresses and must be compared with the coefficient imposed in the input file ($\mu = 0.3$):

$$\mu = \frac{\tau_{\text{eq}}}{\sigma_{33}}, \quad (3-1)$$

where $\tau_{\text{eq}} = \sqrt{\sigma_{13}^2 + \sigma_{23}^2}$ is the equivalent friction stress and σ_{33} is the normal stress (the index 3 corresponds to the normal to the contact surface).

Figure 5 presents the postcalculated friction coefficient μ for four different elements, sketched in white, in contact with the shoulder. It reveals that the element based definition allows μ to be closer to the prescribed friction coefficient than the node based definition even though the value is not exactly the one imposed in the input file. This confirms the ABAQUS/Explicit recommendation.

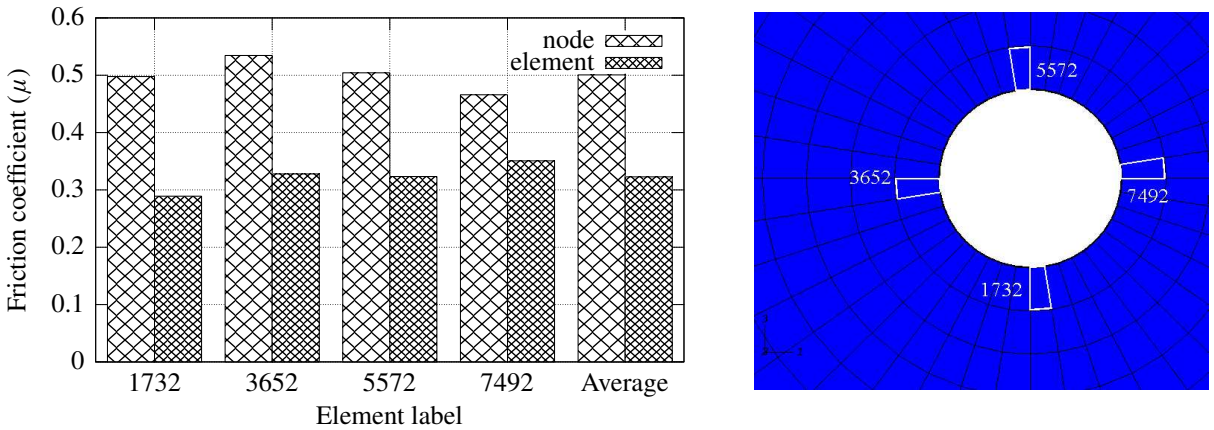


Figure 5. Friction coefficient (left) calculated using node-based and element-based surfaces, for the four elements highlighted on the right.

3.2. Mass scaling influence. In a FEM, the dynamic problem is written for each node of the mesh as

$$P - I = M\ddot{u}, \tag{3-2}$$

where P , I , M , and \ddot{u} represent respectively the applied forces, the internal forces, the mass matrix, and the node acceleration.

ABAQUS/Explicit solves the equilibrium (3-2) by an explicit algorithm. The acceleration of each node is determined as a function of its mass and forces applied to it:

$$\ddot{u}|_t = M^{-1}(P - I)|_t. \tag{3-3}$$

The mass matrix used in this formulation is diagonal, which simplifies the explicit calculation. The node velocities at time $(t + \Delta t/2)$ are determined by

$$\dot{u}|_{t+\Delta t/2} = \dot{u}|_{t-\Delta t/2} + \frac{\Delta t|_{t+\Delta t} + \Delta t|_t}{2} \ddot{u}|_t. \tag{3-4}$$

Then, the displacements of nodes are determined by

$$u|_{t+\Delta t} = u|_t + \Delta t|_{t+\Delta t} \dot{u}|_{t+\Delta t/2}. \tag{3-5}$$

As shown by equations (3-3), (3-4) and (3-5), the dynamic computation in ABAQUS/Explicit does not need a stiffness matrix decomposition. The solutions are given at the end of the time increment without any iteration. So each increment needs little CPU time. However, the algorithm has a time cost related to the mesh size. In fact, the time increment (Δt) must be less than a certain limit, to ensure the stability of the calculation and avoid fluctuations in the solution. This limit is the smallest time needed for an elastic wave to propagate through a mesh element [Schmidt and Hattel 2005; ABAQUS 2006; Liu 2006; Wang et al. 2007]:

$$\Delta t = \frac{L_e}{C_d}, \tag{3-6}$$

where L_e is the length of the smallest mesh element and C_d is the celerity of elastic waves through the element considered. For a solid in an elastic regime, we have

$$C_d = \sqrt{\frac{E}{\rho}}, \tag{3-7}$$

where E is the Young’s modulus and ρ the mass density of the material. Therefore, once the mesh is determined, the simulation time is proportional to the real duration of the process to be simulated.

To reduce computation time without increasing the element size (which would compromise the simulation quality at the contact surface), one can increase artificially the value of the density ρ , and so also, according to (3-6) and (3-7), the value of Δt . This technique is known as mass scaling, and it used also in other explicit software such as LS-DYNA [Liu 2006; Wang et al. 2007] and METAFOR [Papeleux and Ponthot 2002].

To analyse the effect of mass scaling, we compared the friction coefficient obtained using four different time increments in the previously described simulations. (The increment time was selected directly, and ABAQUS/Explicit determines the corresponding mass factor.) Figure 6 shows the results obtained for the same elements used in Figure 5. Clearly, for a high value of mass scaling (n_3 and e_3), fluctuations are important and the results are not physically acceptable, since the calculated friction coefficient is higher than 1. With less drastic mass scaling, the element-based friction coefficient displays a smaller dependence on mass scaling than the node-based one. This again confirms the ABAQUS/Explicit recommendation, and the element-based surface was chosen in further simulations.

3.3. ALE in ABAQUS/Explicit. Since large deformations occur during the process, mesh distortions appear quickly in a Lagrangian finite element simulation. ABAQUS/Explicit offers an adaptive meshing mode called ALE [Song and Kovacevic 2003b; 2003a; Schmidt and Hattel 2004; 2005; Zhang et al. 2005; 2007] which repositions nodes with no change in the number or connectivity of elements. Figure

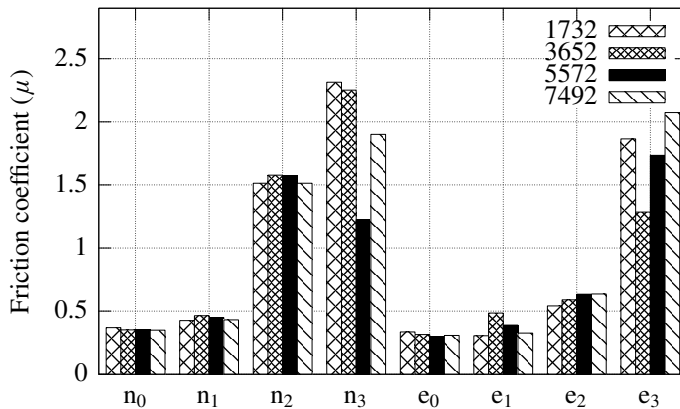


Figure 6. Influence of mass scaling on the calculated friction coefficient. Notation: “n” and “e” stand for node- and element-based surfaces; the subscripts indicate increasingly more drastic mass scaling (0 corresponds to $\Delta t = 0.00008$, or no mass scaling, while 1, 2 and 3 correspond to $\Delta t = 0.0001, 0.0005$, and 0.001 , respectively).

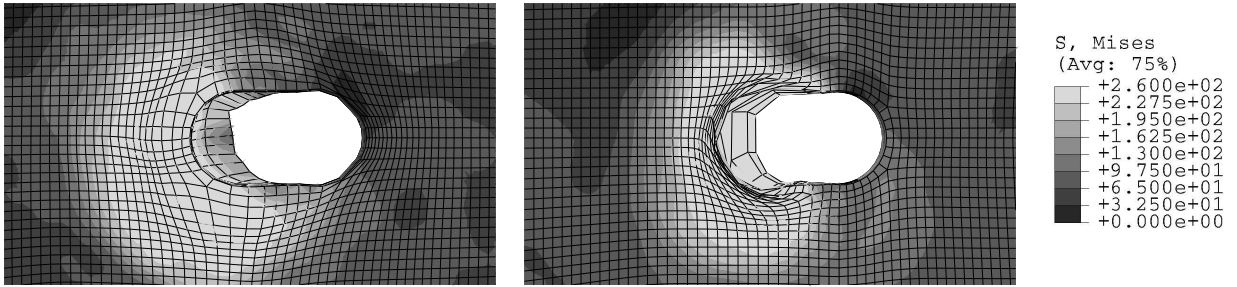


Figure 7. Von Mises stress distribution after a 5 mm translation with adaptive meshing (left) and without it (right).

7 shows the von Mises stress distribution obtained with and without the adaptive meshing for a pure elastoplastic simulation where the tool undergoes a 5 mm translation. The distortions clearly decrease but do not vanish when the ALE method is used. The element in contact with the tool gets thinner and thinner due to the adaptive meshing and finally the small thickness of the elements makes the calculation diverge.

Two parameters must be chosen to perform the adaptive meshing: the frequency of node repositioning and the number of mesh sweeps (the number of times the algorithm is used). Their values have to be high enough to avoid distortions, but not so high as to degrade the information excessively. (When nodes are moved during each sweep, the fields must be transferred to the new mesh, leading to a loss of information due to interpolation.)

4. Numerical analysis

The preceding analysis can be of help in making certain modelling and numerical choices for FSW simulations. Ultimately, a systematic exploration of such choices can clarify how the temperature field is influenced by process parameters such as the plunge force or the translation and rotation speeds of the tool. The results presented here are the first ones obtained after having chosen the parameters presented in earlier sections.

Material and geometry. The tool is considered analytically rigid, which means in ABAQUS/Explicit terminology undeformable and isothermal. It is made of two simple cylinders: an inferior rounded one for the pin, with radius $r_p = 2.5$ mm and height $h_p = 3$ mm, and an upper one for the shoulder, with radius $r_e = 6.5$ mm and height $h_e = 5$ mm as illustrated in [Figure 8](#).

The sheet is modelled by the extrusion of a square of length $L_s = 40$ mm with a hole in the center of diameter $D_s = 5$ mm and thickness $E_s = 4$ mm. The anvil is also modelled by the extrusion of the same square (without the hole) with thickness $E_a = 5$ mm.

The sheet is made of aluminium 2024-T3. The Johnson–Cook constitutive equation, (2-5), is chosen, using the following parameters, taken from [[Schmidt and Hattel 2004](#); [2005](#)]:

$$\begin{aligned}
 A &= 369 \text{ MPa}, & C &= 0.0083, & T_{\text{melt}} &= 502^\circ\text{C}, & m &= 1.7, \\
 B &= 684 \text{ MPa}, & & & T_{\text{ref}} &= 25^\circ\text{C}, & n &= 0.73.
 \end{aligned}$$

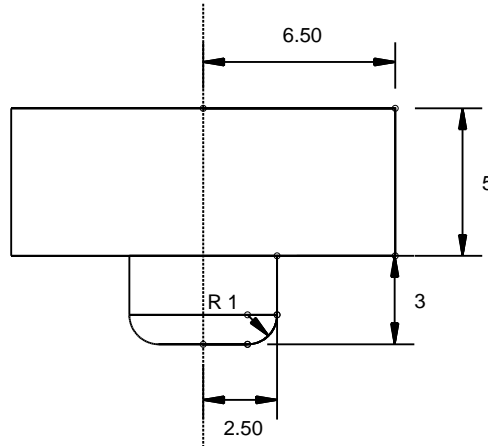


Figure 8. Tool used for simulations (dimensions in mm).

The tool and the anvil are made of classical steel. The physical properties (material conductivity λ , specific heat C_p , and density ρ) used for aluminium and steel are as follows:

| | λ ($\text{W m}^{-1} \text{K}^{-1}$) | C_p ($\text{J kg}^{-1} \text{K}^{-1}$) | ρ (kg m^{-3}) |
|-----------|---|--|-------------------------------|
| Aluminium | 140 | 920 | 2780 |
| Steel | 36 | 450 | 7800 |

Based on experimental observations, the constitutive equation for the anvil is limited to linear elasticity (using a Young’s modulus of 2.1×10^5 MPa and a Poisson’s ratio of 0.3) in order to decrease computation time and because only the behaviour of the sheet is being studied.

The process is decomposed into three steps, presented in Figure 9:

- (1) Plunge (preexisting hole, 1 s): the tool is in contact with the sheet; a force $F = 1200$ N orthogonal to the sheet is applied to the tool.
- (2) Dwelling (beginning of tool rotation, 2 s): an additional angular velocity $\omega = 42$ rad/s is imposed on the tool.

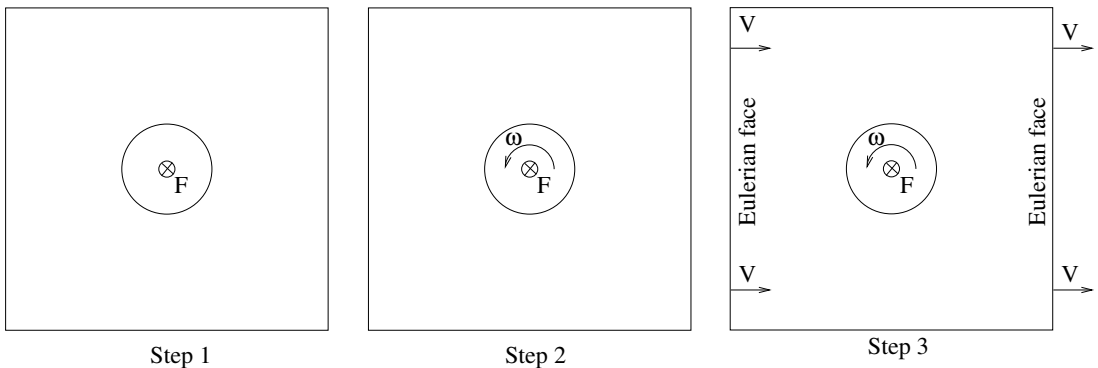


Figure 9. Boundary conditions of each step.

(3) Translation (10 s): Eulerian faces are used to simulate the material flow in the sheet; material flows through these faces with velocity $V = 0.5$ mm/s. This choice is helpful in decreasing mesh distortion.

Adaptive meshing is used only in steps 2 and 3, and is applied to the entire sheet with frequency 1 (at the end of each increment); the number of sweeps is 10. These parameters are chosen from the optimization study performed previously.

The definition of the slave surface of the contact between tool and sheet is element based as justified in [Section 3.1](#).

For these simulations, about 50000 hexahedral elements with reduced integration (C3D8RT) were used, and the computation time varied between 3 and 19 hours. The translation step accounts for most of the CPU time overall.

Temperature results. As discussed, the use of mass scaling can decrease computation time but may lead to loss of precision and hence predictability. Here we consider the effect of mass scaling on the calculated temperature field. We chose three elements on the sheet, shown in [Figure 10](#): element A is under the shoulder and in contact with the pin; element B lies under the shoulder, 6 mm from the centre of the sheet; and element C is 9.5 mm away from the centre of the sheet.

[Figure 11](#) shows the calculated temperature evolution at these elements. The same values of Δt are used as in [Figure 6](#), reflecting increasingly drastic mass scaling. First, the temperature found is of the same order as the experimental one observed in the test using conditions close to the numerical simulation. The parameters are taken from the literature and consequently are not fitted to match experimental results. Second, for elements A and B, fluctuations clearly appear for the whole simulation. This phenomenon is magnified with the use of mass scaling. For element A, the mass scaling leads to an underestimation of the temperature value, and for element B to an overestimation. Thus mass scaling can cause errors in either direction in the calculation of temperature. For element C, which is far from the tool, there is no fluctuation and mass scaling has only a weak influence. So the overall heating due to the tool is well predicted but local results fluctuate and are not really accurate because mass scaling has an effect on the management of the contact.

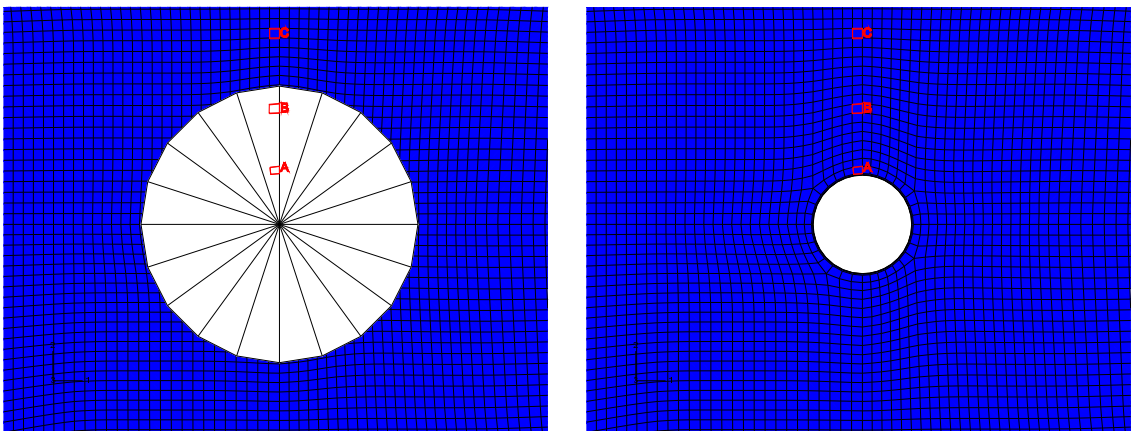


Figure 10. Position of elements in the sheet for the complete simulation.

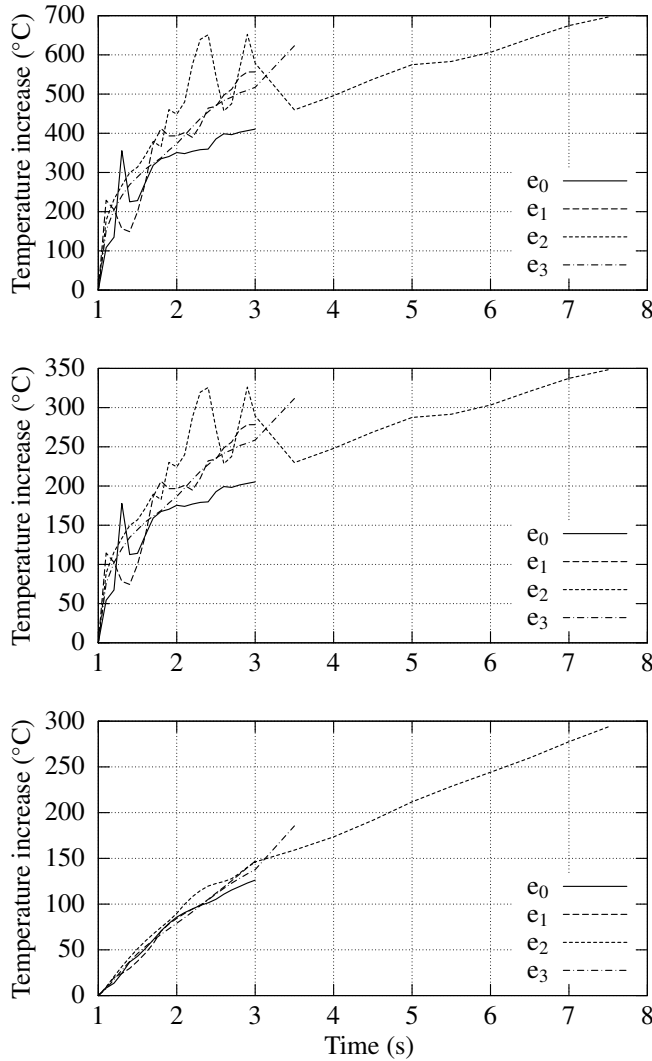


Figure 11. Increase in temperature for elements A (top), B (center), and C (bottom); simulation using element-based surfaces with Δt equal to 0.00008 (e_0), 0.0001 (e_1), 0.0005 (e_2), and 0.001 (e_3). The time scale begins at 1 s because step 1 is omitted.

Material flow results. Understanding and controlling the material flow are crucial to the success of the FSW process, are closely related to the temperature field. In view of the link between microstructure evolution and the strain path of the particle [Xu and Deng 2003; Schmidt and Hattel 2004], we also study the latter: Figure 12 shows the von Mises equivalent plastic strain partition in a cut perpendicular to the welding direction. We can clearly associate the zone of highest equivalent plastic strain with the thermomechanical affected zone (TMAZ). The vase shape of the TMAZ and the lack of symmetry between the advancing and retreating sides are in good agreement with experimental observations and numerical simulations found in [Xu and Deng 2003; Schmidt and Hattel 2004; Mishra and Ma 2005]. We notice the presence of a highly deformed zone at the center of the TMAZ, corresponding to the nugget.

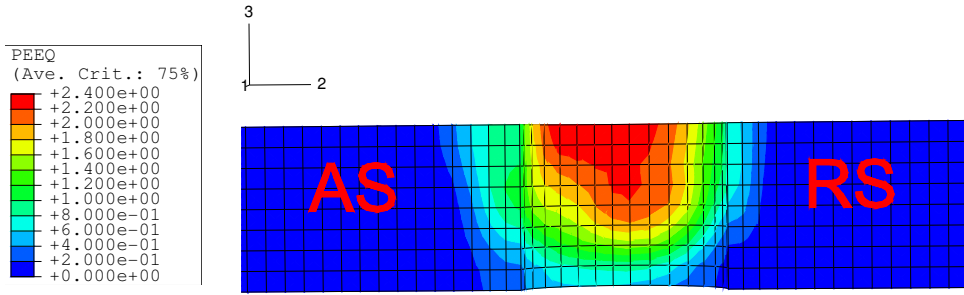


Figure 12. Cross-section of the weld: von Mises equivalent plastic strain partition. AS is the advancing side, RS the retreating side.

Figure 13 shows the results concerning the elements near the pin for different sections through the sheet, perpendicular to the axis of the tool. The results come from an analytical calculation in which the particle velocities are imposed on the contact surface and depend on the advancing speed, the angular velocity, and the pitch of the pin.

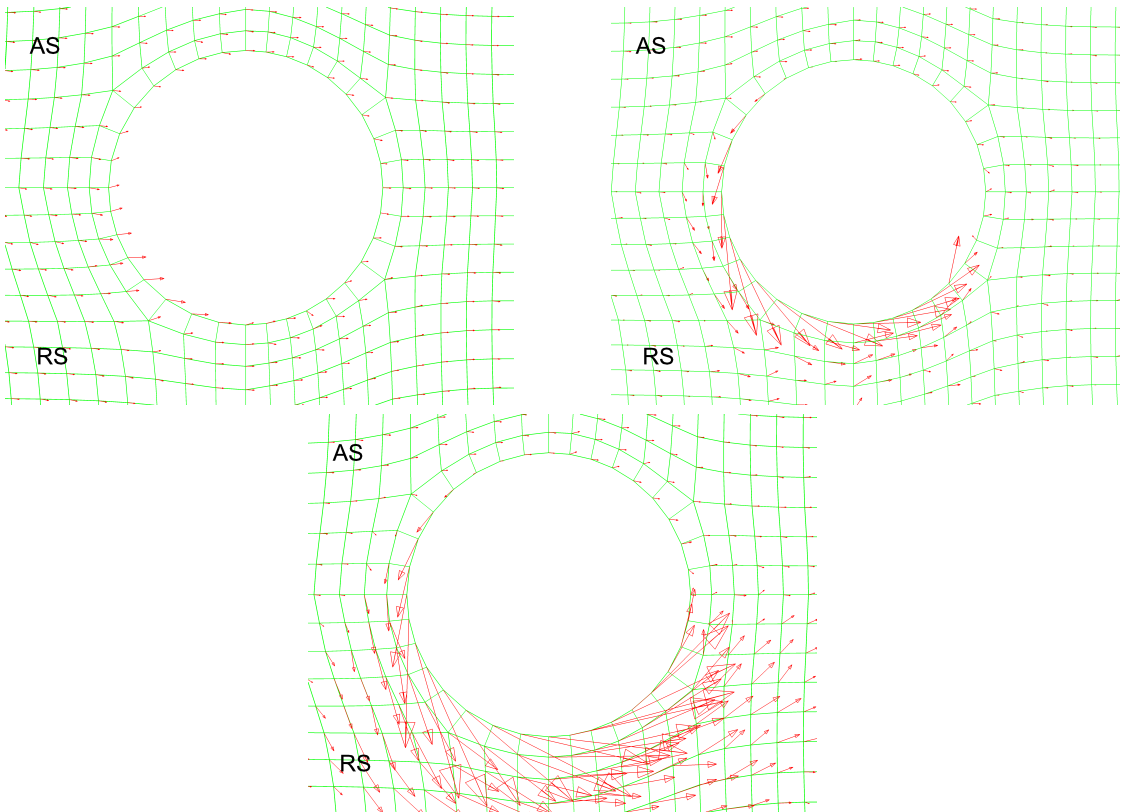


Figure 13. Calculated angular velocity and flow velocity: under the shoulder (top left), mid thickness (top right), and at the bottom of the weld (bottom). For each figure, the advancing side is at the top and the retreating side is at the bottom.

The results agree qualitatively with those in [Nandan et al. 2007]; see especially Figure 15 there, showing how that material moves from the advancing side to the retreating side. This observation was made using thin copper sheets as material markers. Quantitatively, our simulation results obtained are close to those of Nandan et al. for the upper part of the weld (contact with the shoulder), except near the tool on the advancing side where the material flow does not follow the movement of the tool. In addition, in the bottom part of the weld the rotational motion of the pin does not affect the material flow. It is similar to a flow around a fixed obstacle that does not stick to it. This phenomenon may come from a too low normal stress at the pin/sheet interface, which does not provide high enough tangential stress. In the upper part, the normal force is imposed on the tool and stirs the material. Note that a hole is still present behind the tool, even though a small amount of material seems to be stirred behind the tool.

5. Conclusion

We have examined different approaches used in the literature and based on commercial codes to simulate the FSW process, highlighting their strengths and weaknesses. The choice of an approach is motivated by the desired predictive feature for the simulation.

From the viewpoint of kinematics, the ALE representation seems to be the most promising method to reduce mesh distortions; it does not prevent computationally expensive remeshing but reduces it.

One way to reduce computation time is to use mass scaling. However, excessive mass scaling can degrade the quality of the outputs. The technique of coupling two different models seems to provide a good compromise, since it allows us to largely reduce or avoid remeshing.

For the model to be predictive, material mixing must be well described in order to predict defects like void formation in the weld joint. This is possible if the complex geometry of the pin and the shoulder is considered in the simulation.

References

- [ABAQUS 2006] *ABAQUS user's manual*, Hibbit, Karlsson and Sorensen, Pawtucket, RI, 2006.
- [Aggoune et al. 2006] W. Aggoune, H. Zahrouni, and M. Potier-Ferry, “Asymptotic numerical methods for unilateral contact”, *Int. J. Numer. Methods Eng.* **68**:6 (2006), 605–631.
- [Alart and Curnier 1991] P. Alart and A. Curnier, “A mixed formulation for frictional contact problems prone to Newton like solution methods”, *Comput. Methods Appl. Mech. Eng.* **92**:3 (1991), 353–375.
- [Alfaro et al. 2007] I. Alfaro, L. Fratini, E. Cueto, F. Chinesta, and F. Micari, “Meshless simulation of friction stir welding”, pp. 203–208 in *Materials processing and design: modeling, simulation and applications: NUMIFORM '07* (Porto, 2007), edited by J. M. A. César de Sá and A. D. Santos, AIP Conference Proceedings **908**, American Inst. of Physics, Melville, NY, 2007.
- [Alfaro et al. 2008] I. Alfaro, L. Fratini, E. Cueto, and F. Chinesta, “Numerical simulation of friction stir welding by natural element methods”, *Int. J. Mater. Form.* Supplement 1 (2008), 1079–1082.
- [Askes and Sluys 2000] H. Askes and L. J. Sluys, “Remeshing strategies for adaptive ALE analysis of strain localisation”, *Eur. J. Mech. A Solids* **19**:3 (2000), 447–467.
- [Aymone et al. 2001] J. L. F. Aymone, E. Bittencourt, and G. J. Creus, “Simulation of 3D metal-forming using an arbitrary Lagrangian–Eulerian finite element method”, *J. Mater. Process. Technol.* **110**:2 (2001), 218–232.
- [Bastier 2006] A. Bastier, *Modélisation du soudage d'alliages d'aluminium par friction et malaxage*, Ph.D. thesis, École Polytechnique, Palaiseau, 2006, Available at <http://pastel.paristech.org/2089>.
- [Bastier et al. 2006] A. Bastier, M. H. Maitournam, K. Dang Van, and F. Roger, “Steady state thermomechanical modelling of friction stir welding”, *Sci. Technol. Weld. Joi.* **11**:3 (2006), 278–288.

- [Bastier et al. 2008] A. Bastier, M. H. Maitournam, F. Roger, and K. Dang Van, “Modelling of the residual state of friction stir welded plates”, *J. Mater. Process. Technol.* **200**:1–3 (2008), 25–37.
- [Ben Dia 1998] H. Ben Dia, “Multiscale mechanical problems: the Arlequin method”, *C. R. Acad. Sci. II B Mec.* **326** (1998), 899–904.
- [Ben Dia and Rateau 2005] H. Ben Dia and G. Rateau, “The Arlequin method as a flexible engineering design tool”, *Int. J. Numer. Methods Eng.* **62**:11 (2005), 1442–1462.
- [Buffa et al. 2006a] G. Buffa, J. Hua, R. Shivpuri, and L. Fratini, “Design of the friction stir welding tool using the continuum based FEM model”, *Mater. Sci. Eng. A* **419**:1–2 (2006), 381–388.
- [Buffa et al. 2006b] G. Buffa, J. Hua, R. Shivpuri, and L. Fratini, “A continuum based FEM model for friction stir welding: model development”, *Mater. Sci. Eng. A* **419**:1–2 (2006), 389–396.
- [Buffa et al. 2007] G. Buffa, L. Fratini, and R. Shivpuri, “CDRX modelling in friction stir welding of AA7075-T6 aluminum alloy: analytical approaches”, *J. Mater. Process. Technol.* **191**:1–3 (2007), 356–359.
- [Chao and Qi 1999] Y. J. Chao and X. Qi, “Heat transfer and thermo-mechanical analysis of friction stir joining of AA6061-T6 plates”, in *1st International Symposium on Friction Stir Welding* (Thousand Oaks, CA, 1999), TWI, Cambridge, 1999.
- [Chen and Kovacevic 2003a] C. Chen and R. Kovacevic, “Finite element modelling of thermomechanical performance of friction stir welding”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Chen and Kovacevic 2003b] C. M. Chen and R. Kovacevic, “Finite element modeling of friction stir welding: thermal and thermomechanical analysis”, *Int. J. Mach. Tool. Manuf.* **43**:13 (2003), 1319–1326.
- [Colegrove and Shercliff 2004] P. A. Colegrove and H. R. Shercliff, “Modelling the friction stir welding of aerospace alloys”, in *5th International Symposium on Friction Stir Welding* (Metz, 2004), TWI, Cambridge, 2004.
- [Colegrove and Shercliff 2005] P. A. Colegrove and H. R. Shercliff, “3-Dimensional CFD modelling of flow round a threaded friction stir welding tool profile”, *J. Mater. Process. Technol.* **169**:2 (2005), 320–327.
- [Colegrove and Shercliff 2006] P. A. Colegrove and H. R. Shercliff, “CFD modelling of friction stir welding of thick plate 7449 aluminium alloy”, *Sci. Technol. Weld. Joi.* **11**:4 (2006), 429–441.
- [Colegrove et al. 2000] P. A. Colegrove, M. Painter, D. Graham, and T. Miller, “3-dimensional flow and thermal modelling of the friction stir welding process”, in *2nd International Symposium on Friction Stir Welding* (Gothenborg, 2000), TWI, Cambridge, 2000.
- [Colegrove et al. 2003] P. A. Colegrove, H. R. Shercliff, and P. L. Threadgill, “Modelling and development of the Trivex friction stir welding tool”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Dickerson et al. 2003] T. L. Dickerson, Q.-Y. Shi, and H. R. Shercliff, “Heat flow into friction stir welding tools”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Feng et al. 1998] Z. Feng, J. E. Gould, and T. J. Lienert, “A heat flow model for friction stir welding of aluminium alloys”, pp. 149–158 in *Hot deformation of aluminum alloys II: proceedings of the Second Symposium held at the 1998 TMS Fall Meeting* (Rosemont, IL, 1998), edited by T. R. Bieler et al., TMS, Warrendale, PA, 1998.
- [Feulvarch 2006] E. Feulvarch, *Modélisation numérique du soudage par friction-malaxage (friction stir welding)*, Ph.D. thesis, Université Jean Monnet, Saint-Etienne, 2006.
- [Feulvarch et al. 2005a] E. Feulvarch, F. Boitout, and J.-M. Bergheau, “Friction stir welding: modélisation de l’écoulement de la matière pendant la phase de soudage”, in *17ème Congrès Français de Mécanique* (Troyes, 2005), Université Technologie de Troyes, Troyes, 2005.
- [Feulvarch et al. 2005b] E. Feulvarch, F. Boitout, and J.-M. Bergheau, “Modélisation thermomécanique et microstructurale du soudage par friction-malaxage (friction stir welding): développement d’un modèle élément fini”, pp. 347–352 in *Septième Colloque National en Calcul des Structures* (Giens, 2005), vol. 2, edited by R. Ohayon et al., Hermes Science/Lavoisier, Paris, 2005.
- [Feulvarch et al. 2007] E. Feulvarch, V. Robin, F. Boitout, and J.-M. Bergheau, “A 3D finite element modelling for thermofluid flow in friction stir welding”, pp. 711–724 in *Mathematical modelling of weld phenomena* (Graz, 2006), vol. 8, edited by H. Cerjak et al., Verlag der Technischen Universität Graz, Graz, 2007.
- [Fortin and Glowinski 1983] M. Fortin and R. Glowinski, *Augmented Lagrangian methods: applications to the numerical solutions of boundary-value problems*, Studies in Mathematics and its Applications **15**, North-Holland, Amsterdam, 1983.

- [Fourment et al. 2004] L. Fourment, S. Guerdoux, M. Miles, and T. Nelson, “Numerical simulation of the friction stir welding process using both Lagrangian and arbitrary Lagrangian Eulerian formulations”, in *5th International Symposium on Friction Stir Welding* (Metz, 2004), TWI, Cambridge, 2004.
- [Fratini et al. 2007] L. Fratini, G. Buffa, and R. Shivpuri, “Improving friction stir welding of blanks of different thicknesses”, *Mater. Sci. Eng. A* **459**:1–2 (2007), 209–215.
- [Gadala 2004] M. S. Gadala, “Recent trends in ALE formulation and its applications in solid mechanics”, *Comput. Methods Appl. Mech. Eng.* **193**:39–41 (2004), 4247–4275.
- [Gadala et al. 2002] M. S. Gadala, M. R. Movahhedy, and J. Wang, “On the mesh motion for ALE modeling of metal forming processes”, *Finite Elem. Anal. Des.* **38**:5 (2002), 435–459.
- [Gallais et al. 2004] C. Gallais, A. Denquin, A. Pic, A. Simar, T. Pardoën, and Y. Brechet, “Modelling the relationship between process parameters, microstructural evolutions and mechanical behaviour in a friction stir welded 6xxx aluminium alloy”, in *5th International Symposium on Friction Stir Welding* (Metz, 2004), TWI, Cambridge, 2004.
- [Guerdoux and Fourment 2005] S. Guerdoux and L. Fourment, “ALE formulation for the numerical simulation of friction stir welding”, in *Computational plasticity: fundamentals and applications: proceedings of the Eighth International Conference on Computational Plasticity, COMPLAS VIII* (Barcelona, 2005), edited by E. Oñate et al., CIMNE, Barcelona, 2005. Extended abstract available at <http://congress.cimne.upc.es/complas05/admin/Files/FilePaper/p50.pdf>.
- [Khandkar et al. 2006] M. Z. H. Khandkar, J. A. Khan, A. P. Reynolds, and M. A. Sutton, “Predicting residual thermal stresses in friction stir welded metals”, *J. Mater. Process. Technol.* **174**:1–3 (2006), 195–203.
- [Kikuchi 1982] N. Kikuchi, “A smoothing technique for reduced integration penalty methods in contact problems”, *Int. J. Numer. Methods Eng.* **18**:3 (1982), 343–350.
- [Liu 2006] C.-H. Liu, “The simulation of the multi-pass and die-less spinning process”, *J. Mater. Process. Technol.* **192-193** (2006), 518–524.
- [Mandal and Williamson 2006] S. Mandal and K. Williamson, “A thermomechanical hot channel approach for friction stir welding”, *J. Mater. Process. Technol.* **174**:1–3 (2006), 190–194.
- [Mishra and Ma 2005] R. S. Mishra and Z. Y. Ma, “Friction stir welding and processing”, *Mater. Sci. Eng. R* **50**:1–2 (2005), 1–78.
- [Myhr and Grong 1991a] O. R. Myhr and Ø. Grong, “Process modelling applied to 6082-T6 aluminium weldments, 1: Reaction kinetics”, *Acta Metall. Mater.* **39**:11 (1991), 2693–2702.
- [Myhr and Grong 1991b] O. R. Myhr and Ø. Grong, “Process modelling applied to 6082-T6 aluminium weldments, 2: Applications of model”, *Acta Metall. Mater.* **39**:11 (1991), 2703–2708.
- [Nandan et al. 2007] R. Nandan, G. G. Roy, T. J. Lienert, and T. Debroy, “Three-dimensional heat and material flow during friction stir welding of mild steel”, *Acta Mater.* **55**:3 (2007), 883–895.
- [Papadopoulos and Solberg 1998] P. Papadopoulos and J. M. Solberg, “A Lagrange multiplier for the finite element solution of frictionless contact problems”, *Math. Comput. Model.* **28**:4–8 (1998), 373–384.
- [Papeleux and Ponthot 2002] L. Papeleux and J.-P. Ponthot, “Finite element simulation of springback in sheet metal forming”, *J. Mater. Process. Technol.* **125-126** (2002), 785–791.
- [Perić and Owen 1992] D. Perić and D. R. J. Owen, “Computational model for 3-D contact problems with friction based on the penalty method”, *Int. J. Numer. Methods Eng.* **35**:6 (1992), 1289–1309.
- [Rateau 2003] G. Rateau, *Méthode Arlequin pour les problèmes mécaniques multi-échelles: applications à des problèmes de jonction et de fissuration de structures élancées*, Ph.D. thesis, École Centrale Paris, 2003.
- [Rosenthal 1941] D. Rosenthal, “Mathematical theory of heat distribution during welding and cutting”, *Weld. J.* **20**:5 (1941), 220s–234s.
- [Rosenthal and Schmerber 1938] D. Rosenthal and R. Schmerber, “Thermal study of arc welding”, *Weld. J.* **17** (1938).
- [Schmidt and Hattel 2004] H. Schmidt and J. Hattel, “Modelling thermomechanical conditions at the tool/matrix interface in friction stir welding”, in *5th International Symposium on Friction Stir Welding* (Metz, 2004), TWI, Cambridge, 2004.
- [Schmidt and Hattel 2005] H. Schmidt and J. Hattel, “A local model for the thermomechanical conditions in friction stir welding”, *Model. Simul. Mater. Sci. Eng.* **13**:1 (2005), 77–93.
- [Seidel and Reynolds 2003] T. U. Seidel and A. P. Reynolds, “Two-dimensional friction stir welding process model based on fluid mechanics”, *Sci. Technol. Weld. Joi.* **8**:3 (2003), 175–183.

- [Shi et al. 2003] Q. Shi, T. Dickerson, and H. R. Shercliff, “Thermomechanical FE modelling of friction stir welding of Al-2024 including tool loads”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Simar et al. 2004] A. Simar, T. Pardoën, and B. de Meester, “Influence of friction stir welding parameters on the power input and temperature distribution in aluminium alloys”, in *5th International Symposium on Friction Stir Welding* (Metz, 2004), TWI, Cambridge, 2004.
- [Simo and Laursen 1992] J. C. Simo and T. A. Laursen, “An augmented Lagrangian treatment of contact problems involving friction”, *Comput. Struct.* **42**:1 (1992), 97–116.
- [Song and Kovacevic 2003a] M. Song and R. Kovacevic, “A coupled heat transfer model for workpiece and tool in friction stir welding”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Song and Kovacevic 2003b] M. Song and R. Kovacevic, “Thermal modeling of friction stir welding in a moving coordinate system and its validation”, *Int. J. Mach. Tool. Manuf.* **43**:6 (2003), 605–615.
- [Tartakovsky et al. 2006] A. M. Tartakovsky, G. J. Grant, X. Sun, and M. A. Khaleel, “Modeling of friction stir welding (FSW) process with smooth particle hydrodynamics (SPH)”, in *Welding & joining & fastening & friction stir welding* (Detroit, MI, 2006), Society of Automotive Engineers, Warrendale, PA, 2006. Paper # 2006-01-1394.
- [Ulysse 2002] P. Ulysse, “Three-dimensional modeling of the friction stir-welding process”, *Int. J. Mach. Tool. Manuf.* **42**:14 (2002), 1549–1557.
- [Wang et al. 2007] Z. W. Wang, S. Q. Zeng, X. H. Yang, and C. Cheng, “The key technology and realization of virtual ring rolling”, *J. Mater. Process. Technol.* **182**:1–3 (2007), 374–381.
- [Wriggers 1995] P. Wriggers, “Finite element algorithms for contact problems”, *Arch. Comput. Methods Eng.* **2**:4 (1995), 1–49.
- [Xiao and Belytschko 2004] S. P. Xiao and T. Belytschko, “A bridging domain method for coupling continua with molecular dynamics”, *Comput. Methods Appl. Mech. Eng.* **193**:17–20 (2004), 1645–1669.
- [Xu and Deng 2001] S. Xu and X. Deng, “Solid mechanics simulation of friction stir welding process”, pp. 631–638 in *Transactions of NAMRI/SME* (Gainesville, FL, 2001), vol. XXIX, Society of Manufacturing Engineers, Dearborn, MI, 2001.
- [Xu and Deng 2002] S. Xu and X. Deng, “A three-dimensional model for the friction-stir welding process”, pp. 699–704 in *Developments in theoretical and applied mechanics (SECTAM XXI)* (Orlando, FL, 2002), edited by A. Kassab and H. Mahfuz, Rivercross, Orlando, FL, 2002. Paper # 2108.
- [Xu and Deng 2003] S. Xu and X. Deng, “Two and three dimensional finite element models for the friction stir welding process”, in *4th International Symposium on Friction Stir Welding* (Park City, UT, 2003), TWI, Cambridge, 2003.
- [Zhang et al. 2005] H. W. Zhang, Z. Zhang, and J. T. Chen, “The finite element simulation of the friction stir welding process”, *Mater. Sci. Eng. A* **403**:1–2 (2005), 340–348.
- [Zhang et al. 2007] H. W. Zhang, Z. Zhang, and J. T. Chen, “3D modeling of material flow in friction stir welding under different process parameters”, *J. Mater. Process. Technol.* **183**:1 (2007), 62–70.

Received 17 Dec 2007. Revised 12 Dec 2008. Accepted 15 Dec 2008.

OLIVIER LORRAIN: olivier.lorrain@metz.ensam.fr

LPMM UMR CNRS 7554, Arts et Métiers ParisTech Metz, 4 Rue Augustin Fresnel, 57078 Metz, France

JÉRÔME SERRI: serri@lpm.science.univ-metz.fr

LRMA-LE2I/UFR Sciences & Techniques, Antenne d’Auxerre, Route des Plaines de l’Yonne, BP16, 89010 Auxerre, France

VÉRONIQUE FAVIER: veronique.favier@paris.ensam.fr

LIM UMR CNRS 8006, Arts et Métiers ParisTech Paris, 151 Boulevard de l’Hôpital, 75013 Paris, France

HAMID ZAHROUNI: zahrouni@univ-metz.fr

LPMM UMR CNRS 7554, Université de Metz, Ile du Saulcy, 57045 Metz, France

MOURAD EL HADROUZ: mourad.elhadrouz@metz.ensam.fr

LPMM UMR CNRS 7554, Arts et Métiers ParisTech Metz, 4 Rue Augustin Fresnel, 57078 Metz, France

ANISOTROPIC THIN-WALLED BEAM MODELS: A RATIONAL DEDUCTION FROM THREE-DIMENSIONAL ELASTICITY

FRANCO MACERI AND GIUSEPPE VAIRO

In this paper anisotropic thin-walled beam models are rationally deduced from three-dimensional elasticity by means of a constrained approach. Consistent frictionless internal constraints on both stress and strain dual fields are enforced through a modified Hu–Washizu functional obtained by a nonstandard application of Lagrange multipliers. Beam theories accounting for different shear refinement levels are justified, showing that this variational approach enables the development of new refined models, including high-order nonconventional effects and enhancing standard treatments of shear deformation effects. In agreement with the constrained problem, a locally equilibrated approximation of the stress field acting on beam cross-section is recovered in closed form. Finally, cases of laminated thin-walled beams as well as of unilateral conewise constitutive behavior (with special reference to bimodular materials) are investigated.

1. Introduction

Thin-walled beams comprising anisotropic materials, such as thin-walled elements made of composite fiber-reinforced polymer, are increasingly being used as primary and secondary members in structural applications in many areas, including aeronautical, mechanical, biomechanical, and civil engineering. The most common shapes include open cross-sections, such as I-, C-, L-, Z-shaped profiles, usually produced by pultrusion technology.

Because of the beam's specific geometry (two dimensions fairly smaller than the third) the analysis is generally carried out by means of approximate one-dimensional models. However, thin-walled members need treatments different from those adopted for classical rods because they can be very sensitive to shear deformation effects, which challenge both the validity of the classical Bernoulli–Navier assumption on cross-section deformation and Saint-Venant's pure torsion theory [Kollbrunner and Basler 1969].

The first one-dimensional thin-walled beam theory was developed by Vlasov [1961; 1962], who assumed that the section contour is invariant in its plane and that shear stresses in the middle surface of the beam vanish, and so reached a new distribution law of longitudinal stresses in the cross-section (law of sectorial areas). Vlasov's theory includes extension, bending and torsion deformations of isotropic homogeneous beams and describes the torsional shear stress flow as a superposition of two parts: the Saint-Venant primary flow (pure torsion) and a secondary one, associated to the shear stresses induced by the nonuniform warping of the cross-section. In spite of some inherent limitations, this theory represents a general framework in which many authors have developed more refined theoretical and numerical approaches for the analysis of thin-walled beams.

Keywords: thin-walled beams, constrained elasticity.

The static behavior of thin-walled beams has been analyzed for instance in [Bauld and Tzeng 1984; Ascione et al. 2000; De Lorenzis and La Tegola 2003; Erkmén and Mohareb 2006; Lee and Lee 2004; Lee 2005], where different shear refinement levels are included. Bauld and Tzeng [1984] developed a Vlasov-type theory for fiber-reinforced beams with thin-walled open cross-sections, which does not take into account any shear effect. Ascione et al. [2000], improving the kinematics of the classical Vlasov theory, included shear deformability using Timoshenko-type contributions enhanced through a Galerkin approach. De Lorenzis and La Tegola [2003] generalized the exact theory of thin-walled isotropic beams developed by Capurso [1964] to the case of transversely isotropic materials, including the effects of restraints and of concentrated loads. Erkmén and Mohareb [2006], postulating statically admissible stress fields in agreement with those coming out from the Vlasov's theory, proposed a theoretical and numerical torsional analysis for isotropic thin-walled beams including shear deformation effects on the middle surface. Lee [2005] has recently employed a thin-walled beam model for the analysis of laminated beams, based on a first-order shear-deformable beam theory (such as in [Maddur and Chaturvedi 1999]), accounting not only for shear deformation due to flexure but also for nonuniform warping effects (analogously to [Cortínez and Piovan 2002]).

Starting from generalized Vlasov-type theories, including shear effects as well as first- and second-order terms of rotational parameters, stability of composite thin-walled profiles has been analyzed in [Cortínez and Piovan 2002; Fraternali and Feo 2000; Lee and Kim 2001; Saadé et al. 2004; Piovan and Cortínez 2007]. Modified Vlasov-type theories have also been proposed in [Maddur and Chaturvedi 1999; Cortínez and Piovan 2002; Piovan and Cortínez 2005; 2007; Ambrosini et al. 2000] and elsewhere for the theoretical and numerical analysis of the dynamical behavior of such structures.

Nevertheless, the rational deduction and justification of these theories from three-dimensional elasticity and their consistent generalization for anisotropic materials as well as for nonconventional cases (such as laminated beams or unilateral material behavior) can be truly considered as an open task yet. Note that the deduction of thin-walled beam models in a consistent way is not only a speculative issue, but leads to a safer and more complete technical use of these theories.

In the specialized literature, the rational deduction of structural theories is performed mainly through two strategies: the asymptotic method and the constrained approach. The first, started in the mid-seventies through the influence of several works addressing theories of plates and shells, was later applied to rods and thin-walled beams [Trabucho and Viaño 1996; Rodríguez and Viaño 1997; Volovoi et al. 1999; Hamdouni and Millet 2006]. Its main idea is that the three-dimensional solution of the elasticity equations can be approximated through successive terms of a power series, where, for beams, the slenderness ratio (between cross-section diameter and beam length) is taken as a small parameter. Accordingly, under suitable hypotheses that ensure series convergence, different structural theories can be rationally deduced as approximate solutions of an exactly stated problem, varying the series truncation order. The constrained approach, by contrast, is based on a diametrically opposite concept: it looks for an exact solution of a simplified constrained problem, based on approximate representations of the unknown functions. The primary three-dimensional elastostatic problem is reduced to a consistent simplified one (two-dimensional in the case of plates and shells, one-dimensional for beams), enforcing suitable assumptions on strain and/or stress fields as internal frictionless constraints. That approach was successfully employed for deducing classical plate and shell theories [Podio-Guidugli 1989; Bisegna and Sacco 1997b] and theories of beams with solid sections [Lembo and Podio-Guidugli 2001; Maceri

and Bisegna 2002]. To justify Kirchhoff–Love plate theory [Podio-Guidugli 1989] and the Timoshenko beam model [Lembo and Podio-Guidugli 2001], a constrained approach was proposed based on strain assumptions and on the concept of constrained material. By contrast, the constrained approach proposed by Bisegna and Sacco [1997b] for deducing classical plate theories maintains the constitutive law given a priori, and involves consistent assumptions on both strain and stress dual fields. Nevertheless, while strain assumptions can be easily identified if the problem is characterized by special geometries, effective and consistent stress assumptions can be sometimes not obvious. In order to overcome this difficulty, Maceri and Bisegna [2002] showed how plates, shells and planar beams theories can be justified enforcing in a consistent way the same constraints on both stress and strain dual fields.

In this paper, this dual-constraint approach is employed for justifying several anisotropic thin-walled beam models, accounting for different levels of shear refinements. To this aim, the Hu–Washizu variational formulation of the three-dimensional elasticity problem is modified using Lagrange multiplier theory [Antman and Marlow 1991]. Because of the simultaneous presence of constraints on dual spaces, a nonstandard application of Lagrange multipliers is required.

As far as homogeneous beams are considered, an anisotropic model which does not account for any shear deformations is deduced first, justifying the model of Bauld and Tzeng [1984] and reducing to the classical Vlasov’s theory when isotropic constitutive symmetry is considered. The proposed approach enables to take into account also nonconventional high-order effects, related to thickness and curvature of the cross-section centerline. In detail, generalized expressions for warping function and torsion constant are obtained, whose influence is investigated through simple numerical applications. Afterwards, a first-order shear-deformable thin-walled beam model is rationally derived, justifying models involving Timoshenko’s shear effects. This is the case of the model proposed in [Ascione et al. 2000] when Galerkin-type contributions are neglected, and, for homogeneous beams, of the model analyzed in [Maddur and Chaturvedi 1999]. In the framework of the dual-constraint approach, different shear refinement levels can be also accounted for. This is proved including shear effects due to nonuniform warping (e.g., models in [Lee 2005; Cortínez and Piovan 2002]) and drawing some strategies to deduce a new branchwise model, for multibranch beam cross-sections, including branch-depending Timoshenko’s and warping shear effects.

A more accurate evaluation of stresses on beam cross-section with respect to those obtained by constitutive relationship represents an issue of great technical interest when, for instance, damage of composite thin-walled beams is addressed. To this aim, in agreement with the constrained problem, approximated locally equilibrated stresses acting on beam cross-section are recovered in closed form.

Finally, the dual-constraint approach is successfully employed to rationally deduce models for the analysis of laminated thin-walled beams as well as of homogeneous beams comprising nonlinear cone-wise elastic materials [Curnier et al. 1995] (with reference to bimodular ones).

2. Dual-constraint approach

In the framework of infinitesimal deformation theory, the equilibrium problem of a three-dimensional body Ω comprising a linearly elastic material can be recast by adopting the Hu–Washizu variational formulation:

Find the displacement field \mathbf{u} , the strain field $\boldsymbol{\varepsilon}$ and the stress field $\boldsymbol{\sigma}$ that make stationary the functional

$$\begin{aligned} \mathcal{W}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}) = & \int_{\Omega} \frac{1}{2} \mathcal{C} \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon} \, dv - \int_{\Omega} \boldsymbol{\sigma} \cdot \boldsymbol{\varepsilon} \, dv + \int_{\Omega} \boldsymbol{\sigma} \cdot \hat{\nabla} \mathbf{u} - \int_{\Omega} \mathbf{b} \cdot \mathbf{u} \, dv \, dv \\ & - \int_{\partial_f \Omega} \mathbf{p} \cdot \mathbf{u} \, da - \int_{\partial_u \Omega} \boldsymbol{\sigma} \mathbf{n}_{\partial} \cdot (\mathbf{u} - \mathbf{u}_0) \, da, \end{aligned} \quad (2-1)$$

where $\hat{\nabla}$ denotes the symmetrical part of the gradient operator, \cdot the inner product, \mathbf{b} the volume forces, \mathbf{p} the surface forces on $\partial_f \Omega$, \mathbf{u}_0 the displacement assigned on $\partial_u \Omega = \partial \Omega \setminus \partial_f \Omega$, \mathbf{n}_{∂} the outward normal unit vector to $\partial \Omega$, and \mathcal{C} the fourth-order elasticity tensor satisfying major and minor symmetries. The stationary conditions of \mathcal{W} with respect to \mathbf{u} , $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ yield equilibrium, compatibility and constitutive equations governing the three-dimensional elastostatic problem for Ω .

When the body is characterized by special geometrical aspect ratios, the three-dimensional problem can be approximated using suitable assumptions on strain and/or stress fields. If these assumptions are regarded as internal frictionless constraints, reactive fields arise and the original difficult three-dimensional elastic problem can be replaced by a constrained problem that can be often solved more easily.

In order to enforce constraints on both strain and stress dual fields, the constrained equilibrium problem can be suitably formulated employing Lagrange multipliers. Physically, Lagrange multipliers represent reactive actions belonging to the dual space of the one where the constrained variable lives: a Lagrange multiplier of a strain constraint identifies a reactive stress, and conversely a multiplier of a stress constraint has the meaning of reactive strain. The consistent representation of such reactive fields arises as a consequence of the enforced constraints and it is not postulated a priori.

As in [Bisegna and Sacco 1997b], we will adopt the following definitions: the total strain field is the symmetrical part of the gradient of the displacement field; the total stress field satisfies the equilibrium equations; elastic stresses and strains are related to each other by the elastic constitutive law; the total stress (or strain) is sum of its elastic and reactive parts.

In order to build up structural theories, one usually imposes a representation law for the displacement field (which is equivalent to imposing constraints on the total strains), and some hypotheses on the stress field at the constitutive law level (i.e., on the elastic stress field).

Let the total strain and the elastic stress fields be constrained to belong to the kernel of the linear (possibly differential) operators \mathbf{G} and \mathbf{H} , respectively. Following [Bisegna and Sacco 1997b], these constraints acting on dual spaces can be enforced by introducing the Lagrangian functional

$$\mathcal{L}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{\chi}, \boldsymbol{\omega}) = \mathcal{W}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}) - \int_{\Omega} \boldsymbol{\chi} \cdot \mathbf{G} \boldsymbol{\varepsilon} \, dv - \int_{\Omega} \boldsymbol{\omega} \cdot \mathbf{H} \boldsymbol{\sigma} \, dv - \int_{\Omega} \mathbf{G}^* \boldsymbol{\chi} \cdot \mathbf{H}^* \boldsymbol{\omega} \, dv, \quad (2-2)$$

where the vectors $\boldsymbol{\chi}$, $\boldsymbol{\omega}$ are Lagrange multipliers, \mathbf{G}^* and \mathbf{H}^* denote the adjoint operators of \mathbf{G} and \mathbf{H} , respectively. The stationary condition of \mathcal{L} with respect to \mathbf{u} yields the equilibrium equations

$$\begin{aligned} \operatorname{div} \boldsymbol{\sigma} + \mathbf{b} &= \mathbf{0} \quad \text{in } \Omega, \\ \boldsymbol{\sigma} \mathbf{n}_{\partial} &= \mathbf{p} \quad \text{on } \partial_f \Omega; \end{aligned} \quad (2-3)$$

the one with respect to $\boldsymbol{\varepsilon}$ yields the constitutive equations

$$\boldsymbol{\sigma} + \mathbf{G}^* \boldsymbol{\chi} = \mathcal{C} \boldsymbol{\varepsilon} \quad \text{in } \Omega; \quad (2-4)$$

the one with respect to $\boldsymbol{\sigma}$ yields the compatibility equations

$$\begin{aligned}\boldsymbol{\varepsilon} + \mathbf{H}^* \boldsymbol{\omega} &= \hat{\nabla} \mathbf{u} \quad \text{in } \Omega, \\ \mathbf{u} &= \mathbf{u}_0 \quad \text{on } \partial_u \Omega;\end{aligned}\tag{2-5}$$

and finally the stationary conditions with respect to the Lagrange multipliers $\boldsymbol{\chi}$ and $\boldsymbol{\omega}$ yield, respectively, the constraint equations

$$\begin{aligned}\mathbf{G}(\boldsymbol{\varepsilon} + \mathbf{H}^* \boldsymbol{\omega}) &= \mathbf{0} \quad \text{in } \Omega, \\ \mathbf{H}(\boldsymbol{\sigma} + \mathbf{G}^* \boldsymbol{\chi}) &= \mathbf{0} \quad \text{in } \Omega.\end{aligned}\tag{2-6}$$

Accordingly, by (2-3), (2-4) and (2-5), $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon} + \mathbf{H}^* \boldsymbol{\omega}$ turn out to be the total stress and strain fields, respectively, and $\boldsymbol{\sigma} + \mathbf{G}^* \boldsymbol{\chi}$ and $\boldsymbol{\varepsilon}$ the elastic stress and strain fields, respectively. As a consequence, reactive stress and strain fields are $-\mathbf{G}^* \boldsymbol{\chi}$ and $\mathbf{H}^* \boldsymbol{\omega}$, respectively. Note that the reactive stress field is orthogonal to every admissible total strain field and likewise the reactive strain field is orthogonal to every admissible elastic stress field, as it appears from (2-6).

In this dual-constraint framework, Maceri and Bisegna [2002] have shown that classical plates and planar beams theories can be rationally deduced by assuming that the dual constraints on the elastic stress field are the same as those imposed on the total strain field. In this case, operators \mathbf{G} and \mathbf{H} are such that $\mathbf{H}\mathbf{A} = \mathbf{G}\mathbf{A}$ for every symmetrical second order tensor \mathbf{A} and Lagrange multipliers $\boldsymbol{\chi}$ and $\boldsymbol{\omega}$ belong to dual vector subspaces characterized by the same dimensions. In this way, once kinematic constraints are chosen, consistent stress assumptions directly arise.

For what follows, it is more useful transforming the functional \mathcal{L} of (2-2) into a potential energy functional, by enforcing a priori satisfied stationary conditions of \mathcal{L} with respect to $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$, that is (2-5) and (2-4). Accordingly, in the framework of the dual constraint approach proposed in [Maceri and Bisegna 2002], the functional \mathcal{L} becomes

$$\mathcal{E}(\mathbf{u}, \boldsymbol{\chi}, \boldsymbol{\omega}) = \int_{\Omega} \frac{1}{2} \mathcal{C}(\hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega}) \cdot (\hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega}) dv - \int_{\Omega} \boldsymbol{\chi} \cdot \mathbf{G} \hat{\nabla} \mathbf{u} dv - \int_{\Omega} \mathbf{b} \cdot \mathbf{u} dv - \int_{\partial_f \Omega} \mathbf{p} \cdot \mathbf{u} da, \tag{2-7}$$

defined on the manifold $\mathbf{u} = \mathbf{u}_0$ on $\partial_u \Omega$. It can be verified that total strain and elastic strain are $\boldsymbol{\Lambda} = \hat{\nabla} \mathbf{u}$ and $\boldsymbol{\varepsilon} = \hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega}$, respectively, and total stress and elastic stress are $\boldsymbol{\sigma} = \mathcal{C}(\hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega}) - \mathbf{G}^* \boldsymbol{\chi}$ and $\boldsymbol{\sigma}^{(el)} = \mathcal{C}(\hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega})$, respectively.

It should be noted that \mathcal{E} depends on the reactive fields. In order to obtain a potential-energy functional that does not depend on the Lagrange multipliers it is sufficient to make stationary conditions of \mathcal{E} with respect to $\boldsymbol{\chi}$ and $\boldsymbol{\omega}$ a priori satisfied.

3. Thin-walled beam models

3A. Notation. Consider a right cylinder beam-like body $\Omega = \mathcal{P} \times]-L, L[$, whose length is $2L$ and whose cross-section is an open thin-walled profile \mathcal{P} , assumed to be constant along the beam axis-line. Let the global Cartesian frame (O, x, y, z) be assumed with $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ the corresponding orthonormal basis and with z -axis parallel to the cylinder axis. Let p the centerline of \mathcal{P} and $2\delta \ll 2L$ the thickness of the wall. Hence, the middle surface of the beam is $M = p \times]-L, L[$ and, with a little abuse of notation, the cross-section can be described as $\mathcal{P} = p \times]-\delta, \delta[$. The beam is assumed to be in equilibrium when

the thickness is

$$\bar{f}(s, z) = \frac{1}{2\delta} \int_{-\delta}^{\delta} f(s, \eta, z) j(s, \eta) d\eta. \quad (3-5)$$

Moreover, it is easy to verify that, taking into account (3-1), the following identity holds:

$$\nabla f = \frac{f_{/1}}{j} \mathbf{t} + f_{/2} \mathbf{n} + f' \mathbf{k}. \quad (3-6)$$

Here and in what follows we indicate the partial derivatives of a function f with respect to local coordinates s and η as $f_{/1}$ and $f_{/2}$, respectively, whereas the partial derivative with respect to z is denoted by f' . Where necessary, vector and tensor components are denoted by subscripts and Einstein's summation convention is used, in which case Greek indices imply values in $\{1, 2\}$ and denote components in the plane of \mathcal{P} referred to the local tangent frame (1 standing for the tangent-to- p component, i.e., along s , and 2 for the normal-to- p component, i.e., along η), the index 3 indicates components along the z -axis, and components along in-plane Cartesian axes are explicitly denoted by the subscripts x and y .

It is useful to introduce the following in-plane shear strain and stress vectors:

$$\begin{aligned} \boldsymbol{\gamma} &= 2\varepsilon_{13} \mathbf{t} + 2\varepsilon_{23} \mathbf{n} = 2\varepsilon_{x3} \mathbf{i} + 2\varepsilon_{y3} \mathbf{j}, \\ \boldsymbol{\tau} &= \sigma_{13} \mathbf{t} + \sigma_{23} \mathbf{n} = \sigma_{x3} \mathbf{i} + \sigma_{y3} \mathbf{j} \end{aligned} \quad (3-7)$$

for both elastic and total fields.

Finally, in this section the beam is assumed to be homogeneous and comprising a linearly elastic material having at least a monoclinic symmetry, with symmetry plane orthogonal to \mathbf{k} . Accordingly, $\mathcal{C}_{\alpha\beta\gamma 3} = \mathcal{C}_{\alpha 333} = 0$. This material symmetry includes, for instance, the case of thin-walled profiles pultruded along the z -direction.

3B. No-shear beam model. The total strain field is assumed to satisfy Vlasov's assumptions [1961; 1962] on M and the Euler–Bernoulli constraints of classical beam theories:

- (i) The in-plane (dilatation and shear) total strain components vanish everywhere on \mathcal{P} .
- (ii) The shear total strain between the z -axis and $\mathbf{n}(s)$ vanishes everywhere on \mathcal{P} .
- (iii) The flux through the thickness of the in-plane shear total strain vector (see (3-7)) is zero.

Accordingly, the operator \mathbf{G} is such that

$$\mathbf{G}\boldsymbol{\varepsilon} = \{\varepsilon_{11} \ \varepsilon_{22} \ \varepsilon_{12} \ \bar{\varepsilon}_{13} \ \varepsilon_{23}\}^T. \quad (3-8)$$

Following the constrained approach proposed in [Maceri and Bisegna 2002], the dual constraints on the elastic stress field can be directly expressed as follows:

- (iv) The elastic stress vector on every plane parallel to the z -axis is parallel to \mathbf{k} .
- (v) The shear elastic stress along $\mathbf{n}(s)$ vanishes everywhere on \mathcal{P} .
- (vi) The flux through the thickness of the in-plane shear elastic stress vector is zero.

Accordingly, the functional (2-7) can be written as

$$\mathfrak{E}(\mathbf{u}, \boldsymbol{\chi}, \boldsymbol{\omega}) = \frac{1}{2} \int_{\Omega} [\mathfrak{C}_{\alpha\beta\gamma\delta}(\Lambda_{\alpha\beta} - \omega_{\alpha\beta})(\Lambda_{\gamma\delta} - \omega_{\gamma\delta}) + 2\mathfrak{C}_{\alpha\beta 33}(\Lambda_{\alpha\beta} - \omega_{\alpha\beta})u'_3 + \mathfrak{C}_{3333}(u'_3)^2 + 4\mathfrak{C}_{\alpha 3\beta 3}(\Lambda_{\alpha 3} - \omega_{\alpha 3})(\Lambda_{\beta 3} - \omega_{\beta 3})] dv - \int_{\Omega} (\chi_{\alpha\beta}\Lambda_{\alpha\beta} + 2\chi_{\alpha 3}\Lambda_{\alpha 3}) dv - \Pi_{\text{ext}}, \quad (3-9)$$

where $\boldsymbol{\Lambda} = \hat{\nabla}\mathbf{u}$, whose components in the local tangent frame are evaluated through (3-6), and Π_{ext} accounts for external loads:

$$\Pi_{\text{ext}} = \int_{\Omega} (b_{\alpha}u_{\alpha} + b_3u_3) dv + \int_{\Sigma} (\hat{p}_{\alpha}u_{\alpha} + \hat{p}_3u_3) dQ dz + \int_{\mathcal{P}} (p_{\alpha}^{\pm}u_{\alpha}|_{\pm L} + p_3^{\pm}u_3|_{\pm L}) da, \quad (3-10)$$

where dQ is the arc element along $\partial\mathcal{P}$ and the notation $f^{\pm}(\cdot)|_{\pm L}$ means $f^{+}(\cdot)|_{z=L} + f^{-}(\cdot)|_{z=-L}$.

Note that, since constraints (iii) and (vi) are imposed by average condition (3-5), relevant Lagrange multipliers are defined on M , i.e., $\omega_{13} = \omega_{13}(s, z)$, $\chi_{13} = \chi_{13}(s, z)$.

The stationary conditions of the functional (3-9) with respect to $\chi_{\alpha\beta}$ and $\chi_{\alpha 3}$ give, respectively, the following constraints on the displacement field, i.e., on the total strain field:

$$\begin{aligned} \Lambda_{11} &= [u_{1/1} - \kappa u_2]/j = 0, & 4\delta\bar{\Lambda}_{13} &= \int_{-\delta}^{\delta} (u_{3/1} + ju'_1) d\eta = 0, \\ \Lambda_{22} &= u_{2/2} = 0, & & \\ 2\Lambda_{12} &= [u_{2/1} + \kappa u_1]/j + u_{1/2} = 0, & 2\Lambda_{23} &= u'_2 + u_{3/2} = 0, \end{aligned} \quad (3-11)$$

where $u_1 = \mathbf{u} \cdot \mathbf{t}$, $u_2 = \mathbf{u} \cdot \mathbf{n}$ and $u_3 = \mathbf{u} \cdot \mathbf{k}$.

Integration of (3-11) yields the Cartesian representation formulas for the displacement field:

$$\begin{aligned} u_x(s, \eta, z) &= \mathbf{u} \cdot \mathbf{i} = u_c(z) - \theta(z)[y(s, \eta) - y_c], \\ u_y(s, \eta, z) &= \mathbf{u} \cdot \mathbf{j} = v_c(z) + \theta(z)[x(s, \eta) - x_c], \\ u_3(s, \eta, z) &= \mathbf{u} \cdot \mathbf{k} = w_c(z) - u'_c(z)x(s, \eta) - v'_c(z)y(s, \eta) + \theta'(z)\psi_c(s, \eta) \end{aligned} \quad (3-12)$$

where x_c, y_c are the coordinates of an arbitrary point C in the plane $x - y$, assumed as the pole of the in-plane rigid rotation $\theta(z)$ of \mathcal{P} ; $u_c(z), v_c(z)$ and $w_c(z)$ are the components of the displacement of C along Cartesian axes; $\psi_c(s, \eta)$ is the warping function referred to C and defined as

$$\psi_c(s, \eta) = \psi_0 + \int_0^s \mathbf{r} \cdot \mathbf{n} ds - \eta \mathbf{r} \cdot \mathbf{t} - \frac{1}{3} \int_0^s \delta^2 \kappa ds, \quad (3-13)$$

where $\mathbf{r} = [\tilde{x}(s) - x_c]\mathbf{i} + [\tilde{y}(s) - y_c]\mathbf{j}$ is the position vector of a point $\tilde{\mathbf{x}}$ on p with respect to C . If $\int_{\mathcal{P}} \psi_c x da = \int_{\mathcal{P}} \psi_c y da = 0$ then C coincides with the twist-center of \mathcal{P} . Moreover, the constant ψ_0 can be evaluated by equating to zero the average value of ψ_c on \mathcal{P} , i.e., considering the point S coincident to the sectorial centroid of p . The second and third term in (3-13) represent the middle-contour warping (i.e., the primary warping) and the thickness warping (i.e., the secondary warping), respectively. Furthermore, the last contribution takes into account high-order effects relating to thickness and curvature of p , generalizing the warping expression discussed by Lin and Hsiao [2003] and widely used in the specialized literature; see for example [Lee and Lee 2004; Lee 2005; Cortínez and Piovan 2002; Saadé et al. 2004; Piovan and Cortínez 2005; 2007].

The stationary conditions of the functional \mathfrak{E} with respect to $\omega_{\alpha\beta}$ and $\omega_{\alpha 3}$ give, respectively,

$$\begin{aligned} \mathcal{C}_{\alpha\beta\gamma\delta}[\Lambda_{\gamma\delta} - \omega_{\gamma\delta}] + \mathcal{C}_{\alpha\beta 33}u'_3 &= 0, \\ \int_{-\delta}^{\delta} \mathcal{C}_{\alpha 313}[\Lambda_{\alpha 3} - \omega_{\alpha 3}] j(s, \eta) d\eta &= 0, \\ \mathcal{C}_{\alpha 323}[\Lambda_{\alpha 3} - \omega_{\alpha 3}] &= 0. \end{aligned} \quad (3-14)$$

Taking into account (3-11), the Lagrange multipliers $\omega_{\alpha\beta}$ and $\omega_{\alpha 3}$ are uniquely determined by (3-14):

$$\omega_{\alpha\beta} = (\mathcal{C}_{\alpha\beta\gamma\delta})^{-1} \mathcal{C}_{\gamma\delta 33}u'_3, \quad \omega_{13} = 0, \quad \omega_{23} = \mathcal{C}_{1323}\Lambda_{13}/\mathcal{C}_{2323}, \quad (3-15)$$

where $(\mathcal{C}_{\alpha\beta\gamma\delta})^{-1}$ denotes the inverse tensor of $\mathcal{C}_{\alpha\beta\gamma\delta}$.

Substituting (3-12) and (3-15) into the functional (3-9) and performing integration over \mathcal{P} , the potential energy functional for the thin-walled beam in terms of pure displacement unknowns can be written as

$$\hat{\mathcal{E}}(u_c, v_c, w_c, \theta) = \frac{1}{2} \int_{-L}^L \mathbf{D}\mathbf{e} \cdot \mathbf{e} dz - \int_{-L}^L \mathbf{q} \cdot \hat{\mathbf{s}} dz - \mathbf{Q}^{\pm} \cdot \hat{\mathbf{s}}|_{\pm L}, \quad (3-16)$$

where the vector

$$\hat{\mathbf{s}} = \{u_c \ v_c \ w_c \ -v'_c \ u'_c \ \theta \ \theta'\}^T \quad (3-17)$$

collects the generalized displacements,

$$\mathbf{e} = \{w'_c \ -v''_c \ u''_c \ \theta'' \ \theta'\}^T \quad (3-18)$$

denotes the generalized total strain,

$$\mathbf{q} = \{q_x \ q_y \ q_z \ m_x \ m_y \ m_z \ m_{\psi}\}^T, \quad \mathbf{Q}^{\pm} = \{Q_x^{\pm} \ Q_y^{\pm} \ Q_z^{\pm} \ M_x^{\pm} \ M_y^{\pm} \ M_z^{\pm} \ M_{\psi}^{\pm}\}^T \quad (3-19)$$

indicate, respectively, the generalized distributed and end-located forces acting on the beam, with

$$\{q_x, q_y, q_z\} = \int_{\mathcal{P}} \{b_x, b_y, b_z\} da + \int_{\partial\mathcal{P}} \{\hat{p}_x, \hat{p}_y, \hat{p}_z\} d\varrho, \quad (3-20)$$

$$\{m_x, m_y, m_{\psi}\} = \int_{\mathcal{P}} \{y, -x, \psi_c\} b_z da + \int_{\partial\mathcal{P}} \{y, -x, \psi_c\} \hat{p}_z d\varrho, \quad (3-21)$$

$$m_z = \int_{\mathcal{P}} [b_y(x - x_c) - b_x(y - y_c)] da + \int_{\partial\mathcal{P}} [\hat{p}_y(x - x_c) - \hat{p}_x(y - y_c)] d\varrho, \quad (3-22)$$

$$\{Q_x^{\pm}, Q_y^{\pm}, Q_z^{\pm}\} = \int_{\mathcal{P}} \{p_x^{\pm}, p_y^{\pm}, p_z^{\pm}\} da, \quad (3-23)$$

$$\{M_x^{\pm}, M_y^{\pm}, M_{\psi}^{\pm}\} = \int_{\mathcal{P}} \{y, -x, \psi_c\} p_z^{\pm} da, \quad M_z^{\pm} = \int_{\mathcal{P}} [p_y^{\pm}(x - x_c) - p_x^{\pm}(y - y_c)] da, \quad (3-24)$$

and finally

$$\mathbf{D} = \begin{bmatrix} \hat{\mathcal{C}}_{3333}A & \hat{\mathcal{C}}_{3333}S_x & \hat{\mathcal{C}}_{3333}S_y & \hat{\mathcal{C}}_{3333}S_{\psi} & 0 \\ & \hat{\mathcal{C}}_{3333}I_x & \hat{\mathcal{C}}_{3333}I_{xy} & \hat{\mathcal{C}}_{3333}I_{x\psi} & 0 \\ & & \hat{\mathcal{C}}_{3333}I_y & \hat{\mathcal{C}}_{3333}I_{y\psi} & 0 \\ & sym & & \hat{\mathcal{C}}_{3333}I_{\psi} & 0 \\ & & & & \hat{\mathcal{C}}_{1313}J_{\theta} \end{bmatrix} \quad (3-25)$$

is the generalized elasticity matrix, with

$$\{A, S_x, S_y, S_\psi\} = \int_{\mathcal{P}} \{1, y, -x, \psi_c\} da, \quad (3-26)$$

$$\{I_x, I_{xy}, I_y, I_{x\psi}, I_{y\psi}, I_\psi\} = \int_{\mathcal{P}} \{y^2, -xy, x^2, y\psi_c, -x\psi_c, \psi_c^2\} da, \quad (3-27)$$

$$J_\theta = \int_{\mathcal{P}} [\kappa(\eta^2 - \delta^2/3) - 2\eta]^2/j^2 da, \quad (3-28)$$

where J_θ is the torsion constant, which generalizes the one commonly used in classical thin-walled open cross-section beam models, through nonconventional high-order terms depending on curvature of p and thickness.

Note that the coefficients of \mathbf{D} depend on the so-called reduced elastic moduli:

$$\hat{\mathcal{C}}_{1313} = \mathcal{C}_{1313} - \frac{\mathcal{C}_{1323}^2}{\mathcal{C}_{2323}}, \quad \hat{\mathcal{C}}_{3333} = \mathcal{C}_{3333} - \mathcal{C}_{\alpha\beta 33}(\mathcal{C}_{\alpha\beta\gamma\delta})^{-1}\mathcal{C}_{\gamma\delta 33}. \quad (3-29)$$

As is customary with beam theories, the stress resultants on \mathcal{P} (generalized stresses) are introduced:

$$\mathbf{S} = \{N \ C_x \ C_y \ C_\psi \ C_z\}^T = \mathbf{D}\mathbf{e}, \quad (3-30)$$

where \mathbf{S} is the generalized stress vector, defined in terms of normal force N , bending moments C_x and C_y , (primary) twisting moment C_z and bimoment (warping torque) C_ψ :

$$\{N, C_x, C_y, C_\psi\} = \int_{\mathcal{P}} \{1, y, -x, \psi_c\} \sigma_{33} da, \quad C_z = \int_{\mathcal{P}} (\sigma_{y3x} - \sigma_{x3y}) da. \quad (3-31)$$

Assuming that the global Cartesian frame is centered at the centroid of the beam Ω and that its x and y axes are principal for \mathcal{P} , referring ψ_c to the twisting center for \mathcal{P} and taking S coincident with the sectorial centroid of p , the matrix \mathbf{D} reduces to a diagonal matrix and the functional $\hat{\mathcal{E}}$ of (3-16) can be written as

$$\hat{\mathcal{E}}(u_c, v_c, w_c, \theta) = \mathcal{A}(w_c) + \mathcal{F}_x(v_c) + \mathcal{F}_y(u_c) + \mathcal{T}(\theta), \quad (3-32)$$

where

$$\begin{aligned} \mathcal{A}(w_c) &= \frac{1}{2}\hat{\mathcal{C}}_{3333}A \int_{-L}^L (w'_c)^2 dz - \int_{-L}^L q_z w_c dz - Q_z^\pm w_c|_{\pm L}, \\ \mathcal{F}_x(v_c) &= \frac{1}{2}\hat{\mathcal{C}}_{3333}I_x \int_{-L}^L (v''_c)^2 dz - \int_{-L}^L (q_y v_c - m_x v'_c) dz - (Q_y^\pm v_c|_{\pm L} - M_x^\pm v'_c|_{\pm L}), \\ \mathcal{F}_y(u_c) &= \frac{1}{2}\hat{\mathcal{C}}_{3333}I_y \int_{-L}^L (u''_c)^2 dz - \int_{-L}^L (q_x u_c + m_y u'_c) dz - (Q_x^\pm u_c|_{\pm L} + M_y^\pm u'_c|_{\pm L}), \\ \mathcal{T}(\theta) &= \frac{1}{2}\hat{\mathcal{C}}_{3333}I_\psi \int_{-L}^L (\theta'')^2 dz + \frac{1}{2}\hat{\mathcal{C}}_{1313}J_\theta \int_{-L}^L (\theta')^2 dz - \int_{-L}^L (m_z \theta + m_\psi \theta') dz \\ &\quad - (M_z^\pm \theta|_{\pm L} + M_\psi^\pm \theta'|_{\pm L}). \end{aligned} \quad (3-33)$$

The functional \mathcal{A} models the extensional problem of the beam, \mathcal{F}_x and \mathcal{F}_y govern the Euler–Bernoulli flexural model of the beam, \mathcal{T} is relevant to the torsional problem. Observe that these functionals are uncoupled and therefore the corresponding governing equations for extensional, flexural and torsional problems are themselves uncoupled.

The governing equations and natural boundary conditions for the extensional problem of the beam arise from the stationary conditions of \mathcal{A} with respect to w_c :

$$\hat{\mathcal{C}}_{3333} A w_c'' + q_z = 0, \quad \hat{\mathcal{C}}_{3333} A w_c'|_{-L}^L = Q_z^\pm, \quad (3-34)$$

where $f|_{-L}^L$ indicates $f(-L)$ for $z = -L$ or $f(L)$ for $z = L$.

Analogously, the stationary conditions of \mathcal{F}_x and \mathcal{F}_y give governing equations and boundary conditions for flexural problems in the yz and xz planes, respectively:

$$\hat{\mathcal{C}}_{3333} I_x v_c'''' - q_y - m_x' = 0, \quad \begin{cases} \hat{\mathcal{C}}_{3333} I_x v_c'''|_{-L}^L - m_x|_{-L}^L = -Q_y^\pm, \\ \hat{\mathcal{C}}_{3333} I_x v_c''|_{-L}^L = -M_x^\pm, \end{cases} \quad (3-35)$$

$$\hat{\mathcal{C}}_{3333} I_y u_c'''' - q_x + m_y' = 0, \quad \begin{cases} \hat{\mathcal{C}}_{3333} I_y u_c'''|_{-L}^L + m_y|_{-L}^L = -Q_x^\pm, \\ \hat{\mathcal{C}}_{3333} I_y u_c''|_{-L}^L = M_y^\pm. \end{cases} \quad (3-36)$$

Finally, the stationary conditions of the functional \mathcal{T} yield the governing equations and boundary conditions for the torsional problem:

$$\hat{\mathcal{C}}_{3333} I_\psi \theta'''' - \hat{\mathcal{C}}_{1313} J_\theta \theta'' - m_z + m_\psi' = 0, \quad \begin{cases} [\hat{\mathcal{C}}_{3333} I_\psi \theta''' - \hat{\mathcal{C}}_{1313} J_\theta \theta' + m_\psi]|_{-L}^L = -M_z^\pm, \\ \hat{\mathcal{C}}_{3333} I_\psi \theta''|_{-L}^L = M_\psi^\pm. \end{cases} \quad (3-37)$$

Solution of (3-34) to (3-37) gives the unknown functions u_c, v_c, w_c, θ from which total, reactive and elastic strain fields can be computed. Therefore, the elastic stress field $\boldsymbol{\sigma}^{(el)} = \mathcal{C}(\hat{\nabla}\mathbf{u} - \mathbf{G}^*\boldsymbol{\omega})$ referred to the local tangent frame turns out to be

$$\begin{aligned} \sigma_{\alpha\beta}^{(el)} &= \sigma_{23}^{(el)} = 0, \\ \sigma_{13}^{(el)} &= \hat{\mathcal{C}}_{1313} \theta' [\kappa(\eta^2 - \delta^2/3) - 2\eta]/j, \\ \sigma_{33}^{(el)} &= \hat{\mathcal{C}}_{3333} (w_c' - u_c''x - v_c''y + \theta''\psi_c). \end{aligned} \quad (3-38)$$

The total stress field coincides with the elastic one only for the component σ_{33} , whereas total stress components $\sigma_{\alpha\beta}$ and $\sigma_{\alpha 3}$ can be not uniquely recovered in a general case from the equilibrium equations (2-3). Nevertheless, (3-30) furnish stress resultants on \mathcal{P} which satisfy global equilibrium. Moreover, as will be showed in the following, a more accurate evaluation of total shear stress components on \mathcal{P} may be recovered from the field equilibrium equation along z .

We emphasize that the so-called reduced constitutive law enters the expression of $\hat{\mathcal{C}}$ and the generalized constitutive law (3-30) by means of the reduced elastic moduli (3-29). The present derivation clearly shows how the appearance of these quantities is a straightforward and rational consequence of constraints on dual fields, without contradiction because they act on fields (total strain and elastic stress) which are not related by constitutive law. In other words, the reduced constitutive law comes out from the procedure adopted and is not a priori enforced by means of a constrained constitutive law.

Note also that when constraints on the total strain field only (i.e., without constraints (iv), (v) and (vi)) are enforced, the functional $\hat{\mathcal{C}}$ contains nonreduced elastic moduli, whereas when condition (iv) is added to total strain assumptions, $\hat{\mathcal{C}}_{3333}$ appears instead of \mathcal{C}_{3333} .

Finally, it is easy to verify that the discussed model, rationally deduced by means of the dual-constraint approach, corresponds to that proposed by Bauld and Tzeng [1984] when nonconventional high-order

effects relating to curvature κ and thickness are neglected, and it reduces to Vlasov's classical one [1961; 1962] when an isotropic material is considered.

3C. First-order shear-deformable beam model. Since the wall thickness is small compared with other cross-section dimensions, for the sake of brevity we assume in what follows that

$$\delta \ll \int_p ds, \quad \delta\kappa \ll 1, \quad (3-39)$$

and then $j \cong 1$.

In order to include first-order shear deformation effects, constraints (ii) and (iii) of Section 3B can be replaced by the following assumptions:

(ii') The shear total strain between z -axis and $\mathbf{n}(s)$ is constant over \mathcal{P} .

(iii') The flux through the thickness of the in-plane shear total strain vector is constant over \mathcal{P} .

The corresponding dual constraints on the elastic stress field are:

(v') The shear elastic stress along $\mathbf{n}(s)$ is constant over \mathcal{P} .

(vi') The flux through the thickness of the in-plane shear elastic stress vector is constant over \mathcal{P} .

Accordingly, the operator \mathbf{G} turns out to be such that

$$\mathbf{G}\boldsymbol{\varepsilon} = \{\varepsilon_{11} \ \varepsilon_{22} \ \varepsilon_{12} \ (2\delta\bar{\varepsilon}_{13})_{/1} \ \varepsilon_{23/1} \ \varepsilon_{23/2}\}^T. \quad (3-40)$$

Under assumptions (3-39) and recalling (3-3), the following equality can be stated:

$$(2\delta\bar{\Lambda}_{x3})_{/1}\mathbf{i} + (2\delta\bar{\Lambda}_{y3})_{/1}\mathbf{j} \cong (2\delta\bar{\Lambda}_{13})_{/1}\mathbf{t} + (2\delta\bar{\Lambda}_{23})_{/1}\mathbf{n} = \mathbf{0}. \quad (3-41)$$

Therefore, constraints (ii') and (iii') are equivalent to requiring that the integral over the thickness of total shear strain between the z -axis and every in-plane direction depends only on z , that is, to the Timoshenko-type kinematic shear assumptions

$$2\bar{\Lambda}_{x3} = \gamma_x(z), \quad 2\bar{\Lambda}_{y3} = \gamma_y(z). \quad (3-42)$$

Analogously, the dual constraints (v') and (vi') can be thought as referring to $\sigma_{x3}^{(el)}$ and $\sigma_{y3}^{(el)}$.

Hence, the functional (2-7) can be written as

$$\begin{aligned} \mathcal{E}(\mathbf{u}, \boldsymbol{\chi}, \boldsymbol{\omega}) = & \frac{1}{2} \int_{\Omega} \{ \mathcal{C}_{\alpha\beta\gamma\delta} (\Lambda_{\alpha\beta} - \omega_{\alpha\beta}) (\Lambda_{\gamma\delta} - \omega_{\gamma\delta}) \\ & + 2\mathcal{C}_{\alpha\beta 33} (\Lambda_{\alpha\beta} - \omega_{\alpha\beta}) u'_3 + \mathcal{C}_{3333} (u'_3)^2 + 4\mathcal{C}_{1313} [\Lambda_{13} + (2\delta\omega_{13})_{/1}]^2 \\ & + 8\mathcal{C}_{1323} [\Lambda_{13} + (2\delta\omega_{13})_{/1}] [\Lambda_{23} + \tilde{\omega}_{\alpha 3/\alpha}] + 4\mathcal{C}_{2323} [\Lambda_{23} + \tilde{\omega}_{\alpha 3/\alpha}]^2 \} dv \\ & - \int_{\Omega} [\chi_{\alpha\beta} \Lambda_{\alpha\beta} + 2\chi_{13} (2\delta\bar{\Lambda}_{13})_{/1} + 2\tilde{\chi}_{\alpha 3} \Lambda_{23/\alpha}] dv - \Pi_{\text{ext}}, \quad (3-43) \end{aligned}$$

where $\tilde{\omega}_{\alpha 3}$, $\tilde{\chi}_{\alpha 3}$, $\omega_{13}(s, z)$, and $\chi_{13}(s, z)$ are the Lagrange multipliers relating to constraints (ii'), (v'), (iii') and (vi'), respectively. The functional \mathcal{E} is defined on the manifold: $\tilde{\chi}_{\alpha 3} = \tilde{\omega}_{\alpha 3} = 0$ on Σ and $\omega_{13} = \chi_{13} = 0$ on $\partial p \times]-L, L[$.

The stationary conditions of (3-43) with respect to χ return constraints on the displacement field:

$$u_{1/1} - \kappa u_2 = 0, \quad u_{2/2} = 0, \quad u_{2/1} + \kappa u_1 + u_{1/2} = 0, \quad (3-44)$$

$$\int_{-\delta}^{\delta} (u_{3/11} + u'_{1/1}) d\eta = 0, \quad u'_{2/\alpha} + u_{3/2\alpha} = 0, \quad (3-45)$$

whose integration yields the Cartesian representation formulas

$$\begin{aligned} u_x(s, \eta, z) &= u_c(z) - \theta(z)[y(s, \eta) - y_c], \\ u_y(s, \eta, z) &= v_c(z) + \theta(z)[x(s, \eta) - x_c], \\ u_3(s, \eta, z) &= w_c(z) - \phi_y(z)x(s, \eta) + \phi_x(z)y(s, \eta) + \theta'(z)\psi_c(s, \eta), \end{aligned} \quad (3-46)$$

where the positions introduced in (3-12) are assumed to be valid and where

$$\phi_x(z) = \gamma_y(z) - v'_c(z), \quad \phi_y(z) = u'_c(z) - \gamma_x(z). \quad (3-47)$$

The stationary conditions of \mathcal{E} in (3-43) with respect to ω give the equations

$$\begin{aligned} \mathcal{C}_{\alpha\beta\gamma\delta}[\Lambda_{\gamma\delta} - \omega_{\gamma\delta}] + \mathcal{C}_{\alpha\beta 33}u'_3 &= 0, \\ \left\{ \int_{-\delta}^{\delta} \mathcal{C}_{1313}[\Lambda_{13} + (2\delta\omega_{13})_{/1}] + \mathcal{C}_{1323}[\Lambda_{23} + \tilde{\omega}_{\alpha 3/\alpha}] d\eta \right\}_{/1} &= 0, \\ \{\mathcal{C}_{1323}[\Lambda_{13} + (2\delta\omega_{13})_{/1}] + \mathcal{C}_{2323}[\Lambda_{23} + \tilde{\omega}_{\alpha 3/\alpha}]\}_{/\beta} &= 0, \end{aligned} \quad (3-48)$$

from which the Lagrange multipliers $\omega_{\alpha\beta}$, ω_{13} and $\tilde{\omega}_{\alpha 3}$ can be found:

$$\omega_{\alpha\beta} = (\mathcal{C}_{\alpha\beta\gamma\delta})^{-1} \mathcal{C}_{\gamma\delta 33}u'_3, \quad \omega_{13} = \tilde{\omega}_{13} = 0, \quad \tilde{\omega}_{23} = \frac{\mathcal{C}_{1323}}{\mathcal{C}_{2323}} \theta'(\eta^2 - \delta^2). \quad (3-49)$$

Substituting equations (3-46) and (3-49) into the functional (3-43), the potential energy functional $\hat{\mathcal{E}}$ is obtained in terms of pure displacement unknowns $u_c, v_c, w_c, \phi_x, \phi_y, \theta$, and it can be expressed as in (3-16). In this case, the generalized displacements $\hat{\mathbf{s}}$ and generalized strains \mathbf{e} are

$$\hat{\mathbf{s}} = \{u_c \ v_c \ w_c \ \phi_x \ \phi_y \ \theta \ \theta'\}^T, \quad \mathbf{e} = \{w'_c \ \phi'_x \ \phi'_y \ \theta'' \ \theta' \ \gamma_x \ \gamma_y\}^T,$$

and the generalized elastic matrix of the beam, of size 7×7 , takes on the form

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}^{(0)} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{(1)} \end{bmatrix} \quad (3-50)$$

where the 5×5 submatrix $\mathbf{D}^{(0)}$ coincides with the matrix \mathbf{D} of (3-25), derived for the no-shear case, and the symmetrical 2×2 submatrix $\mathbf{D}^{(1)}$ is defined by

$$\begin{aligned} D_{11}^{(1)} &= \mathcal{C}_{1313} \int_{\mathcal{Q}} (\tilde{x}_{/1})^2 da + \mathcal{C}_{2323} \int_{\mathcal{Q}} (\tilde{y}_{/1})^2 da - 2\mathcal{C}_{1323} \int_{\mathcal{Q}} \tilde{x}_{/1} \tilde{y}_{/1} da, \\ D_{12}^{(1)} &= (\mathcal{C}_{1313} - \mathcal{C}_{2323}) \int_{\mathcal{Q}} \tilde{x}_{/1} \tilde{y}_{/1} da + \mathcal{C}_{1323} \int_{\mathcal{Q}} [(\tilde{x}_{/1})^2 - (\tilde{y}_{/1})^2] da, \\ D_{22}^{(1)} &= \mathcal{C}_{2323} \int_{\mathcal{Q}} (\tilde{x}_{/1})^2 da + \mathcal{C}_{1313} \int_{\mathcal{Q}} (\tilde{y}_{/1})^2 da + 2\mathcal{C}_{1323} \int_{\mathcal{Q}} \tilde{x}_{/1} \tilde{y}_{/1} da. \end{aligned} \quad (3-51)$$

Moreover, in this model the generalized stress vector $\mathbf{S} = \mathbf{D}\mathbf{e}$ contains shear resultant forces T_x and T_y , representing static quantities associated to the generalized strains γ_x and γ_y , respectively:

$$\{T_x, T_y\} = \int_{\mathcal{P}} \boldsymbol{\tau} \cdot \{\mathbf{i}, \mathbf{j}\} da = \int_{\mathcal{P}} \sigma_{13}\{\tilde{x}/1, \tilde{y}/1\} da + \int_{\mathcal{P}} \sigma_{23}\{-\tilde{y}/1, \tilde{x}/1\} da. \quad (3-52)$$

Under geometrical assumptions making $\mathbf{D}^{(0)}$ diagonal, the functional $\hat{\mathcal{E}}$ can be written in the form

$$\hat{\mathcal{E}}(u_c, v_c, w_c, \phi_x, \phi_y, \theta) = \mathcal{A}(w_c) + \mathcal{F}(u_c, v_c, \phi_x, \phi_y) + \mathcal{T}(\theta), \quad (3-53)$$

where \mathcal{A} and \mathcal{T} are defined as in (3-33), and the functional \mathcal{F} is

$$\begin{aligned} \mathcal{F}(u_c, v_c, \phi_x, \phi_y) = & \frac{1}{2} \hat{\mathcal{C}}_{3333} \left[I_x \int_{-L}^L (\phi'_x)^2 dz + I_y \int_{-L}^L (\phi'_y)^2 dz \right] + \frac{1}{2} D_{11}^{(1)} \int_{-L}^L (u'_c - \phi_y)^2 dz \\ & + D_{12}^{(1)} \int_{-L}^L (u'_c - \phi_y)(v'_c + \phi_x) dz + \frac{1}{2} D_{22}^{(1)} \int_{-L}^L (v'_c + \phi_x)^2 dz. \end{aligned} \quad (3-54)$$

The stationary conditions of the functional $\hat{\mathcal{E}}$ of (3-53) supply governing equations and natural boundary conditions for the beam, allowing us to compute the unknown functions $u_c, v_c, w_c, \phi_x, \phi_y$, and θ . In detail, the extensional and torsional problems are governed by (3-34) and (3-37), and the stationary conditions of the functional \mathcal{F} yield governing equations of the Timoshenko flexural beam model, which, for the sake of brevity, are not reported here.

The elastic stress field turns out to be

$$\begin{aligned} \sigma_{13}^{(el)} &= (\mathcal{C}_{1313}\tilde{x}/1 - \mathcal{C}_{1323}\tilde{y}/1)\gamma_x + (\mathcal{C}_{1313}\tilde{y}/1 + \mathcal{C}_{1323}\tilde{x}/1)\gamma_y - 2\hat{\mathcal{C}}_{1313}\theta'\eta, \\ \sigma_{\alpha\beta}^{(el)} = 0, \quad \sigma_{23}^{(el)} &= (\mathcal{C}_{1323}\tilde{x}/1 - \mathcal{C}_{2323}\tilde{y}/1)\gamma_x + (\mathcal{C}_{1323}\tilde{y}/1 + \mathcal{C}_{2323}\tilde{x}/1)\gamma_y, \\ \sigma_{33}^{(el)} &= \hat{\mathcal{C}}_{3333}(w'_c - \phi'_y x + \phi'_x y + \theta'' \psi_c). \end{aligned} \quad (3-55)$$

The total stress field coincides with the elastic one only for the component σ_{33} , whereas, as for the case of the no-shear model, total stress components $\sigma_{\alpha\beta}$ and $\sigma_{\alpha 3}$ can not be uniquely recovered in a general case from the equilibrium equations (2-3).

We note that the present model corresponds to the one proposed in [Ascione et al. 2000] when Galerkin-type enhancements are neglected and, when applied to homogeneous beams, to the model analyzed in [Maddur and Chaturvedi 1999]. Moreover, it reduces to the no-shear model discussed in Section 3B when γ_x and γ_y are enforced to vanish.

3D. A shear refinement. The procedure until now employed can be successfully applied for deducing in a consistent way different thin-walled beam models, in which shear deformation effects can be taken into account through different refinement levels. For instance, the kinematics employed in [Lee 2005; Cortínez and Piovan 2002; Piovan and Cortínez 2007] and including shear effects due to nonuniform warping by a first-order warping shear term, can be deduced, under assumptions (3-39), by enforcing on the total strain field, together with (i) and (iii'), the following constraint:

(ii'') The shear total strain between z -axis and $\mathbf{n}(s)$ at every position $(s, \eta) \in \mathcal{P}$ does not depend upon the thickness coordinate η and varies linearly along p .

Constraints (ii'') and (iii') are equivalent to prescribing the equalities

$$2\bar{\Lambda}_{xz} = \gamma_x(z) - \gamma_\psi(z)(\tilde{y} - y_c), \quad 2\bar{\Lambda}_{yz} = \gamma_y(z) + \gamma_\psi(z)(\tilde{x} - x_c), \quad (3-56)$$

where γ_ψ is the first-order warping shear unknown function. In this case, the operator \mathbf{G} is given by

$$\mathbf{G}\boldsymbol{\varepsilon} = \{\varepsilon_{11} \ \varepsilon_{22} \ \varepsilon_{12} \ (2\delta\bar{\varepsilon}_{13})_{/1} \ \varepsilon_{23/11} \ \varepsilon_{23/2}\}^T. \quad (3-57)$$

Thus the admissible displacement field belonging to the kernel of \mathbf{G} can be represented through (3-46) for the components u_x and u_y , and the displacement along beam axis z becomes

$$u_3(s, \eta, z) = w_c(z) - \phi_y(z)x(s, \eta) + \phi_x(z)y(s, \eta) + \phi_z(z)\psi_c(s, \eta), \quad (3-58)$$

where

$$\phi_z(z) = \theta'(z) - \gamma_\psi(z). \quad (3-59)$$

Following the present dual-constraint approach, the constraints on the elastic stress field are directly stated and the potential energy functional is

$$\begin{aligned} \mathcal{E}(\mathbf{u}, \boldsymbol{\chi}, \boldsymbol{\omega}) = & \frac{1}{2} \int_{\Omega} \{ \mathcal{C}_{\alpha\beta\gamma\delta} (\Lambda_{\alpha\beta} - \omega_{\alpha\beta})(\Lambda_{\gamma\delta} - \omega_{\gamma\delta}) + 2\mathcal{C}_{\alpha\beta 33} (\Lambda_{\alpha\beta} - \omega_{\alpha\beta})u'_3 \\ & + \mathcal{C}_{3333}(u'_3)^2 + 8\mathcal{C}_{1323}[\Lambda_{13} + (2\delta\omega_{13})_{/1}][\Lambda_{23} - \tilde{\omega}_{13/11} + \tilde{\omega}_{23/2}] \\ & + 4\mathcal{C}_{1313}[\Lambda_{13} + (2\delta\omega_{13})_{/1}]^2 + 4\mathcal{C}_{2323}[\Lambda_{23} - \tilde{\omega}_{13/11} + \tilde{\omega}_{23/2}]^2 \} dv \\ & - \int_{\Omega} [\chi_{\alpha\beta}\Lambda_{\alpha\beta} + 2\chi_{13}(2\delta\bar{\Lambda}_{13})_{/1} + 2(\tilde{\chi}_{13}\Lambda_{23/11} + \tilde{\chi}_{23}\Lambda_{23/2})] dv - \Pi_{\text{ext}}, \quad (3-60) \end{aligned}$$

defined on the manifold: $\tilde{\chi}_{\alpha 3} = \tilde{\chi}_{23/1} = \tilde{\omega}_{\alpha 3} = \tilde{\omega}_{13/1} = 0$ on Σ , and $\omega_{13} = \chi_{13} = 0$ on $\partial p \times]-L, L[$.

Consistent generalized stress-strain relationship and governing equations can be rationally obtained as in the previous sections, and now the vectors of generalized displacements, strains and stresses become

$$\begin{aligned} \hat{\mathbf{s}} &= \{u_c \ v_c \ w_c \ \phi_x \ \phi_y \ \theta \ \phi_z\}^T, \\ \mathbf{e} &= \{w'_c \ \phi'_x \ \phi'_y \ \phi'_z \ \theta' \ \gamma_x \ \gamma_y \ \gamma_\psi\}^T, \\ \mathbf{S} &= \{N \ C_x \ C_y \ C_\psi \ C_z \ T_x \ T_y \ C_z^\psi\}^T, \end{aligned} \quad (3-61)$$

where the static quantity C_z^ψ associated to γ_ψ represents the secondary twisting moment, defined as

$$C_z^\psi = \int_{\mathcal{P}} (\sigma_{13} \mathbf{r} \cdot \mathbf{t} - \sigma_{23} \mathbf{r} \cdot \mathbf{n}) da. \quad (3-62)$$

Accordingly, generalized elastic matrix \mathbf{D} is expressed as in (3-50), with a symmetrical submatrix $\mathbf{D}^{(1)}$ of size 3×3 ; the new elements of \mathbf{D} are

$$\begin{aligned} D_{13}^{(1)} &= -\mathcal{C}_{1313} \int_{\mathcal{P}} \mathbf{r} \cdot \mathbf{n} \tilde{x}_{/1} da + \mathcal{C}_{1323} \int_{\mathcal{P}} (\mathbf{r} \cdot \mathbf{t} \tilde{x}_{/1} + \mathbf{r} \cdot \mathbf{n} \tilde{y}_{/1}) da - \mathcal{C}_{2323} \int_{\mathcal{P}} \mathbf{r} \cdot \mathbf{t} \tilde{y}_{/1} da, \\ D_{23}^{(1)} &= -\mathcal{C}_{1313} \int_{\mathcal{P}} \mathbf{r} \cdot \mathbf{n} \tilde{y}_{/1} da + \mathcal{C}_{1323} \int_{\mathcal{P}} (\mathbf{r} \cdot \mathbf{t} \tilde{y}_{/1} - \mathbf{r} \cdot \mathbf{n} \tilde{x}_{/1}) da + \mathcal{C}_{2323} \int_{\mathcal{P}} \mathbf{r} \cdot \mathbf{t} \tilde{x}_{/1} da, \\ D_{33}^{(1)} &= \mathcal{C}_{1313} \int_{\mathcal{P}} (\mathbf{r} \cdot \mathbf{n})^2 da - 2\mathcal{C}_{1323} \int_{\mathcal{P}} (\mathbf{r} \cdot \mathbf{t})(\mathbf{r} \cdot \mathbf{n}) da + \mathcal{C}_{2323} \int_{\mathcal{P}} (\mathbf{r} \cdot \mathbf{t})^2 da. \end{aligned} \quad (3-63)$$

The elastic stress field comes out as

$$\begin{aligned} \sigma_{13}^{(el)} &= \hat{\sigma}_{13}^{(el)} - (\mathcal{C}_{1313} \mathbf{r} \cdot \mathbf{n} - \mathcal{C}_{1323} \mathbf{r} \cdot \mathbf{t}) \gamma_\psi, \\ \sigma_{\alpha\beta}^{(el)} &= 0, \quad \sigma_{23}^{(el)} = \hat{\sigma}_{23}^{(el)} - (\mathcal{C}_{1323} \mathbf{r} \cdot \mathbf{n} - \mathcal{C}_{2323} \mathbf{r} \cdot \mathbf{t}) \gamma_\psi, \\ \sigma_{33}^{(el)} &= \hat{\mathcal{C}}_{3333} (w'_c - \phi'_y x + \phi'_x y + \phi'_z \psi_c), \end{aligned} \quad (3-64)$$

where $\hat{\sigma}_{\alpha 3}^{(el)}$ is the elastic in-plane shear stress components expressed in (3-55).

The present model reduces, for homogeneous thin-walled beams, to the one employed, for instance, in [Lee 2005; Cortínez and Piován 2002] and it coincides with the model deduced in Section 3C when γ_ψ is enforced to vanish.

3E. A branchwise model. If the thin-walled beam cross-section is subdivided in N_b branches (as it naturally occurs for I, C, L, Z -type profiles) a new and more refined consistent model could be developed by enforcing dual constraints associated to (i), (ii'') and (iii') for every branch b . As for layerwise laminated plate models [Bisegna and Sacco 1997a], a branchwise compatible displacement field comes out with usual components u_x and u_y and with u_3 represented as

$$u_3^{(b)}(s, \eta, z) = w^{(b)}(z) - \phi_y^{(b)}(z)x(s, \eta) + \phi_x^{(b)}(z)y(s, \eta) + \phi_z^{(b)}(z)\psi_c(s, \eta). \quad (3-65)$$

The quantities $w^{(b)}(z)$ can be easily expressed in terms of unknown functions $w_c(z)$, $\phi_x^{(b)}(z)$, $\phi_y^{(b)}(z)$ and $\phi_z^{(b)}(z)$ by imposing continuity conditions on u_3 at every centerline joint between contiguous branches. Accordingly, the number of displacement unknown functions is $4 + 3N_b$ (namely u_c , v_c , w_c , θ , $\phi_x^{(b)}$, $\phi_y^{(b)}$, $\phi_z^{(b)}$, where b varies from 1 to N_b) and the corresponding governing equations can be deduced following the previous variational constrained approach. Further developments will be given in a future paper.

4. Recovering total shear stresses

In many technical problems the primary interest is on the total stress field (elastic plus reactive stress field), that is the stress field which is in equilibrium with the external loads. Indeed, just total stress values should be compared against material's strength in order to prevent failure. Therefore, once the elastic stress field is obtained by means of the dual-constraint approach, the reactive stress field should be determined. Unfortunately, the constraint (i) employed in the previously discussed thin-walled beam models and prescribing the rigid rotation of the beam cross-section in its representation plane, does not enable to recover equilibrated stress components $\sigma_{\alpha\beta}$. Nevertheless, a technical analysis of beam-type structural elements usually involves the stress vector on the plane of \mathcal{P} , i.e., only the stress components σ_{33} , $\sigma_{\alpha 3}$.

For all cases under investigation the component $\sigma_{33}^{(el)}$ of the elastic stress field coincides with the total σ_{33} , and an useful estimate of the total shear stress components $\sigma_{\alpha 3}$, much more accurate in a local sense than the elastic ones, can be obtained as follows.

The field and boundary (on Σ) equilibrium equations along z can be written, respectively, as

$$\begin{aligned} \text{Div } \boldsymbol{\tau} + \sigma_{33/3} + b_3 &= 0 \quad \text{in } \Omega, \\ \boldsymbol{\tau} \cdot \mathbf{n}_\partial &= \hat{p}_3 \quad \text{on } \Sigma, \end{aligned} \quad (4-1)$$

where Div denotes the divergence operator acting on \mathcal{P} and the in-plane total shear stress vector $\boldsymbol{\tau}$ is defined as in (3-7). Due to the stated assumptions (see Section 3A) the outward normal unit vector \mathbf{n}_∂ to the mantle Σ (i.e., to $\partial\mathcal{P}$) can be approximated by means of normal and tangent unit vectors to the center-line p , that is

$$\mathbf{n}_\partial = \begin{cases} \pm \mathbf{n} & \text{on } \partial\mathcal{P}|_{\eta=\pm\delta}, \\ -\mathbf{t} & \text{on } \partial\mathcal{P}|_{s=s_0}, \\ \mathbf{t} & \text{on } \partial\mathcal{P}|_{s=s_\ell}. \end{cases} \quad (4-2)$$

Consider an internal beam cross-section $\mathcal{P}^i = \mathcal{P} \times \{z^i\}$, where $z^i \in]-L, L[$, and let $s^i \in]s_0, s_\ell[$ to be an internal coordinate on p . Then \mathcal{P}^i can be subdivided into two complementary parts \mathcal{P}_1^i and \mathcal{P}_2^i , such that $\mathcal{P}^i = \mathcal{P}_1^i \cup \mathcal{P}_2^i$ and $\mathcal{P}_1^i \cap \mathcal{P}_2^i = \xi$, where ξ denotes the thickness chord at s^i , i.e., $\xi = \{s^i\} \times]-\delta, \delta[$, whose unit normal vector is $\mathbf{t}(s^i)$, outward directed from \mathcal{P}_1^i .

Integrating (4-1)₁ over \mathcal{P}_1^i , applying the divergence theorem and taking (4-1)₂ into account, the exact average value over the thickness of the total shear stress σ_{13} is obtained as

$$\bar{\sigma}_{13}(s^i, z) = -\frac{1}{2\delta} \left(\int_{\mathcal{P}_1^i} (\sigma_{33/3} + b_3) da + \int_{\partial\mathcal{P}_1^i \setminus \xi} \hat{p}_3 d\varrho \right), \quad (4-3)$$

where, expressing σ_{33} via (3-64), we have

$$\int_{\mathcal{P}_1^i} \sigma_{33/3} da = \hat{\mathcal{C}}_{3333} (w_c'' A_1 + \phi_x'' S_{x1} + \phi_y'' S_{y1} + \phi_z'' S_{\psi 1}); \quad (4-4)$$

here the quantities A_1 and $S_{.1}$ are defined as in (3-26) and refer to \mathcal{P}_1^i , i.e., as functions of s^i .

Since the wall thickness is small, the reactive part relating to σ_{13} can be identified with its average value over the thickness. Accordingly, the total shear stress σ_{13} can be approximated as

$$\sigma_{13}(s^i, \eta, z) \cong \sigma_{13}^{(\text{el})} + (\bar{\sigma}_{13} - \bar{\sigma}_{13}^{(\text{el})}) = -2\hat{\mathcal{C}}_{1313}\theta' + \bar{\sigma}_{13}. \quad (4-5)$$

Observe that (4-5) satisfies boundary conditions at $\partial\mathcal{P}^i|_{s=s_0}$ and $\partial\mathcal{P}^i|_{s=s_\ell}$ in an integral sense, i.e., considering the resultant over the thickness of distribution \hat{p}_3 at $s = s_0$ and $s = s_\ell$. Nevertheless, in the framework of assumptions (3-39) and considering s^i sufficiently far from ∂p , estimate (4-5) can be applied without significant error.

In order to give significant indications about the total stress component σ_{23} , due to (4-3), (4-4) and (4-5), and observing that $\sigma_{13/1} = \bar{\sigma}_{13/1}$, the field equilibrium along z in (4-1) can be written as

$$\sigma_{23/2} + (\sigma_{33/3} - \bar{\sigma}_{33/3}) + (b_3 - \bar{b}_3) - \frac{\hat{p}_3^+ + \hat{p}_3^-}{2\delta} = 0 \quad \text{in } \Omega \quad (4-6)$$

where \hat{p}_3^+ and \hat{p}_3^- stand for the components of surface forces $\hat{\mathbf{p}}$ along z at $\partial\mathcal{P}_1^i|_{\eta=\delta}$ and $\partial\mathcal{P}_1^i|_{\eta=-\delta}$, respectively.

Let the volume force component b_3 be almost linear in η -coordinate or let such an approximation be possible. Using (3-64) and taking into account boundary conditions $\sigma_{23}|_{\eta=\pm\delta} = \pm\hat{p}_3^\pm$, we integrate (4-6) with respect to η to obtain

$$\sigma_{23}(s^i, \eta, z) = \frac{\Xi}{2} (\delta^2 - \eta^2) + \frac{\hat{p}_3^+ + \hat{p}_3^-}{2\delta} \eta + \frac{\hat{p}_3^+ - \hat{p}_3^-}{2} \quad (4-7)$$

with

$$\Xi = \hat{\mathcal{C}}_{3333}[\phi''_x \tilde{x}_{/1} + \phi''_y \tilde{y}_{/1} - \phi''_z \mathbf{r} \cdot \mathbf{t}] + \left. \frac{\partial b_3}{\partial \eta} \right|_{\eta=0}. \tag{4-8}$$

Therefore, in the limit of positions (3-39) and assuming that the components of the surface forces \mathbf{p} acting upon the beam ends \mathcal{P}^\pm coincide with the total stress components at \mathcal{P}^\pm (minus at \mathcal{P}^-), equations (4-5) and (4-7) recover the in-plane total shear stress field. However, even when the boundary equilibrium on \mathcal{P}^\pm is not locally satisfied, (4-5) and (4-7) can be considered as an useful approximation in the spirit of Saint-Venant’s principle [Toupin 1965].

5. Some generalizations

We now turn to some generalized models in the case of laminated thin-walled beam and of nonlinear conewise elastic materials, with particular reference to bimodular ones.

5A. Laminated thin-walled beams. Assume the thin-walled beam Ω is formed by perfectly bonded layers and the beam cross-section \mathcal{P} is obtained as the union of N_b perfectly bonded branches $\mathcal{P}^{(b)}$, such that the centerline curve p can be defined as the union of the branch centerlines $p^{(b)}$. Each branch is formed by $N_l^{(b)}$ layers and in what follows any quantity referring to the l -th layer of the b -th branch is marked by superscript (b, l) . Shape and dimensions of each branch $\mathcal{P}^{(b)}$ are assumed to be independent on z . Accordingly, the l -th layer of the b -th branch occupies the region

$$\Omega^{(b,l)} = \mathcal{P}^{(b)} \times]-L, L[= p^{(b)} \times]\eta^{(b,l-1)}, \eta^{(b,l)}[\times]-L, L[,$$

with $\eta^{(b,0)} = -\delta^{(b)}$ and $\eta^{(b,N_l^{(b)})} = \delta^{(b)}$, where $2\delta^{(b)}$ is the overall thickness of the b -th branch. Each layer of the laminated beam is assumed to be homogeneous and comprising a linearly elastic material, having at least a monoclinic symmetry with symmetry plane orthogonal to \mathbf{k} . This material symmetry enables to model multilayer composite profiles commonly used in civil engineering, as for example fiber-reinforced beams produced with pultrusion technology, characterized by orthotropic symmetry with one of the orthotropy axes coincident with the beam axis-line.

If the first-order shear-deformable beam model including shear effects due to nonuniform warping is considered, constraints (i), (ii'') and (iii') (see Sections 3B, 3C and 3D) on the total strain field have to be enforced on Ω . Correspondingly, the dual constraints on the elastic stress field are (iv), (v'') and (vi') and they refer to each region $\Omega^{(b,l)}$. Accordingly, the potential energy functional \mathcal{E} is expressed by (3-60), where the integrals over \mathcal{P} have to be regarded as the summation of integrals over $\mathcal{P}^{(b,l)}$, and it is defined on the manifold: $\tilde{\chi}_{\alpha 3} = \tilde{\chi}_{23/1} = 0$ on Σ , $\tilde{\omega}_{\alpha 3} = \tilde{\omega}_{13/1} = 0$ on $\partial\mathcal{P}^{(b,l)} \times]-L, L[$, $\chi_{13} = 0$ on $\partial p \times]-L, L[$, and $\omega_{13} = 0$ on $\partial p^{(b)} \times]-L, L[$.

Following the procedure discussed previously and eliminating the dependency on ω and χ in \mathcal{E} , the kinematics of the beam is described by equations (3-46)₁, (3-46)₂, and (3-58); the generalized vectors are defined as in (3-19) and (3-61); the elastic stress field for each region $\Omega^{(b,l)}$ is expressed by (3-64) and the nonzero elements of the 8×8 symmetric generalized elastic matrix \mathbf{D} are

$$\begin{aligned}
D_{11} &= \sum_b \int_{p^{(b)}} \hat{\mathcal{A}}_{3333}^{(b)} ds, & D_{12} &= \sum_b \int_{p^{(b)}} (\hat{\mathcal{A}}_{3333}^{(b)} \tilde{y} + \hat{\mathcal{B}}_{3333}^{(b)} \tilde{x}_{/1}) ds, \\
D_{13} &= - \sum_b \int_{p^{(b)}} (\hat{\mathcal{A}}_{3333}^{(b)} \tilde{x} - \hat{\mathcal{B}}_{3333}^{(b)} \tilde{y}_{/1}) ds, & D_{14} &= \sum_b \int_{p^{(b)}} (\hat{\mathcal{A}}_{3333}^{(b)} \tilde{\psi}_c - \hat{\mathcal{B}}_{3333}^{(b)} \mathbf{r} \cdot \mathbf{t}) ds, \\
D_{22} &= \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{y}^2 + 2\hat{\mathcal{B}}_{3333}^{(b)} \tilde{x}_{/1} \tilde{y} + \hat{\mathcal{D}}_{3333}^{(b)} (\tilde{x}_{/1})^2] ds, \\
D_{23} &= - \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{x} \tilde{y} + \hat{\mathcal{B}}_{3333}^{(b)} (\tilde{x} \tilde{x}_{/1} - \tilde{y} \tilde{y}_{/1}) - \hat{\mathcal{D}}_{3333}^{(b)} \tilde{x}_{/1} \tilde{y}_{/1}] ds, \\
D_{24} &= \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{y} \tilde{\psi}_c + \hat{\mathcal{B}}_{3333}^{(b)} (\tilde{x}_{/1} \tilde{\psi}_c - \tilde{y} \mathbf{r} \cdot \mathbf{t}) - \hat{\mathcal{D}}_{3333}^{(b)} \tilde{x}_{/1} \mathbf{r} \cdot \mathbf{t}] ds, \\
D_{33} &= \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{x}^2 - 2\hat{\mathcal{B}}_{3333}^{(b)} \tilde{x} \tilde{y}_{/1} + \hat{\mathcal{D}}_{3333}^{(b)} (\tilde{y}_{/1})^2] ds, \\
D_{34} &= - \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{x} \tilde{\psi}_c - \hat{\mathcal{B}}_{3333}^{(b)} (\tilde{y}_{/1} \tilde{\psi}_c + \tilde{x} \mathbf{r} \cdot \mathbf{t}) + \hat{\mathcal{D}}_{3333}^{(b)} \tilde{y}_{/1} \mathbf{r} \cdot \mathbf{t}] ds, \\
D_{44} &= \sum_b \int_{p^{(b)}} [\hat{\mathcal{A}}_{3333}^{(b)} \tilde{\psi}_c^2 - 2\hat{\mathcal{B}}_{3333}^{(b)} \tilde{\psi}_c \mathbf{r} \cdot \mathbf{t} + \hat{\mathcal{D}}_{3333}^{(b)} (\mathbf{r} \cdot \mathbf{t})^2] ds, \\
D_{55} &= 4 \sum_b \int_{p^{(b)}} \hat{\mathcal{D}}_{1313}^{(b)} ds, & D_{56} &= -2 \sum_b \int_{p^{(b)}} \hat{\mathcal{B}}_{1313}^{(b)} \tilde{x}_{/1} ds, \\
D_{57} &= -2 \sum_b \int_{p^{(b)}} \hat{\mathcal{B}}_{1313}^{(b)} \tilde{y}_{/1} ds, & D_{58} &= 2 \sum_b \int_{p^{(b)}} \hat{\mathcal{B}}_{1313}^{(b)} \mathbf{r} \cdot \mathbf{n} ds, \\
D_{66} &= \sum_b \int_{p^{(b)}} [\mathcal{A}_{1313}^{(b)} (\tilde{x}_{/1})^2 - 2\mathcal{A}_{1323}^{(b)} \tilde{x}_{/1} \tilde{y}_{/1} + \mathcal{A}_{2323}^{(b)} (\tilde{y}_{/1})^2] ds, \\
D_{67} &= \sum_b \int_{p^{(b)}} \{(\mathcal{A}_{1313}^{(b)} - \mathcal{A}_{2323}^{(b)}) \tilde{x}_{/1} \tilde{y}_{/1} + \mathcal{A}_{1323}^{(b)} [(\tilde{x}_{/1})^2 - (\tilde{y}_{/1})^2]\} ds, \\
D_{68} &= - \sum_b \int_{p^{(b)}} [\mathcal{A}_{1313}^{(b)} \tilde{x}_{/1} \mathbf{r} \cdot \mathbf{n} + \mathcal{A}_{2323}^{(b)} \tilde{y}_{/1} \mathbf{r} \cdot \mathbf{t} - \mathcal{A}_{1323}^{(b)} (\tilde{x}_{/1} \mathbf{r} \cdot \mathbf{t} + \tilde{y}_{/1} \mathbf{r} \cdot \mathbf{n})] ds, \\
D_{77} &= \sum_b \int_{p^{(b)}} [\mathcal{A}_{1313}^{(b)} (\tilde{y}_{/1})^2 + 2\mathcal{A}_{1323}^{(b)} \tilde{x}_{/1} \tilde{y}_{/1} + \mathcal{A}_{2323}^{(b)} (\tilde{x}_{/1})^2] ds, \\
D_{78} &= - \sum_b \int_{p^{(b)}} [\mathcal{A}_{1313}^{(b)} \tilde{y}_{/1} \mathbf{r} \cdot \mathbf{n} + \mathcal{A}_{1323}^{(b)} (\tilde{x}_{/1} \mathbf{r} \cdot \mathbf{n} - \tilde{y}_{/1} \mathbf{r} \cdot \mathbf{t}) - \mathcal{A}_{2323}^{(b)} \tilde{x}_{/1} \mathbf{r} \cdot \mathbf{t}] ds, \\
D_{88} &= \sum_b \int_{p^{(b)}} [\mathcal{A}_{1313}^{(b)} (\mathbf{r} \cdot \mathbf{n})^2 - 2\mathcal{A}_{1323}^{(b)} (\mathbf{r} \cdot \mathbf{n})(\mathbf{r} \cdot \mathbf{t}) + \mathcal{A}_{2323}^{(b)} (\mathbf{r} \cdot \mathbf{t})^2] ds,
\end{aligned}$$

where $\tilde{\psi}_c = \psi_c|_{\eta=0}$ and where the laminate extensional, coupling and bending stiffness coefficients \mathcal{A} ., \mathcal{B} ., \mathcal{D} . are defined by

$$\{\hat{\mathcal{A}}^{(b)}, \hat{\mathcal{B}}^{(b)}, \hat{\mathcal{D}}^{(b)}\} = \sum_l \int_{\eta^{(b,l-1)}}^{\eta^{(b,l)}} \hat{\mathcal{C}}_{\{1, \eta, \eta^2\}} d\eta, \quad \mathcal{A}_{\alpha\beta\gamma}^{(b)} = \sum_l \int_{\eta^{(b,l-1)}}^{\eta^{(b,l)}} \hat{\mathcal{C}}_{\alpha\beta\gamma} d\eta. \quad (5-1)$$

This model, in the case of the addressed material symmetry, justifies [Lee 2005] and it is easy to verify that it reduces to the one discussed in the Section 3D, when $N_b = N_l^{(b)} = 1$.

Finally, following the approach discussed in Section 4 and referring, for the sake of brevity, to the case $N_b = 1$, the total shear stresses on $\mathcal{P}^{i(l)}$ can be recovered by means of the equations

$$\begin{aligned} \sigma_{13}^{(l)}(s^i, \eta, z) &\cong \sigma_{13}^{(el,l)} + (\bar{\sigma}_{13}^{(l)} - \bar{\sigma}_{13}^{(el,l)}), \\ \sigma_{23}^{(l)}(s^i, \eta, z) &= \sigma_{23}^{-(l)} + \frac{\sigma_{23}^{+(l)} - \sigma_{23}^{-(l)}}{\eta^{(l)} - \eta^{(l-1)}}(\eta - \eta^{(l-1)}) - \Xi^{(l)} \left(\frac{\eta^2}{2} - \eta \bar{\eta}^{(l)} + \frac{\eta^{(l-1)} \eta^{(l)}}{2} \right), \end{aligned} \quad (5-2)$$

where $\Xi^{(l)}$ is as in (4-8), $\sigma_{23}^{+(l)}$ and $\sigma_{23}^{-(l)}$ stand for σ_{23} at $\partial\mathcal{P}^{i(l)}|_{\eta=\eta^{(l)}}$ and $\partial\mathcal{P}^{i(l)}|_{\eta=\eta^{(l-1)}}$ respectively, $\bar{\eta}^{(l)} = (\eta^{(l)} + \eta^{(l-1)})/2$, and

$$\bar{\sigma}_{13}^{(l)} = -\frac{1}{\eta^{(l)} - \eta^{(l-1)}} \left(\int_{\mathcal{P}^{i(l)}} (\sigma_{33/3}^{(l)} + b_3) da + \int_{\eta^{(l-1)}}^{\eta^{(l)}} \hat{p}_3|_{s=s_0} d\eta + \int_{s_0}^{s^i} (\sigma_{23}^{+(l)} - \sigma_{23}^{-(l)}) da \right). \quad (5-3)$$

Observing that

$$\sigma_{13/1}^{(l)}(s^i, \eta, z) = -\bar{\sigma}_{33/3}^{(l)} - \bar{b}_3 - \frac{\sigma_{23}^{+(l)} - \sigma_{23}^{-(l)}}{\eta^{(l)} - \eta^{(l-1)}}, \quad (5-4)$$

where here \bar{f} denotes the average value over the l -th layer thickness of f , the closure of (5-2) and (5-3) has to be performed determining $\sigma_{23}^{\pm(l)}$ by means of the continuity conditions

$$\sigma_{23}^{+(l)} = \sigma_{23}^{-(l+1)}, \quad \sigma_{13/1}^{(l)}|_{\eta^{(l)}} = \sigma_{13/1}^{(l+1)}|_{\eta^{(l)}} \quad (5-5)$$

for l from 1 to N_l and with $\sigma_{23}^{+(N_l)} = \hat{p}_3^+$ and $\sigma_{23}^{+(0)} = \hat{p}_3^-$.

5B. Nonlinear conewise elastic materials: the fiber-governed bimodular case. The variational dual-constraint approach can be successfully employed also when a thin-walled beam comprises nonlinear elastic materials characterized by a continuous and convex elastic potential (not necessarily differentiable), such as the conewise materials addressed in [Curnier et al. 1995]. Bimodular behavior, which generally may characterize the constitutive response of a number of composite materials, belong to this category. In this case a nonlinear elastic response appears, identified by a linear relationship between stress and strain both in tension and in compression, but with different elastic moduli. We will refer to the special case of fiber-reinforced composite materials, where the bimodularity depends on the sign of the unit elongation in the fiber-direction [Curnier et al. 1995; Bert 1977; Maceri and Sacco 1990; Bisegna et al. 1995].

Let $\boldsymbol{\varepsilon}$ be the symmetric strain tensor and \mathbf{f} the unit vector along the fiber direction. Denoting as $\epsilon_f = \boldsymbol{\varepsilon} \mathbf{f} \cdot \mathbf{f}$ the extension in the fiber direction, the following definitions are introduced:

$$E^+ = \{\boldsymbol{\varepsilon} : \epsilon_f > 0\}, \quad E^o = \{\boldsymbol{\varepsilon} : \epsilon_f = 0\}, \quad E^- = \{\boldsymbol{\varepsilon} : \epsilon_f < 0\}. \quad (5-6)$$

As showed in [Bisegna et al. 1995], a bimodular constitutive law can be deduced in a consistent way by assuming that an elastic potential Φ exists. Accordingly, restrictions of Φ to E^+ and E^- are the potentials for the mappings:

$$\boldsymbol{\varepsilon} \in E^+ \mapsto \mathcal{C}^+ \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \in E^- \mapsto \mathcal{C}^- \boldsymbol{\varepsilon}, \quad (5-7)$$

and the fourth order constitutive tensors \mathcal{C}^+ and \mathcal{C}^- (relevant to tension and compression behavior, respectively) are symmetric. Moreover, due to the definition of a conservative (or hyperelastic) material, the potential $\Phi(\boldsymbol{\varepsilon})$ is continuous [Bisegna et al. 1995]. As a consequence, restrictions of Φ to E^+ and E^- can be extended by continuity to E^o , providing the following equality to be satisfied:

$$\mathcal{C}^+ \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon} = \mathcal{C}^- \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}, \quad \forall \boldsymbol{\varepsilon} \in E^o. \quad (5-8)$$

Equations (5-7) and (5-8) imply that the material strain energy density Φ can be written (omitting constant contributions) in the form:

$$\Phi(\boldsymbol{\varepsilon}) = \frac{1}{2}[h\mathcal{C}^+ + (1-h)\mathcal{C}^-]\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}, \quad (5-9)$$

where $h(\epsilon_f)$ is the Heaviside function, such that $h = 1$ if ϵ_f is positive, $h = 0$ otherwise.

Thus, the first term of the Hu–Washizu functional \mathcal{W} (2-1) has to be replaced by the integral over Ω of $\Phi(\boldsymbol{\varepsilon})$ and then, introducing a Lagrangian functional \mathcal{L} as in (2-2), the potential energy functional \mathcal{E} turns out to be

$$\mathcal{E}(\mathbf{u}, \boldsymbol{\chi}, \boldsymbol{\omega}) = \int_{\Omega} \Phi(\hat{\nabla} \mathbf{u} - \mathbf{G}^* \boldsymbol{\omega}) dv - \int_{\Omega} \boldsymbol{\chi} \cdot \mathbf{G} \hat{\nabla} \mathbf{u} dv - \Pi_{\text{ext}}. \quad (5-10)$$

The stationary conditions of \mathcal{L} with respect to \mathbf{u} , $\boldsymbol{\varepsilon}$, $\boldsymbol{\sigma}$, $\boldsymbol{\chi}$ and $\boldsymbol{\omega}$ formally give the same equations (2-3) to (2-6), where now \mathcal{C} has to be regarded as

$$\mathcal{C} = [h\mathcal{C}^+ + (1-h)\mathcal{C}^-]. \quad (5-11)$$

Accordingly, the dual-constraint approach can be employed in the case of bimodular fiber-governed materials.

In this case, the stationary condition of the Lagrangian functional \mathcal{L} (2-2) with respect of $\boldsymbol{\varepsilon}$ yields (2-4) without jumping terms (depending on the difference between \mathcal{C}^+ and \mathcal{C}^-) as a consequence of the continuity condition (5-8). Moreover, strain along the fiber-direction ϵ_f , whose sign discriminates the unilateral constitutive behavior, has to be considered as deduced from the elastic strain field.

Referring to the first-order shear-deformable beam model accounting for warping shear, when a homogeneous beam comprises a bimodular material at least monoclinic, the generalized elastic matrix \mathbf{D} can be evaluated as in Section 5A taking \mathcal{P} as $\mathcal{P} = \mathcal{P}|_{\mathcal{C}^+} \cup \mathcal{P}|_{\mathcal{C}^-}$. Clearly, this partition of \mathcal{P} needs the preliminary knowledge for every z -coordinate value of the neutral region satisfying $\epsilon_f = 0$. Therefore, a free-boundary problem underlies this formulation and an iterative procedure has to be employed in order to evaluate elements of \mathbf{D} .

6. Influence of the curvature

In order to assess the influence of high-order contributions associated with the curvature κ of p (see (3-13) and (3-28)), some results relating to very simple numerical applications are herein presented. In detail, we refer to the no-shear model discussed in Section 3B when it is applied to a pure torsion problem for a thin-walled open section beam, with a circular centerline ($\kappa = \text{const}$).

Assuming that $\hat{\mathcal{C}}_{3333}/\hat{\mathcal{C}}_{1313} = 2.5$, the following two cases are analyzed:

- a simply supported beam ($\theta(\pm L) = \theta''(\pm L) = 0$, i.e., warping free) loaded by an uniform twisting moment distribution m_z ;

- a cantilever beam ($\theta(-L) = \theta'(-L) = 0$, i.e., warping constrained) loaded by a twisting moment M_z applied at its free end.

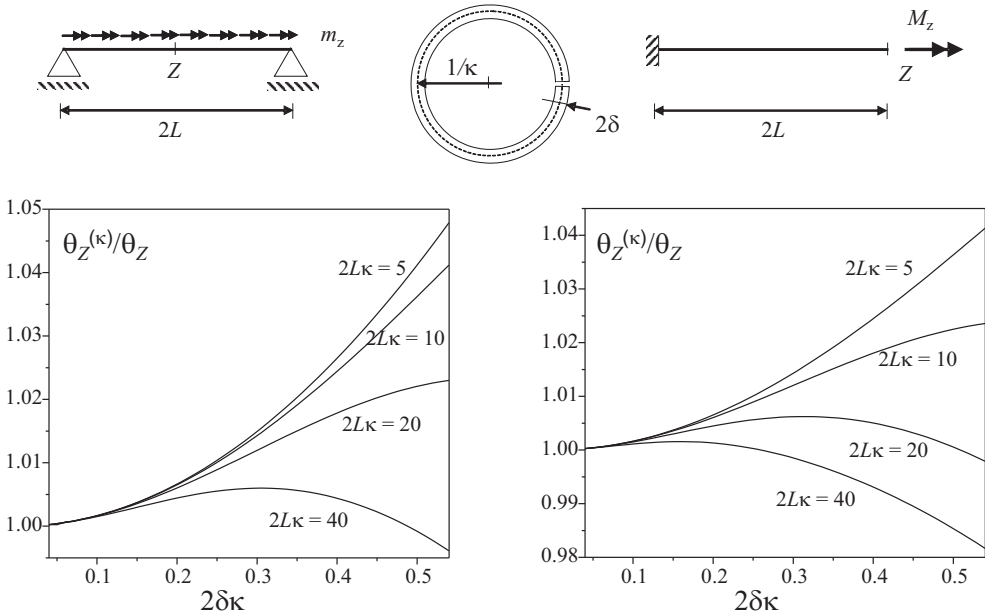


Figure 2. Influence of high-order nonconventional effects relating to curvature κ of p in torsional analysis of a simply supported (on the left) and a cantilever (on the right) thin-walled beam with a circular open cross-section.

Figure 2 sketches the cases under investigation and depicts the ratios between the maximum rotation angle around the z -axis evaluated taking into account curvature effects ($\theta^{(\kappa)}$) versus that computed disregarding curvature effects (θ). It can be noted that, varying the dimensionless thickness $2\delta\kappa$ and the dimensionless beam length $2L\kappa$, the influence of curvature increases when the beam slenderness ratio decreases (i.e., when $L \rightarrow \sim 1/\kappa$), amounting to approximately 4–5% and increasing when the simply supported scheme is experienced. Therefore, high-order effects associated to κ and δ might significantly affect, from an engineering point of view, thin-walled beam static response.

7. Concluding remarks

This paper presents a consistent deduction of anisotropic thin-walled beam models from three-dimensional elasticity. Employing a constrained approach and following the strategy first outlined in [Maceri and Bisegna 2002], a modified Hu–Washizu variational formulation has been proposed. This provides a rational and unified foundation for thin-walled beam theories widely used in the recent literature, and allows us to account for different shear refinement levels.

This variational approach enables to take into account high-order nonconventional effects related to curvature of cross-section profile and wall thickness, leading to a generalized warping function and torsional constant. Influence of these effects has been proved through simple numerical applications, highlighting that in some cases it could be not completely negligible from an engineering point of view.

It was shown in the case of a branchwise model for the analysis of multibranch cross-sections that this approach allows the consistent deduction of new refined models. Moreover, due to its variational character, this formulation opens the possibility of building new consistent and refined thin-walled beam finite elements. Finally, a more accurate evaluation of stresses acting on the beam cross-section with respect to those obtained by constitutive relationship has been also proposed. In agreement with the dual-constraint approach and starting from the one-dimensional generalized unknown functions, approximated locally equilibrated stresses are recovered in closed form.

The dual-constraint variational framework has also proved effective for the treatment of laminated pultruded thin-walled beams and nonlinear conewise elastic materials (with special reference to bimodular materials).

Acknowledgements

The authors thank professor Paolo Bisegna for valuable suggestions and fruitful discussions on this paper.

This work was developed within the framework of Lagrange Laboratory, a European research group comprising CNRS, CNR, the Universities of Rome “Tor Vergata”, Calabria, Cassino, Pavia, and Salerno, École Polytechnique, University of Montpellier II, ENPC, LCPC, and ENTPE.

References

- [Ambrosini et al. 2000] R. D. Ambrosini, J. D. Riera, and R. F. Danesi, “A modified Vlasov theory for dynamic analysis of thin-walled and variable open section beams”, *Eng. Struct.* **22**:8 (2000), 890–900.
- [Antman and Marlow 1991] S. S. Antman and R. S. Marlow, “Material constraints, Lagrange multipliers, and compatibility: applications to rod and shell theories”, *Arch. Ration. Mech. An.* **116**:3 (1991), 257–299.
- [Ascione et al. 2000] L. Ascione, L. Feo, and G. Mancusi, “On the statical behaviour of fibre-reinforced polymer thin-walled beams”, *Compos. B Eng.* **31**:8 (2000), 643–654.
- [Bauld and Tzeng 1984] N. R. Bauld and L. S. Tzeng, “A Vlasov theory for fiber-reinforced beams with thin-walled open cross sections”, *Int. J. Solids Struct.* **20**:3 (1984), 277–297.
- [Bert 1977] C. W. Bert, “Models for fibrous composites with different properties in tension and in compression”, *J. Eng. Mater. Technol. (ASME)* **99** (1977), 344–349.
- [Bisegna and Sacco 1997a] P. Bisegna and E. Sacco, “A layer-wise laminate theory rationally deduced from the three-dimensional elasticity”, *J. Appl. Mech. (ASME)* **64**:3 (1997), 538–545.
- [Bisegna and Sacco 1997b] P. Bisegna and E. Sacco, “A rational deduction of plate theories from the three-dimensional linear elasticity”, *Z. Angew. Math. Mech.* **77**:5 (1997), 349–366.
- [Bisegna et al. 1995] P. Bisegna, F. Maceri, and E. Sacco, “On the fiber-governed bimodular constitutive models”, pp. 113–128 in *Advanced technology for design and fabrication of composite materials and structures: applications to the automotive, marine, aerospace, and construction industry*, edited by G. C. Sih et al., Kluwer, Dordrecht, 1995.
- [Capurso 1964] M. Capurso, “Sul calcolo delle travi di parete sottile in presenza di forze e distorsioni, I: La teoria rigorosa e le equazioni fondamentali”, *Ricerca Sci. 2A* **34** (1964), 213–240.
- [Cortínez and Piovan 2002] V. H. Cortínez and M. T. Piovan, “Vibration and buckling of composite thin-walled beams with shear deformability”, *J. Sound Vib.* **258**:4 (2002), 701–723.
- [Curnier et al. 1995] A. Curnier, Q.-C. He, and P. Zysset, “Conewise linear elastic materials”, *J. Elasticity* **37**:1 (1995), 1–38.
- [De Lorenzis and La Tegola 2003] L. De Lorenzis and A. La Tegola, “Static behavior of pultruded fiber-reinforced polymer thin-walled beams”, *Compos. Struct.* **60**:3 (2003), 231–244.
- [Erkmen and Mohareb 2006] R. E. Erkmen and M. Mohareb, “Torsion analysis of thin-walled beams including shear deformation effects”, *Thin-Walled Struct.* **44**:10 (2006), 1096–1108.

- [Fraternali and Feo 2000] F. Fraternali and L. Feo, “On a moderate rotation theory of thin-walled composite beams”, *Compos. B Eng.* **31**:2 (2000), 141–158.
- [Hamdouni and Millet 2006] A. Hamdouni and O. Millet, “An asymptotic non-linear model for thin-walled rods with strongly curved open cross-section”, *Int. J. Non-Linear Mech.* **41**:3 (2006), 396–416.
- [Kollbrunner and Basler 1969] C. F. Kollbrunner and K. Basler, *Torsion in structures: an engineering approach*, Springer, Berlin, 1969.
- [Lee 2005] J. Lee, “Flexural analysis of thin-walled composite beams using shear-deformable beam theory”, *Compos. Struct.* **70**:2 (2005), 212–222.
- [Lee and Kim 2001] J. Lee and S.-E. Kim, “Flexural-torsional buckling of thin-walled I-section composites”, *Comput. Struct.* **79**:10 (2001), 987–995.
- [Lee and Lee 2004] J. Lee and S. Lee, “Flexural-torsional behavior of thin-walled composite beams”, *Thin-Walled Struct.* **42**:9 (2004), 1293–1305.
- [Lembo and Podio-Guidugli 2001] M. Lembo and P. Podio-Guidugli, “Internal constraints, reactive stresses, and the Timoshenko beam theory”, *J. Elasticity* **65**:1-3 (2001), 131–148.
- [Lin and Hsiao 2003] W. Y. Lin and K. M. Hsiao, “More general expression for the torsional warping of a thin-walled open-section beam”, *Int. J. Mech. Sci.* **45**:5 (2003), 831–849.
- [Maceri and Bisegna 2002] F. Maceri and P. Bisegna, “Modellazione strutturale”, pp. 1–90 in *Ingegneria delle strutture*, vol. II, edited by E. Giangreco, Utet, Torino, 2002.
- [Maceri and Sacco 1990] F. Maceri and E. Sacco, “A contribution to the mechanics of bimodular materials”, *Mater. Eng.* **1** (1990), 189–199.
- [Maddur and Chaturvedi 1999] S. S. Maddur and S. K. Chaturvedi, “Laminated composite open profile sections: first order shear deformation theory”, *Compos. Struct.* **45**:2 (1999), 105–114.
- [Piovan and Cortínez 2005] M. T. Piovan and V. H. Cortínez, “Transverse shear deformability in the dynamics of thin-walled composite beams: consistency of different approaches”, *J. Sound Vib.* **285**:3 (2005), 721–733.
- [Piovan and Cortínez 2007] M. T. Piovan and V. H. Cortínez, “Mechanics of shear deformable thin-walled beams made of composite materials”, *Thin-Walled Struct.* **45**:1 (2007), 37–62.
- [Podio-Guidugli 1989] P. Podio-Guidugli, “An exact derivation of the thin plate equation”, *J. Elasticity* **22**:2-3 (1989), 121–133.
- [Rodríguez and Viaño 1997] J. M. Rodríguez and J. M. Viaño, “Asymptotic derivation of a general linear model for thin-walled elastic rods”, *Comput. Methods Appl. Mech. Eng.* **147**:3-4 (1997), 287–321.
- [Saadé et al. 2004] K. Saadé, B. Espion, and G. Warzée, “Non-uniform torsional behavior and stability of thin-walled elastic beams with arbitrary cross sections”, *Thin-Walled Struct.* **42**:6 (2004), 857–881.
- [Toupin 1965] R. A. Toupin, “Saint-Venant’s principle”, *Arch. Ration. Mech. An.* **18**:2 (1965), 83–96.
- [Trabucho and Viaño 1996] L. Trabucho and J. M. Viaño, “Mathematical modelling of rods”, pp. 487–974 in *Handbook of numerical analysis*, vol. IV, edited by P. G. Ciarlet and J. L. Lions, Elsevier, Amsterdam, 1996.
- [Vlasov 1961] V. Z. Vlasov, *Thin walled elastic beams*, 2nd ed., Israel Program for Scientific Translations, Jerusalem, 1961.
- [Vlasov 1962] V. Z. Vlasov, *Pièces longues en voiles minces*, Eyrolles, Paris, 1962.
- [Volovoi et al. 1999] V. V. Volovoi, D. H. Hodges, V. L. Berdichevsky, and V. G. Sutyryn, “Asymptotic theory for static behavior of elastic anisotropic I-beams”, *Int. J. Solids Struct.* **36**:7 (1999), 1017–1043.

Received 26 May 2008. Accepted 16 Oct 2008.

FRANCO MACERI: franco.maceri@uniroma2.it

Università degli Studi di Roma “Tor Vergata”, Dipartimento di Ingegneria Civile, via del Politecnico 1, 00133 Roma, Italy

GIUSEPPE VAIRO: vairo@ing.uniroma2.it

Università degli Studi di Roma “Tor Vergata”, Dipartimento di Ingegneria Civile, via del Politecnico 1, 00133 Roma, Italy

CYCLIC BEHAVIOR AND ENERGY APPROACH TO THE FATIGUE OF SHAPE MEMORY ALLOYS

ZIAD MOUMNI, WAEL ZAKI AND HABIBOU MAITOURNAM

We present an energy-based low-cycle fatigue criterion that can be used in analyzing and designing structures made from shape memory alloys subjected to cyclic loading. Experimentally, a response similar to plastic shakedown is observed. During the first cycles the stress-strain curve shows a hysteresis loop which evolves during the first few cycles before stabilizing. By adopting an analogy with plastic fatigue, it is shown that the dissipated energy of the stabilized cycle is a relevant parameter for estimating the number of cycles to failure of such materials. Following these observations, we provide an application of the cyclic model, previously developed by the authors within the framework of generalized standard materials with internal constraints in order to evaluate such parameter. Numerical simulations are presented along with a validation against experimental data in case of cyclic superelasticity.

1. Introduction

The interesting behavior of SMAs (shape memory alloys) is essentially due to their capability of undergoing a reversible diffusionless solid–solid phase transition known as the martensitic transformation [Wayman and Otsuka 1999; Shaw and Kyriakides 1995; Moumni et al. 2008]. This transition is characterized at the microscopic level by a modification of the crystallographic lattice structure which can be induced by altering either the material temperature or the stress to which it is subjected or both, hence a strong thermomechanical coupling. At high temperatures, a shape memory alloy consists of a relatively ordered parent phase called austenite which transforms when cooled into a less ordered product phase called martensite. In the absence of stress, this leads to self-accommodation of martensite plates, that is, to the formation of lattice twins without any macroscopic deformation.

Mechanical loading may lead to detwinning of self-accommodating martensite. In this case, martensite plates become oriented according to privileged directions that depend on the applied stress. The resulting inelastic macroscopic strain usually reaches several percent; it can be recovered by heating, in which case the SMA regains its initial undeformed austenitic shape. Simple way shape memory refers to the ability of a shape memory alloy to remember its high temperature state.

Besides the characteristic shape memory behavior, SMAs exhibit other interesting effects, namely: superelasticity or pseudoelasticity, which is the ability of a shape memory alloy to accommodate large strains due to stress-induced phase change at a constant and sufficiently high temperature and to recover its undeformed shape upon unloading; and the superthermal effect, which is the ability to deform an initially austenitic SMA by cooling under constant stress and then to recover the austenitic shape by heating. The magnitude of the temperature-induced strains depends on the applied stress.

Keywords: cyclic pseudoelasticity, shape memory alloys, SMA fatigue, cyclic loading, residual strain, internal stress, dissipation.

Furthermore, cyclic loading may allow SMAs to have a two-way shape memory effect. In this case, the material can change its shape reversibly due to cyclic heating-cooling.

Since components made of SMAs usually operate under cyclic thermomechanical loading; their design requires reliable prediction of the material's cyclic three-dimensional response and fatigue resistance. In this regard, several models exist that are capable of handling cyclic, mainly superelastic, SMA behavior; see, for instance, [Liu et al. 1999; Abeyaratne and Kim 1997; Lim and McDowell 2002; Tanaka et al. 1995; Lexcellent and Bourbon 1996; Bo and Lagoudas 1999; Zaki and Moumni 2007a].

The interesting properties of SMAs promoted their use in several fields, especially in outer space (antennas, braces) and in medical applications (orthodontia, cardiology, implant miniaturization, etc.). SMAs are also becoming increasingly attractive for automotive, nuclear, and civil engineering applications, mainly due to their high damping capacity.

One of the main difficulties facing their use in technologically advanced applications with high safety specifications is the poorly known fatigue behavior of these alloys; another is the amnesia phenomenon. Better knowledge and control of these two aspects should promote their use. Two types of fatigue have to be considered.

First, classical mechanical fatigue due to mechanical cycling in the pseudoelastic domain [Miyazaki et al. 1986; Van Humbeeck 1991]. The objective is to determine the number of cycles before failure. For instance, SMAs are used in the biomedical field to manufacture stents, endovascular prostheses inserted in blood vessels to avoid thrombosis and occlusion of the vessels. In stents, cyclic loads would arise from the difference in systolic and diastolic blood pressures and from the stress associated with the contraction of the heart muscle [McKelvey and Ritchie 1999]. It is of primary importance to know the number of cycles before any damage occurs to the stent.

Second, thermal fatigue or amnesia of the material due to a degradation of the material characteristics responsible for the shape memory effect like the transformation temperatures. The question is to determine if the material remains able to remember its initial shape.

Fatigue of shape memory alloys has also attracted considerable attention [Melton and Mercier 1979a; 1979b; 1979c; Wagner et al. 2004; Vaidyanathan et al. 2000; Porter et al. 2001]; it is still, however, not very well understood. In particular, fatigue mechanisms at the microscopic level are still being investigated [Siredey et al. 2005; Predki et al. 2006]. Nevertheless, Manson-Coffing-type criteria have been successfully used for predicting fatigue induced failure of simple SMA structures subjected to uniaxial loading [Melton and Mercier 1979a; Siredey et al. 2005; Wagner et al. 2004]. Like in classical elastoplastic materials, such as steel, the number of cycles to failure of a SMA varies depending on its composition and on the applied loading, among other factors. This number may range from 10^4 cycles for thermal valves using one-way shape memory effect [Eggeler et al. 2004] to a nominal 4×10^8 cycles for stents [Morgan 2004].

Using the analogy with plastic fatigue (low-cycle fatigue) [Halford 1966; Charkaluk et al. 2002], a relation was established in [Moumni et al. 2005] between the amount of dissipated energy associated with the stabilized hysteresis cycle and the number of cycles to failure. In this paper, we provide an application of the cyclic model, developed previously by the authors within the framework of generalized standards materials with internal constraints [Moumni et al. 2008], in order to simulate the dissipated energy at the stabilized cycle. Our aim here is to show that the cyclic model can be combined with the fatigue criterion in order to predict low-cycle failure of superelastic shape memory structures.

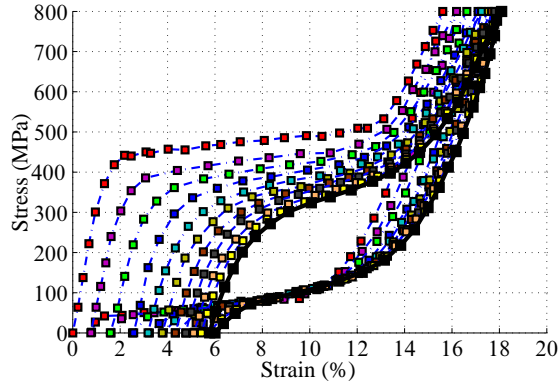


Figure 1. Cyclic superelastic tensile response of a NiTi wire. The characteristic hysteresis loop tends to stabilize when the number of cycles increases.

Section 2 is devoted to the presentation of the cyclic Zaki–Moumni model, where the behavior is described using three state variables representing residual strain induced by cyclic loading, internal stress induced by repeated phase change, and the cumulated martensite volume fraction. See [Zaki and Moumni 2007a; 2007b] for a detailed discussion of the model. Section 3 discusses the low cycle fatigue law and numerical simulations of the dissipated energy at the stabilized cycle, along with a validation against experimental data in the case of cyclic superelasticity.

2. The cyclic model of SMA behavior

2.1. Experimental observations. The response of a nickel-titanium wire to repeated tension, illustrated in Figure 1, displays the following features:

- Recovery of inelastic strain is not complete at the end of each cycle; after complete unloading, some residual strain remains. This strain increases exponentially with the number of cycles, as shown in Figure 2, left.
- Forward phase change yield stress decreases with increasing number of cycles (Figure 2, right).

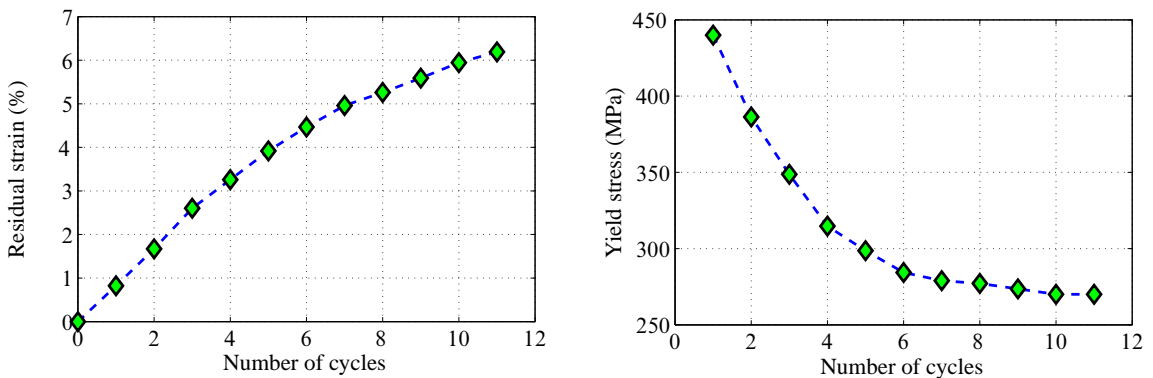


Figure 2. Left: residual strain versus number of tensile loading cycles. Right: forward phase change yield stress versus number of cycles.

- The hysteresis loop evolves progressively with the number of cycles before stabilizing. In [Figure 1](#), the stabilized loop is shown by a continuous dark line.

Residual strain is generally considered to be due to some oriented martensite not transforming back into austenite during reverse phase change [[Lexcelent and Bourbon 1996](#); [Auricchio et al. 2003](#)]. Repeated forward and reverse phase changes create some defects within the material [[Abeyaratne and Kim 1997](#)], which result in localized internal stresses [[Tanaka et al. 1995](#)], allowing SMAs to exhibit two-way shape memory. In this case the material is said to be trained (training phenomena). The internal stress eliminates the need for external loading in order to orient martensite variants. As a result, the shape memory structure can assume two different shapes when the temperature varies: an austenitic undeformed shape at high temperatures and a deformed low-temperature shape resulting from martensite orientation due to internal stress.

2.2. Phenomenological model. For full details regarding the model presented in this section, the reader is referred to [[Zaki 2006](#); [Zaki and Moumni 2007a](#); [2007b](#); [Moumni et al. 2008](#)].

As seen in the previous section, the macroscopic cyclic response of superelastic SMAs induces residual inelastic strains and localized internal stresses within the material. Hence, two state variables are introduced: a residual strain tensor $\boldsymbol{\varepsilon}_r$ and an internal stress tensor \mathbf{B} . A third variable z_e representing cumulated martensite volume fraction is also used:

$$z_e = \int_0^t |\dot{z}| d\tau,$$

where t is a kinematic time. The effect of cyclic loading on the material parameters can be modeled by considering these parameters to depend on z_e .

2.2.1. State variables and free energy. The following state variables are considered

- Macroscopic strain $\boldsymbol{\varepsilon}$ and temperature T .
- Volume fraction z and cumulated volume fraction z_e of martensite.
- Local strain tensors: $\boldsymbol{\varepsilon}_A$ for austenite and $\boldsymbol{\varepsilon}_M$ for martensite.
- Local martensite transformation strain tensor $\boldsymbol{\varepsilon}_{tr}$.
- Internal stress \mathbf{B} and residual strain $\boldsymbol{\varepsilon}_r$.

Phase change latent heat is assumed to depend on the cumulated fraction z_e . The respective free energy densities of austenite and martensite are taken to be

$${}^a W_A \stackrel{\text{def}}{=} W_A(\boldsymbol{\varepsilon}_A, \boldsymbol{\varepsilon}_r) = \frac{1}{2}(\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) : \mathbf{K}_A : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r),$$

and

$${}^m W_M = W_M(\boldsymbol{\varepsilon}_M, \boldsymbol{\varepsilon}_{tr}, \boldsymbol{\varepsilon}_r, T, \mathbf{B}, z_e) = \frac{1}{2}(\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{tr} - \boldsymbol{\varepsilon}_r) : \mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{tr} - \boldsymbol{\varepsilon}_r) + C(T, z_e) - \frac{2}{3}\mathbf{B} : \boldsymbol{\varepsilon}_{tr}.$$

$\mathbf{B} : \boldsymbol{\varepsilon}_{tr}$ in the expression of ${}^m W_M$ allows modeling the two-way shape memory effect. It represents a modification of the free energy of martensite due to the creation of internal stresses, which allows austenite to transform more easily into oriented martensite. Subsequent sections will help clarify this idea, especially when phase change criteria are established.

The contribution of austenite–martensite interaction to the SMA free energy density is assumed to be

$$\mathcal{F} = I(z, z_e, \boldsymbol{\varepsilon}_{\text{tr}}) = G \frac{z^2}{2} + \frac{z}{2} [\alpha z + \beta(1-z)] \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right).$$

G , α and β are material parameters functions of the cumulated martensite volume fraction z_e ($G = \hat{G}(z_e)$, $\alpha = \hat{\alpha}(z_e)$, and $\beta = \hat{\beta}(z_e)$)

- $\beta \frac{z(1-z)}{2} \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right)$ represents interaction between austenite and martensite. Following many published works [Leclercq and Lexcellent 1996; Raniecki et al. 1992; Patoor and Berveiller 1993, etc.], this interaction is taken to be proportional to the volume fractions of interacting phases. β determines how a mechanical loading applied to an initially austenitic shape memory material affects martensite orientation during phase change.
- $G \frac{z^2}{2}$ quantifies orientation-independent interaction between martensite variants.
- $\alpha \frac{z^2}{2} \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right)$ accounts for the interaction increase due to orientation of martensite plates; its expression is similar to that of the energy contribution due to linear kinematic hardening of an elastoplastic material with hardening coefficient α . α controls the slope of the stress-strain curve corresponding to martensite orientation.

Finally, the free energy density of the material is given by

$$\begin{aligned} \mathcal{W} &\stackrel{\text{def}}{=} W(\boldsymbol{\varepsilon}, T, \boldsymbol{\varepsilon}_A, \boldsymbol{\varepsilon}_M, z, \boldsymbol{\varepsilon}_{\text{tr}}, \boldsymbol{\varepsilon}_r, \mathbf{B}, z_e) \\ &= (1-z) \left[\frac{1}{2} (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) : \mathbf{K}_A : \boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r \right] + z \left[\frac{1}{2} (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) : \mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) + C(T, z_e) \right] \\ &\quad + \frac{1}{2} G z^2 + \frac{1}{2} z [\alpha z + \beta(1-z)] \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right) - \frac{2}{3} z \mathbf{B} : \boldsymbol{\varepsilon}_{\text{tr}}. \quad (1) \end{aligned}$$

2.2.2. Internal constraints and Lagrangian. State variables obey the following constraints

- Martensite volume fraction is necessarily within the $[0, 1]$ interval.

$$z \geq 0, \quad (1-z) \geq 0. \quad (2)$$

- The equivalent transformation strain $\sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}}$ has a maximum value γ that varies with respect to the cumulated volume fraction z_e

$$\gamma - \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}} \geq 0, \quad \gamma \stackrel{\text{def}}{=} \hat{\gamma}(z_e). \quad (3)$$

Constraints given by (2) and (3) are assumed to be perfect. They derive from a constraints potential

$$\mathcal{W}_l = -\lambda : [(1-z)\boldsymbol{\varepsilon}_A + z\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}] - \mu \left(\gamma - \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}} \right) - \nu_1 z - \nu_2 (1-z),$$

where λ , ν_1 , ν_2 , and μ are Lagrange multipliers. ν_1 , ν_2 and μ , associated with unilateral constraints, must obey the following conditions

$$\nu_1 \geq 0, \nu_1 z = 0, \quad \nu_2 \geq 0, \nu_2 (1-z) = 0 \quad \text{and} \quad \mu \geq 0, \mu \left(\gamma - \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}} \right) = 0. \quad (4)$$

The sum of the free energy \mathcal{W} and of the constraints potential \mathcal{W}_l gives the Lagrangian

$$\begin{aligned} \mathcal{L} &= \mathcal{W} + \mathcal{W}_l \stackrel{\text{def}}{=} L(\boldsymbol{\varepsilon}, T, \boldsymbol{\varepsilon}_A, \boldsymbol{\varepsilon}_M, z, \boldsymbol{\varepsilon}_{\text{tr}}, \boldsymbol{\varepsilon}_r, \mathbf{B}, z_e) \\ &= (1-z) \left[\frac{1}{2} (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) : \mathbf{K}_A : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) \right] + z \left[\frac{1}{2} (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) : \mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) + C(T, z_e) - \frac{2}{3} \mathbf{B} : \boldsymbol{\varepsilon}_{\text{tr}} \right] \\ &\quad + G \frac{1}{2} z^2 + \frac{1}{2} z [\alpha z + \beta (1-z)] \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right) \\ &\quad - \boldsymbol{\lambda} : [(1-z)\boldsymbol{\varepsilon}_1 + z\boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}] - \mu \left(\gamma - \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}} \right) - \nu_1 z - \nu_2 (1-z), \end{aligned} \quad (5)$$

where conditions (4) must be met.

2.2.3. State equations. Phase change, martensite orientation, training, as well as the creation of residual strain and internal stress are dissipative processes. Thus, if \mathcal{A}_z , \mathcal{A}_{tr} , \mathcal{A}_e , \mathcal{A}_r , and \mathcal{A}_B represent thermodynamic forces associated with state variables z , $\boldsymbol{\varepsilon}_{\text{tr}}$, z_e , $\boldsymbol{\varepsilon}_r$, and \mathbf{B} respectively; only these forces may take nonzero values during a given transformation. Hence the state equations

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}} = \boldsymbol{\sigma} \quad \Rightarrow \quad \boldsymbol{\lambda} - \boldsymbol{\sigma} = \mathbf{0}, \quad (6)$$

$$-\frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}_A} = \mathbf{0} \quad \Rightarrow \quad (1-z) [\mathbf{K}_A : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) - \boldsymbol{\lambda}] = \mathbf{0}, \quad (7)$$

$$-\frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}_M} = \mathbf{0} \quad \Rightarrow \quad z [\mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) - \boldsymbol{\lambda}] = \mathbf{0}, \quad (8)$$

$$\begin{aligned} -\frac{\partial \mathcal{L}}{\partial z} = \mathcal{A}_z \quad \Rightarrow \quad \mathcal{A}_z &= \frac{1}{2} [(\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) : \mathbf{K}_A : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) - (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) : \mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r)] \\ &\quad - C(T, z_e) - Gz - [(\alpha - \beta)z + \frac{1}{2}\beta] \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right) - \boldsymbol{\lambda} : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_M) + \frac{2}{3} \mathbf{B} : \boldsymbol{\varepsilon}_{\text{tr}}, \end{aligned}$$

$$-\frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}_{\text{tr}}} = \mathcal{A}_{\text{tr}} \quad \Rightarrow \quad \mathcal{A}_{\text{tr}} = z \{ \mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r) - \frac{2}{3} [\alpha z + \beta (1-z)] \boldsymbol{\varepsilon}_{\text{tr}} \} + \frac{2}{3} z \mathbf{B} - \frac{2}{3} \mu \boldsymbol{\varepsilon}_{\text{tr}} / \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}},$$

$$-\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \mathbf{0} \quad \Rightarrow \quad (1-z)\boldsymbol{\varepsilon}_A + z\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon} = \mathbf{0}, \quad (9)$$

$$-\frac{\partial \mathcal{L}}{\partial \boldsymbol{\varepsilon}_r} = \mathcal{A}_r \quad \Rightarrow \quad \mathcal{A}_r = (1-z)\mathbf{K}_A : (\boldsymbol{\varepsilon}_A - \boldsymbol{\varepsilon}_r) + z\mathbf{K}_M : (\boldsymbol{\varepsilon}_M - \boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r),$$

$$-\frac{\partial \mathcal{L}}{\partial \mathbf{B}} = \mathcal{A}_B \quad \Rightarrow \quad \mathcal{A}_B = \frac{2}{3} z \boldsymbol{\varepsilon}_{\text{tr}},$$

$$-\frac{\partial \mathcal{L}}{\partial z_e} = \mathcal{A}_e \quad \Rightarrow \quad \mathcal{A}_e = -z \frac{\partial C(T, z_e)}{\partial z_e} - \frac{\partial G}{\partial z_e} \frac{z^2}{2} - \frac{z}{2} \left[\frac{\partial \alpha}{\partial z_e} z + \frac{\partial \beta}{\partial z_e} (1-z) \right] \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}} \right) - \mu \frac{\partial \gamma}{\partial z_e}.$$

Equations (6), (7), (8), and (9) allow us to establish the stress-strain relation

$$\boldsymbol{\sigma} = \mathbf{K} : (\boldsymbol{\varepsilon} - z\boldsymbol{\varepsilon}_{\text{tr}} - \boldsymbol{\varepsilon}_r).$$

\mathbf{K} , the equivalent SMA elastic moduli tensor, is given by

$$\mathbf{K} = [(1-z)\mathbf{K}_A^{-1} + z\mathbf{K}_M^{-1}]^{-1}.$$

2.2.4. Yield functions and evolution laws. Residual strain and internal stress depend on the number of loading cycles. Indeed, as shown in [Figure 2](#), left, residual strain is found to increase exponentially with respect to the number of cycles up to an asymptotic value $\boldsymbol{\varepsilon}_r^{\text{sat}}$. From a theoretical point of view, this can be simulated using the evolution law

$$\dot{\boldsymbol{\varepsilon}}_r = \frac{\varepsilon_r^{\text{sat}}}{\tau} \left(\frac{3}{2} \frac{\mathbf{s}}{\sigma_{\text{VM}}} \right) \exp\left(-\frac{z_e}{\tau}\right) \dot{z}_e.$$

Residual strain is considered to be deviatoric and so it does not induce any volume change. τ is a time constant, $\varepsilon_r^{\text{sat}}$ is the maximum residual strain in tension when the hysteresis loop stabilizes, \mathbf{s} is the deviatoric part of the stress tensor, and σ_{VM} is the equivalent Von Mises measure on the stress tensor. Dissipation due to the evolution of $\boldsymbol{\varepsilon}_r$ is necessarily positive. Indeed, it can easily be shown that \mathcal{A}_r is unconditionally equal to the stress tensor $\boldsymbol{\sigma}$. $\mathcal{A}_r : \dot{\boldsymbol{\varepsilon}}_r$ is therefore positive. Similarly, the evolution of internal stress \mathbf{B} is assumed to be governed by the equation

$$\dot{\mathbf{B}} = \frac{B^{\text{sat}}}{\tau} \left(\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} / \sqrt{\frac{2}{3} \boldsymbol{\varepsilon}_{\text{tr}} : \boldsymbol{\varepsilon}_{\text{tr}}} \right) \exp\left(-\frac{z_e}{\tau}\right) \dot{z}_e, \quad (10)$$

where B^{sat} is a positive scalar.

[Equation \(10\)](#) expresses an increase in equivalent internal stress with respect to the number of cycles. Given the expression of the thermodynamic force \mathcal{A}_B associated with state variable \mathbf{B} , dissipation $\mathcal{A}_B : \dot{\mathbf{B}}$ is positive. Because the evolution of state variables $\boldsymbol{\varepsilon}_r$, \mathbf{B} , and z_e is related to that of the martensite volume fraction, one does not need to define specific yield functions for each of these variables. Nevertheless, three yield functions: \mathcal{F}_z^1 , \mathcal{F}_z^2 , and \mathcal{F}_{ori} are needed in order to describe forward phase change, reverse phase change, and orientation of martensite variants. Thermodynamic forces \mathcal{A}_z and \mathcal{A}_{tr} are chosen to be such that

$$\mathcal{A}_z \in \partial_z \mathcal{D}, \quad \mathcal{A}_{\text{tr}} \in \partial_{\dot{\boldsymbol{\varepsilon}}_{\text{tr}}} \mathcal{D}, \quad (11)$$

where \mathcal{D} is a convex positive continuous function that is equal to zero at the origin:

$$\mathcal{D} \stackrel{\text{def}}{=} D(\dot{z}, \dot{\boldsymbol{\varepsilon}}_{\text{tr}}) = P(z, z_e, \dot{z}) \dot{z} + R(z) \sqrt{\frac{2}{3} \dot{\boldsymbol{\varepsilon}}_{\text{tr}} : \dot{\boldsymbol{\varepsilon}}_{\text{tr}}}, \quad (12)$$

with

$$P(z, z_e, \dot{z}) \stackrel{\text{def}}{=} [a(1-z) + bz] \text{sign } \dot{z},$$

for certain z_e -dependent parameters $a = \hat{a}(z_e)$ and $b = \hat{b}(z_e)$. Moreover, $R(z) = z^2 Y$, where Y is a constant material parameter. It is easy to show that Y , a , and b are always positive; \mathcal{D} is therefore positive. Hence, evolution equations [\(11\)](#) necessarily satisfy the Clausius–Duhem inequality.

From this point on, austenite and martensite are considered to be homogeneous and isotropic media, having the same Poisson coefficient ν . Specifically,

$$\nu_A = \nu_M \stackrel{\text{def}}{=} \nu.$$

[Table 1](#) summarizes the notations used throughout this paper. Using equations [\(11\)](#), one can establish

| Meaning or Expression | | Meaning | Expression |
|-----------------------|-----------------------------|--------------------------|---|
| E_A | Young modulus of austenite | E_{eq} | Equivalent Young modulus $\left(\frac{1-z}{E_A} + \frac{z}{E_M}\right)^{-1}$ |
| E_M | Young modulus of martensite | $\text{tr } \mathbf{M}$ | Trace of symmetrical tensor \mathbf{M} $\sum_i \mathbf{M}_{ii}$ |
| ν | Poisson coefficient | $\text{dev } \mathbf{M}$ | Deviator of \mathbf{M} $\mathbf{M} - \frac{1}{3}(\text{tr } \mathbf{M}) \mathbf{I}$ |
| El_A | $(1+\nu)/E_A$ | M_{VM} | Von Mises equivalent of \mathbf{M} $\sqrt{\frac{3}{2} \text{dev } \mathbf{M} : \text{dev } \mathbf{M}}$ |
| El_M | $(1+\nu)/E_M$ | \mathbf{s} | Stress deviator tensor $\text{dev } \boldsymbol{\sigma}$ |
| P_A | $-\nu/E_A$ | σ_{VM} | Von Mises equivalent stress $\sqrt{\frac{3}{2} \mathbf{s} : \mathbf{s}}$ |
| P_M | $-\nu/E_M$ | μ_A | Austenite shear modulus $\frac{1}{2} E_A / (1+\nu)$ |
| El_{MA} | $El_M - El_A$ | μ_M | Martensite shear modulus $\frac{1}{2} E_M / (1+\nu)$ |
| P_{MA} | $P_M - P_A$ | μ_{eq} | Equivalent shear modulus $\frac{1}{2} E_{eq} / (1+\nu)$ |

Table 1. Notations used in this paper.

expressions for \mathcal{F}_z^1 , \mathcal{F}_z^2 , and \mathcal{F}_{ori} :

$$\mathcal{F}_z^1 = \left\{ \frac{1}{3} El_{MA} \sigma_{VM}^2 + \frac{1}{2} \left(\frac{1}{3} El_{MA} + P_{MA} \right) (\text{tr } \boldsymbol{\sigma})^2 - C(T, z_e) \right\} \\ + \left(\boldsymbol{\sigma} + \frac{2}{3} \mathbf{B} \right) : \boldsymbol{\epsilon}_{tr} - (G+b)z - a(1-z) - \left[(\alpha - \beta)z + \frac{1}{2}\beta \right] \left(\frac{2}{3} \boldsymbol{\epsilon}_{tr} : \boldsymbol{\epsilon}_{tr} \right),$$

$$\mathcal{F}_z^2 = - \left\{ \frac{1}{3} El_{MA} \sigma_{VM}^2 + \frac{1}{2} \left(\frac{1}{3} El_{MA} + P_{MA} \right) (\text{tr } \boldsymbol{\sigma})^2 - C(T, z_e) \right\} \\ - \left(\boldsymbol{\sigma} + \frac{2}{3} \mathbf{B} \right) : \boldsymbol{\epsilon}_{tr} + (G-b)z - a(1-z) + \left[(\alpha - \beta)z + \frac{1}{2}\beta \right] \left(\frac{2}{3} \boldsymbol{\epsilon}_{tr} : \boldsymbol{\epsilon}_{tr} \right),$$

$$\mathcal{F}_{ori} = \left\| \left(\boldsymbol{\sigma} + \frac{2}{3} \mathbf{B} \right) - \frac{2}{3} [\alpha z + \beta(1-z)] \boldsymbol{\epsilon}_{tr} - \frac{2\mu}{3z} \boldsymbol{\epsilon}_{tr} / \sqrt{\frac{2}{3} \boldsymbol{\epsilon}_{tr} : \boldsymbol{\epsilon}_{tr}} \right\|_{VM} - zY.$$

It is worth noting that in each of these expressions, a quantity $\frac{2}{3} \mathbf{B}$, proportional to internal stress due to training, is added to the stress tensor $\boldsymbol{\sigma}$. Thus both \mathbf{B} and $\boldsymbol{\sigma}$ have similar effects on phase change and martensite orientation.

Phase change evolution laws obey certain conditions

- If $\mathcal{F}_z^1 < 0$ and $\mathcal{F}_z^2 < 0$, no phase change can occur. Hence, $\dot{z} = 0$.
- If forward phase change is triggered, \mathcal{F}_z^1 is equal to zero. In this case, \dot{z} is equal to zero if $\dot{\mathcal{F}}_z^1 < 0$; otherwise, \dot{z} is given by the consistency condition $\dot{\mathcal{F}}_z^1 = 0$.
- If reverse phase change is triggered, \mathcal{F}_z^2 is equal to zero. In this case, \dot{z} is equal to zero if $\dot{\mathcal{F}}_z^2 < 0$; otherwise, \dot{z} is given by the consistency condition $\dot{\mathcal{F}}_z^2 = 0$.

Let

$$\mathbf{X} \stackrel{\text{def}}{=} \left(\mathbf{s} + \frac{2}{3} \mathbf{B} \right) - \frac{2}{3} [\alpha z + \beta(1-z)] \boldsymbol{\epsilon}_{tr} - \frac{2\mu}{3z} \boldsymbol{\epsilon}_{tr} / \sqrt{\frac{2}{3} \boldsymbol{\epsilon}_{tr} : \boldsymbol{\epsilon}_{tr}}.$$

The yield function \mathcal{F}_{ori} , associated with martensite orientation, can be written as $\mathcal{F}_{\text{ori}} = X_{\text{VM}} - zY$. Evolution of the local inelastic strain tensor $\boldsymbol{\epsilon}_{\text{tr}}$ satisfies the normality law:

$$\dot{\boldsymbol{\epsilon}}_{\text{tr}} = \eta \frac{\partial \mathcal{F}_{\text{ori}}}{\partial \mathbf{X}} = \frac{3}{2} \eta \frac{\mathbf{X}}{X_{\text{VM}}}. \quad (13)$$

Here η is a positive scalar satisfying the Kuhn–Tucker conditions

$$\eta \geq 0, \quad \mathcal{F}_{\text{ori}} \leq 0 \quad \text{and} \quad \eta \mathcal{F}_{\text{ori}} = 0.$$

Let σ_{rs} and σ_{rf} be orientation start and finish stresses of self-accommodating martensite

- When orientation starts, the yield function \mathcal{F}_{ori} is necessarily equal to zero for $\mu = 0$, $\boldsymbol{\epsilon}_{\text{tr}} = \mathbf{0}$, and $\|\boldsymbol{\sigma} + \mathbf{B}\|_{\text{VM}} = \sigma_{\text{rs}}$. It follows that

$$Y = \sigma_{\text{rs}}; \quad (14)$$

- When orientation is complete, $\mathcal{F}_{\text{ori}} = 0$ for $\sqrt{\frac{2}{3}} \boldsymbol{\epsilon}_{\text{tr}} : \boldsymbol{\epsilon}_{\text{tr}} = \gamma$. If $\sqrt{\frac{2}{3}} \boldsymbol{\epsilon}_{\text{tr}} : \boldsymbol{\epsilon}_{\text{tr}}$ tends towards γ with lower values, μ is equal to zero for $\|\boldsymbol{\sigma} + \mathbf{B}\|_{\text{VM}} = \sigma_{\text{rf}}$. In case of uniaxial tension, it follows that

$$|\sigma_{\text{rf}} - \alpha \gamma| = Y;$$

the above equation, together with Equation (14), gives α as a function of γ

$$\alpha = \frac{\sigma_{\text{rf}} - \sigma_{\text{rs}}}{\gamma}.$$

- When austenite transforms into oriented martensite, orientation is complete when the stress becomes greater or equal to σ_{rf} . Particularly, if the stress tends towards σ_{rf} with lower values, μ remains equal to zero. In this case,

$$z = 0, \quad \|\boldsymbol{\sigma} + \mathbf{B}\|_{\text{VM}} = \sigma_{\text{rf}}, \quad \sqrt{\frac{2}{3}} \boldsymbol{\epsilon}_{\text{tr}} : \boldsymbol{\epsilon}_{\text{tr}} = \gamma, \quad \text{and} \quad \mu = 0.$$

In case of uniaxial tension

$$\beta = \frac{\sigma_{\text{rf}}}{\gamma}.$$

α and β are both functions of z_e due to their dependence on γ . Finally, explicit expressions of the evolution laws can be derived using the consistency conditions.

2.3. Numerical simulation. Figure 3, left shows the experimental response of a NiTi test sample to repeated uniaxial tension. Residual strain evolution with respect to the number of cycles is shown in Figure 3, right. The nickel-titanium used in the experiments has an orientation start stress of 80 MPa and an orientation finish stress equal to 160 MPa. The reverse transformation finish temperature of the untrained material at zero stress is equal to 42°C; this temperature does not evolve considerably with the number of cycles. Figure 4 shows good agreement between experimental and numerical results in the case of repeated tension. For clarity, only loops 2, 4, and 12 are shown. Figure 5 (left) illustrates the evolution of internal strain B with respect to the cumulated martensite volume fraction z_e . Stress-strain response for all 20 loading cycles is shown in Figure 5 (right).

The next section is devoted to the presentation of an energy approach to the fatigue of shape memory alloys. It is shown how the cyclic model presented in the previous section can be combined with the

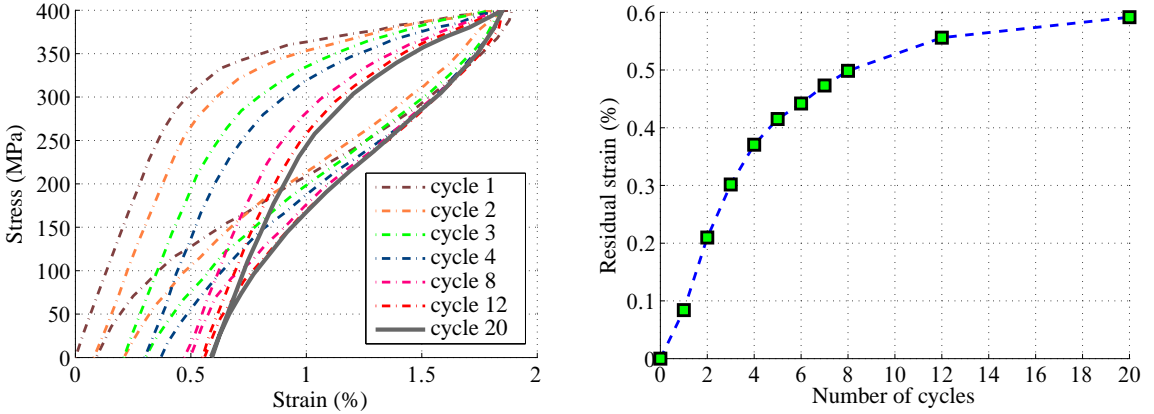


Figure 3. Left: experimental stress-strain response of a NiTi wire to repeated tension. The stabilized cycle is shown by a continuous line. Right: residual strain evolution with respect to the number of tension cycles.

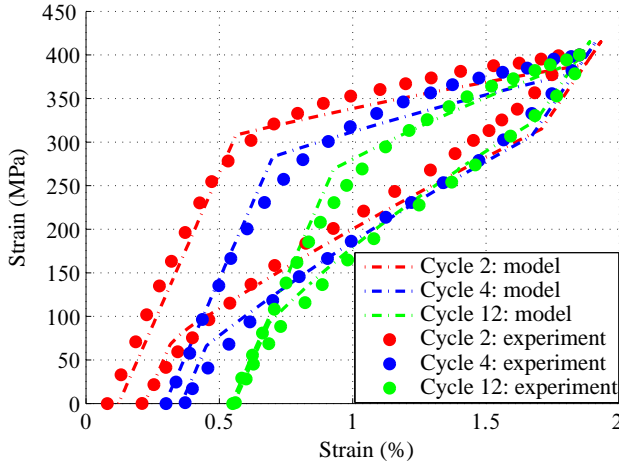


Figure 4. Experimental versus numerical results: evolution of the superelastic loop.

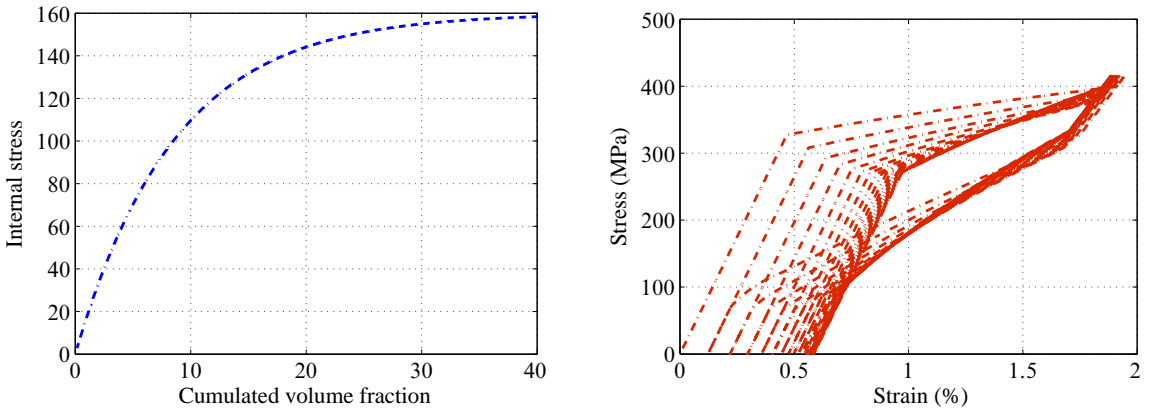


Figure 5. Prediction of nickel-titanium response to repeated tension with the evolution of internal stress (left) and the stress-strain response cycles (right).

fatigue criterion in order to perform numerical calculations of the fatigue parameters necessary for the evaluation of the life time of structures made on SMAs.

3. An energy approach to the fatigue of shape memory alloys

Fatigue of shape memory alloys is generally explained by the creation and propagation of defects within the material at the microscopic level [Melton and Mercier 1979a]. A more rigorous understanding is complicated, however, due to phenomena like formation of residual martensite [Siredey et al. 2005] and local phase changes at the tip of microscopic cracks [Wagner et al. 2004] resulting in slower crack propagation in martensite [Eggeler et al. 2004]. Because of similar damage creation and propagation mechanisms, it is interesting to investigate the fatigue of SMAs within a framework similar to that of usual elastoplastic materials (like steel). It is, hence, useful to distinguish between low-cycle fatigue, finite fatigue life in high-cycle fatigue, and high-cycle fatigue (infinite life). We will focus on low-cycle fatigue associated with cyclic superelasticity.

3.1. Experimental analysis.

3.1.1. Material and thermomechanical treatment. The material is a 51.3% Ti–48.7% Ni in mass nickel-titanium with a grain size between $60\ \mu\text{m}$ and $70\ \mu\text{m}$ (see Figure 6). The phase change temperatures at zero stress are as follows $M_f^0 = 25^\circ\text{C}$, $M_s^0 = 39^\circ\text{C}$, $A_s^0 = 29^\circ\text{C}$ and $A_f^0 = 42^\circ\text{C}$. All the test specimens

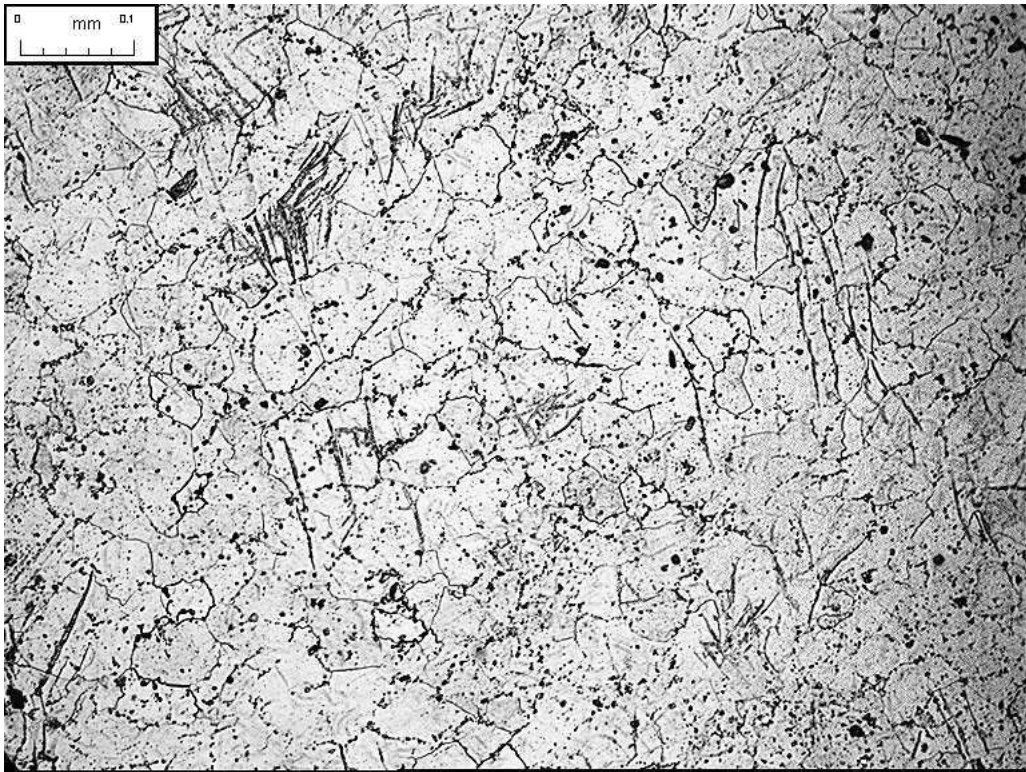


Figure 6. Metallographic structure of the NiTi alloy.

were worked up to 20% in tension while cold, then heat treated at 400°C for one hour. This kind of treatment increases the plastic yield limit of the material while improving its superelasticity [Wayman and Otsuka 1999].

3.1.2. Experimental setup. The testing machine used is a force controlled MTS810/100KN. The strains are measured using an MTS extensometer model 632-13C-21. Load and extensometer signals are captured by an MTS TestStarII data acquisition board and processed by a computer. Experiments are Stress-controlled push-pull tests with a constant amplitude σ_a . They are carried out at a frequency of 0.3 Hz. In order to ensure that the tests are performed in the pseudoelastic domain the temperature is kept constant at $T = 50^\circ\text{C}$ which is higher than the austenite finish temperature A_f^0 . The experiments were conducted in a SERVATHIN hermetic enclosure where the temperature can be regulated and kept constant through a range values from -50°C to 200°C . Maximum stress was kept below the critical stress for slip (750 MPa) to ensure that no macroscopic plastic deformations occurs. The tests were taken through to rupture of the specimen.

In order to examine the effect of mean stress on the fatigue of nickel-titanium, three load ratios ($R = \sigma_{\min}/\sigma_{\max}$) were considered, equal to 0, 0.2, and -1 respectively. The geometry of the specimens we used is illustrated in Figure 7.

3.2. Results and discussions.

3.2.1. S-N curves. The S-N curves (Wöhler curves) relating the number of cycles to failure under uniaxial loading to the amplitude of the applied stress obtained for these experiments are shown on Figure 8. The effect of mean stress on fatigue life can be observed. Indeed, a higher mean stress corresponds to a lower number of cycles to failure. These Wöhler curves can be used for fatigue life prediction when the applied loading is uniaxial and for a given mean stress. They are inadequate, however, for fatigue analysis of shape memory structures under multiaxial loading.

3.2.2. Low-cycle fatigue life prediction for superelastic SMAs. Existing SMA low-cycle fatigue life prediction models are mostly of the Manson–Coffin type where the number of cycles to failure is related to the amplitude of plastic strain. As early as 1979, [Melton and Mercier 1979a] showed that, for different types of shape memory alloys, fatigue life of wires follows the Manson–Coffin law. This result has been confirmed in several subsequent papers [Tolomeo et al. 2000; Wagner et al. 2004]. Even though multiaxial loading can be accounted for theoretically by means of a generalized Manson–Coffin relationship using

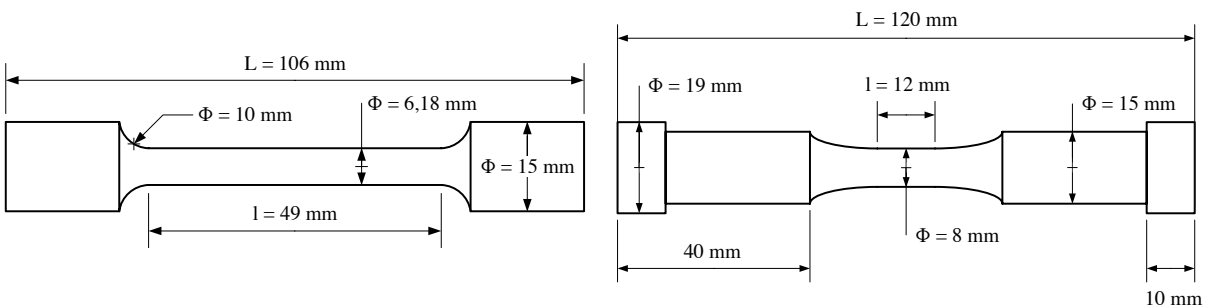


Figure 7. Geometries of specimens used for cyclic experiments for $R = 0$ (left) and $R = 0.2, -1$ (right), respectively.

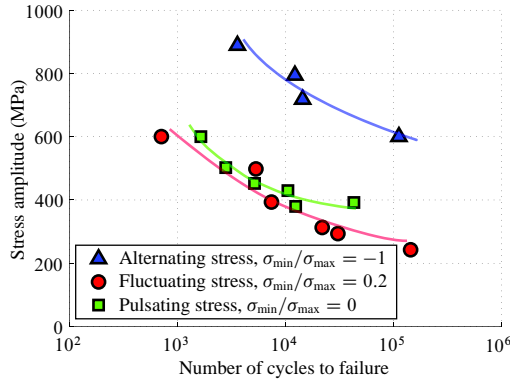


Figure 8. Wöhler curves of nickel-titanium for different mean stress values.

equivalent plastic strain, the ability to predict three-dimensional structure failure using this approach has not been proved.

In the scope of this paper, an energy-based approach is used for estimating low-cycle fatigue life of superelastic SMAs. This is inspired from [Charkaluk et al. \[2002\]](#), where a similar approach was successfully applied on cast iron. Indeed, superelastic hysteresis stabilization for a shape memory material is similar to plastic shakedown of cast iron. In both cases, inelastic (plastic deformation for iron cast and transformation strain for SMAs) deformation is confined but the material continues to dissipate energy. Energy dissipation is usually explained, in the case of superelastic SMAs, by strain incompatibilities across the boundaries of the grains [[Melton and Mercier 1979b](#)] which in time lead to the creation and accumulation of defects at the boundaries [[Melton and Mercier 1979a](#)].

In this paper, the martensitic transformation, responsible for the creation of martensite grains, is accounted for using state variables z and ε_{tr} . Given the above interpretation of energy dissipation in SMAs, the inelastic deformation $z\varepsilon_{tr}$, being proportional to the number of martensite grain boundaries (through state variable z) and to the level of orientation of martensite variants within these grains, seems an adequate parameter for predicting fatigue failure of superelastic SMAs.

The amount of dissipated energy per loading cycle, wW_d , is given by

$${}^wW_d = \oint \boldsymbol{\sigma} : d\boldsymbol{\varepsilon}.$$

Once the material response stabilizes, this energy becomes constant, as shown by the experimental result given in [Figure 9](#), left (stress ratio = 0).

In the case of the studied nickel-titanium, an example of the evolution of the dissipated energy with respect to the number of cycles is illustrated in [Figure 9](#), right. Even though the use of dissipated energy per cycle for fatigue life prediction has been criticized [[Halford 1966](#); [Pineau and Pétrequin 1997](#)], its usefulness in practice is well proven. The work of [[Charkaluk et al. 2002](#)] has been successfully applied to predicting failure of automotive components subjected to complex thermomechanical loading.

The data points in [Figure 10](#), left, represent the amount of dissipated energy per cycle wW_d with respect to the number of cycles to failure N_f , plotted on a log-log scale. The figure shows a quasilinear dependence of $\log {}^wW_d$ on $\log N_f$ for several values of the mean stress. Hence it is interesting to approximate

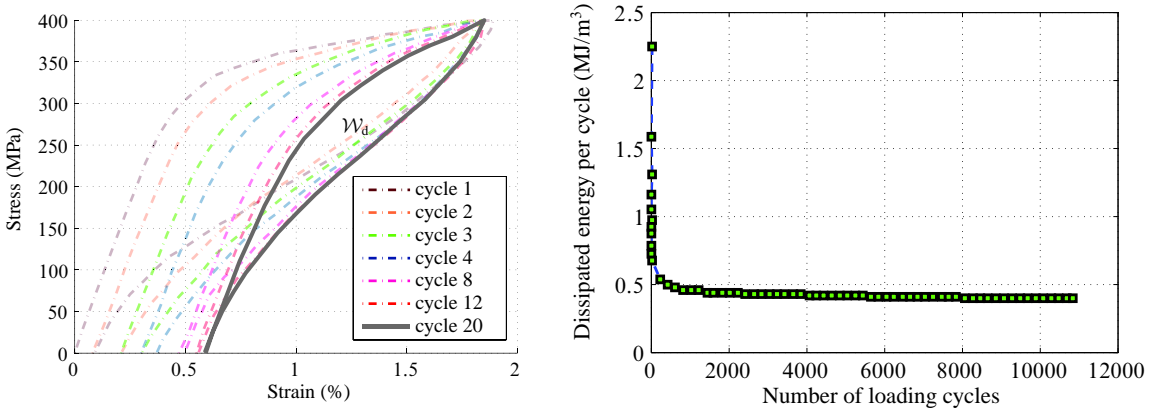


Figure 9. Left: experimental stabilization of the amount of dissipated energy per cycle, \mathcal{W}_d , with the number of cycles. Right: dissipated energy per cycle versus number of loading cycles: experimental result for $\Delta\sigma = 400\text{Mpa}$.

the experimental results using the following curve from [Moumni et al. 2005]:

$$\mathcal{W}_d = \alpha N_f^\beta, \tag{15}$$

where α and β are material parameters. This is shown as the gray line in the same figure. Expression (15) can readily be used for three-dimensional structure analysis because dissipated energy per cycle is well defined and its calculation is straightforward.

The stabilized cycle and the corresponding amount dissipated energy to the stabilized cycle can be numerically determined using the model presented in this paper as shown in Figure 10, right. The number of cycles to failure of a SMA structure can be estimated using the suggested fatigue criterion.

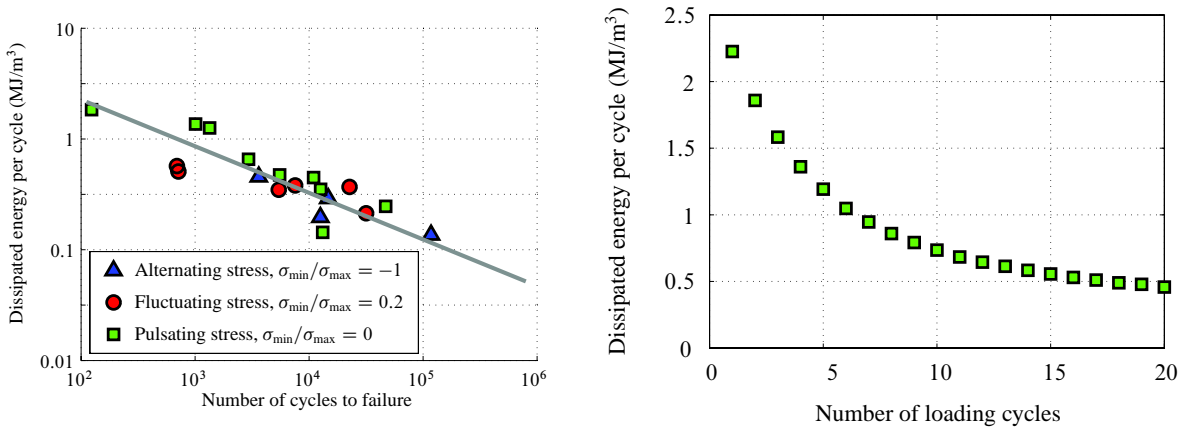


Figure 10. Left: experimental results for dissipated energy per cycle as a function of the number of cycles to failure, for $\alpha = 11$ and $\beta = -0.377$, and fit using Equation (15). Right: numerical results for dissipated energy per cycle with respect to the number of cycles, with $\Delta\sigma = 400\text{Mpa}$.

It is important to note, however, that the validity of this criterion has only been proven in the case of uniaxial loading; its ability to predict failure of structures subjected to complicated loading conditions remains to be established.

4. Conclusion

In the first part of this paper, a model capable of simulating several phenomena associated with shape memory materials subjected to cyclic loading is presented. The modeling process is based on a simple observation. On the macroscopic level, SMA training can be interpreted as a thermomechanically-induced transition from an unstable virgin material configuration into a stable one. From a theoretical point of view, it is easy to account for this transition by making some of the model parameters depend on a cumulated martensite volume fraction which evolves with the applied loading. Inelastic residual strain, which appear during repeated phase change, is accounted for by introducing a state variable similar to plastic deformation strain of classical elastoplastic materials. Numerical results show good agreement with available experimental data.

The second part of the paper investigates the fatigue of SMAs by analogy with plastic fatigue. It has been shown that the dissipated energy at the stabilized cycle during a cyclic loading is a relevant parameter for fatigue life prediction. A relationship between this parameter and the number of cycles to failure has been derived from experimental results. It has also been shown that the cyclic model can be combined with the fatigue criterion in order to predict low-cycle failure of superelastic shape memory structures. Nevertheless, it is clear that the model must be improved on some points. First, it is important to investigate the fatigue of SMAs criterion for another type of loading, namely torsion. Second, it is interesting to check the validity of the model against experimental results for complex structures under complex loading. This work is undertaken and will be presented in future papers.

References

- [Abeyaratne and Kim 1997] R. Abeyaratne and S.-J. Kim, “Cyclic effects in shape-memory alloys: a one-dimensional continuum model”, *Int. J. Solids Struct.* **34**:25 (1997), 3273–3289.
- [Auricchio et al. 2003] F. Auricchio, S. Marfia, and E. Sacco, “Modelling of SMA materials: training and two way memory effects”, *Comput. Struct.* **81**:24-25 (2003), 2301–2317.
- [Bo and Lagoudas 1999] Z. Bo and D. C. Lagoudas, “Thermomechanical modeling of polycrystalline SMAs under cyclic loading, I: theoretical derivations”, *Int. J. Eng. Sci.* **37**:9 (1999), 1089–1140.
- [Charkaluk et al. 2002] E. Charkaluk, A. Bignonnet, A. Constantinescu, and K. D. Van, “Fatigue design of structures under thermomechanical loadings”, *Fatigue Fract. Eng. Mater. Struct.* **25**:12 (2002), 1199–1206.
- [Eggeler et al. 2004] G. Eggeler, E. Hornbogen, A. Yawny, A. Heckmann, and M. Wagner, “Structural and functional fatigue of NiTi shape memory alloys”, *Mater. Sci. Eng. A* **378**:1-2 (2004), 24–33.
- [Halford 1966] G. Halford, “The energy required for fatigue”, *J. Mater.* **1** (1966), 3–18.
- [Leclerq and LExcellent 1996] S. Leclerq and C. LExcellent, “A general macroscopic description of the thermomechanical behavior of shape memory alloys”, *J. Mech. Phys. Solids* **44**:6 (1996), 953–957.
- [LExcellent and Bourbon 1996] C. LExcellent and G. Bourbon, “Thermodynamical model of cyclic behaviour of ti-ni and cu-zn-al shape memory alloys under isothermal undulated tensile tests”, *Mech. Mater.* **24**:1 (1996), 59–73.
- [Lim and McDowell 2002] T. J. Lim and D. L. McDowell, “Cyclic thermomechanical behavior of a polycrystalline pseudoelastic shape memory alloy”, *J. Mech. Phys. Solids* **50**:3 (2002), 651–676.

- [Liu et al. 1999] Y. Liu, Z. Xie, and J. Van Humbeeck, “Cyclic deformation of NiTi shape memory alloys”, *Mater. Sci. Eng. A* **273–275** (1999), 673–678.
- [McKelvey and Ritchie 1999] A. L. McKelvey and R. O. Ritchie, “Fatigue-crack propagation in Nitinol, a shape-memory and superelastic endovascular stent material”, *J. Biomed. Mater. Res.* **47:3** (1999), 301–308.
- [Melton and Mercier 1979a] K. N. Melton and O. Mercier, “The effect of the martensitic phase transformation on the low cycle fatigue behaviour of polycrystalline Ni–Ti and Cu–Zn–Al alloys”, *Mat. Sci. Eng.* **40:1** (1979), 81–87.
- [Melton and Mercier 1979b] K. N. Melton and O. Mercier, “Fatigue life of cuznal alloys”, *Scr. Metall. Mater.* **13:1** (1979), 73–75.
- [Melton and Mercier 1979c] K. N. Melton and O. Mercier, “Fatigue of NiTi thermoelastic martensites”, *Acta Metall.* **27:1** (1979), 137–144.
- [Miyazaki et al. 1986] S. Miyazaki, T. Imai, Y. Igo, and K. Otsuka, “Effect of cyclic deformation on the pseudoelasticity characteristics of Ti-Ni alloys”, *Metall. Mater. Trans. A* **17:1** (1986), 115–120.
- [Morgan 2004] N. B. Morgan, “Medical shape memory alloy applications-the market and its products”, *Mater. Sci. Eng. A* **378:1-2** (2004), 16–23.
- [Moumni et al. 2005] Z. Moumni, A. V. Herpen, and P. Riberty, “Fatigue analysis of shape memory alloys: energy approach”, *Smart Mater. Struct.* **14:5** (2005), S287–S292.
- [Moumni et al. 2008] Z. Moumni, W. Zaki, and Q. S. Nguyen, “Theoretical and numerical modeling of solid-solid phase change: application to the description of the thermomechanical behavior of shape memory alloys”, *Int. J. Plast.* **24:4** (2008), 614–645.
- [Patoor and Berveiller 1993] E. Patoor and M. Berveiller, “Lois de comportement et calcul de structures en alliage à mémoire de forme”, Chapter 9, pp. 195–224 in *Technologie des alliages à mémoire de forme*, Hermes, Paris, 1993.
- [Pineau and Pétrequin 1997] A. Pineau and P. Pétrequin, *La fatigue des matériaux et des structures*, 2^e ed., Chapter 4, pp. 155–214, Hermes, Paris, 1997.
- [Porter et al. 2001] G. A. Porter, P. K. Liaw, T. N. Tieg, and K. H. Wu, “Fatigue and fracture behavior of nickel-titanium shape-memory alloy reinforced aluminum composites”, *Mater. Sci. Eng. A* **314:1-2** (2001), 186–193.
- [Predki et al. 2006] W. Predki, M. Klönne, and A. Knopik, “Cyclic torsional loading of pseudoelastic NiTi shape memory alloys: damping and fatigue failure”, *Mater. Sci. Eng. A* **417:1-2** (2006), 182–189.
- [Raniecki et al. 1992] B. Raniecki, C. LExcellent, and K. Tanaka, “Thermodynamic models of pseudoelastic behaviour of shape memory alloys”, *Arch. Mech.* **44** (1992), 261–284.
- [Shaw and Kyriakides 1995] J. A. Shaw and S. Kyriakides, “Thermomechanical aspects of NiTi”, *J. Mech. Phys. Solids* **43:8** (1995), 1243–1281.
- [Siredey et al. 2005] N. Siredey, A. Hautcoeur, and A. Eberhardt, “Lifetime of superelastic Cu–Al–Be single crystal wires under bending fatigue”, *Mater. Sci. Eng. A* **396:1-2** (2005), 296–301.
- [Tanaka et al. 1995] K. Tanaka, F. Nishimura, T. Hayashi, H. Tobushi, and C. LExcellent, “Phenomenological analysis on subloops and cyclic behavior in shape memory alloys under mechanical and/or thermal loads”, *Mech. Mater.* **19:4** (1995), 281–292.
- [Tolomeo et al. 2000] D. Tolomeo, S. Davidson, and M. Santinoranont, “Cyclic properties of superelastic Nitinol: design implications”, pp. 471–476 in *SMST-2000: Proceedings of the International Conference on Shape Memory and Superelastic Technologies* (Pacific Grove, CA), edited by S. M. Russell and A. R. Pelton, International Organization on SMST, 2000.
- [Vaidyanathan et al. 2000] R. Vaidyanathan, D. C. Dunand, and U. Ramamurty, “Fatigue crack-growth in shape-memory Ni-Ti and NiTi-TiC composites”, *Mater. Sci. Eng. A* **289:1-2** (2000), 208–216.
- [Van Humbeeck 1991] J. Van Humbeeck, “Cycling effects, fatigue and degradation of shape memory alloys”, *J. Phys. (France) IV* **1:C4** (1991), 189–197.
- [Wagner et al. 2004] M. Wagner, T. Sawaguchi, G. Kausträter, D. Höffken, and G. Eggeler, “Structural fatigue of pseudoelastic NiTi shape memory wires”, *Mater. Sci. Eng. A* **378:1-2** (2004), 105–109.
- [Wayman and Otsuka 1999] C. M. Wayman and K. Otsuka (editors), *Shape memory alloys*, Cambridge University Press, Cambridge, 1999.

[Zaki 2006] W. Zaki, *Comportement thermo-mécanique des matériaux à mémoire e forme*, Ph.D. thesis, École Polytechnique, 2006, Available at <http://www.imprimerie.polytechnique.fr/Theses/Files/Zaki.pdf>.

[Zaki and Moumni 2007a] W. Zaki and Z. Moumni, “A 3d model of the cyclic thermomechanical behavior of shape memory alloys”, *J. Mech. Phys. Solids* **55**:11 (2007), 2427–2454.

[Zaki and Moumni 2007b] W. Zaki and Z. Moumni, “A three-dimensional model of the thermomechanical behavior of shape memory alloys”, *J. Mech. Phys. Solids* **55**:11 (2007), 2455–2490.

Received 14 Dec 2007. Revised 14 Apr 2008. Accepted 6 Apr 2008.

ZIAD MOUMNI: moumni@ensta.fr

UME-MS, École Nationale Supérieure de Techniques Avancées, 91761 Palaiseau Cedex, France

Wael Zaki: wael.zaki@tudor.lu

Laboratory for Industrial Technology, Henri Tudor Public Research Center, 66 rue de Luxembourg, L-4221 Esch-sur-Alzette, Luxembourg

HABIBOU MAITOURNAM: habibou@lms.polytechnique.fr

Laboratoire de Mécanique des Solides, UMR 7649, École Polytechnique, 91128 Palaiseau Cedex, France

THERMOMECHANICAL MODELLING OF FRICTION EFFECTS IN GRANULAR FLOWS USING THE DISCRETE ELEMENT METHOD

VIET DUNG NGUYEN, JÉRÔME FORTIN, MOHAMED GUESSASMA,
EMMANUEL BELLENGER AND CLAUDIA COGNÉ

This study deals with the modelling of the thermomechanical phenomena due to friction effects during granular flow. A two-dimensional model using the discrete element method (DEM) and taking into account the contact detection and heat transfers between grains has been developed. Through this study, we have modelled the heat transfer by conductance and the energy dissipation by friction into a granular medium. This modelling enables better understanding of the phenomena at the contact point between grains as well as the energy dissipation by friction of a great number of grains in motion. The validity of the proposed model has been studied by considering some numerical simulations in quasistatic and dynamic regimes.

1. Introduction

The published literature analysis has shown the importance of thermal energy in granular media for industrial processes in applications as diversified as powder metallurgy, chemical reactors (catalysts beds), food technology [Laguerre et al. 2006], thermal insulation [Melka and Bézian 1997], or even simply storing particles in a silo after drying [Ketterhagen et al. 2007]. Only few studies are interested in the understanding of heat transfers resulting from thermomechanical effects. However, these complex phenomena with multiphysical characteristics are an essential stake in the world of industry and transports. For instance, strong frictions (braking, jamming) are responsible for half of ignitions of explosive atmospheres, and are also a dreadful cause of fires in vehicles and accidents. One of the difficulties lies in predicting the friction forces and the temperatures in the friction zone based on the intrinsic properties of bodies in contact.

From a thermal energy point of view, sliding contact is the source of a heat generation by friction, whose distribution between the different bodies is difficult to estimate. Besides, the determination of the contact area, which plays an important role in the value of the transferred heat flow, is also difficult to estimate and depends on various parameters like porosity, rugosity, the distribution of contact forces, and the structure of the media. Mechanical engineers are at the origin of the greatest number of works [Slavin et al. 2002; Vargas-Escobar and McCarthy 2002a; Bahrami et al. 2006]. Slavin et al. [2002] and Bahrami et al. [2006] have developed models to estimate the effective thermal conductivity of a particle packing from the intrinsic properties of solids and fluids. These models enable us to determine the apparent thermal conductivity evolution of a granular medium according to the mechanical loading applied to the particle bed. Vargas-Escobar and McCarthy [2002a] have studied more particularly the influence of contact forces on the apparent conductivity of a bed with a small, but finite, area of contact.

Keywords: heat transfer, contact, conductance, friction, DEM.

The first part of this work consists of using the discrete element method (DEM) for contact detection, determination of contact forces, and kinematic parameters. A computational program, MULTICOR, that can treat an important number of particles (10^6), has been developed to solve the mechanical equations. In the second part, heat transfers by contact as well as energy generation by friction have been studied and implemented in MULTICOR. Finally, through some examples, we underscore the phenomena of thermomechanical interaction.

2. Mechanical resolution by DEM

The conventional DEM allows us to model really deformable particles as well as complex shapes, from the ellipsoid to the polygon. Here, we have studied the simple case of nondeformable and nonpenetrable particles in two dimensions with the computational programme MULTICOR [Fortin et al. 2005]. The coordinates and the Euler rotation angles are the configuration parameters q . The gyroscopic forces are three-dimensional but the centrifugal forces exist even in two-dimensional problems (see Figure 1). The matrix of generalized mass M of the system doesn't depend on q , and hence is a diagonal block. The mechanical equation can be written in the form

$$M\ddot{q} = F_{\text{ext}}(q, \dot{q}, t) + R^\alpha,$$

where F_{ext} represents the known external forces and R^α the unknown interior forces related to contact reactions, with α being the contact number.

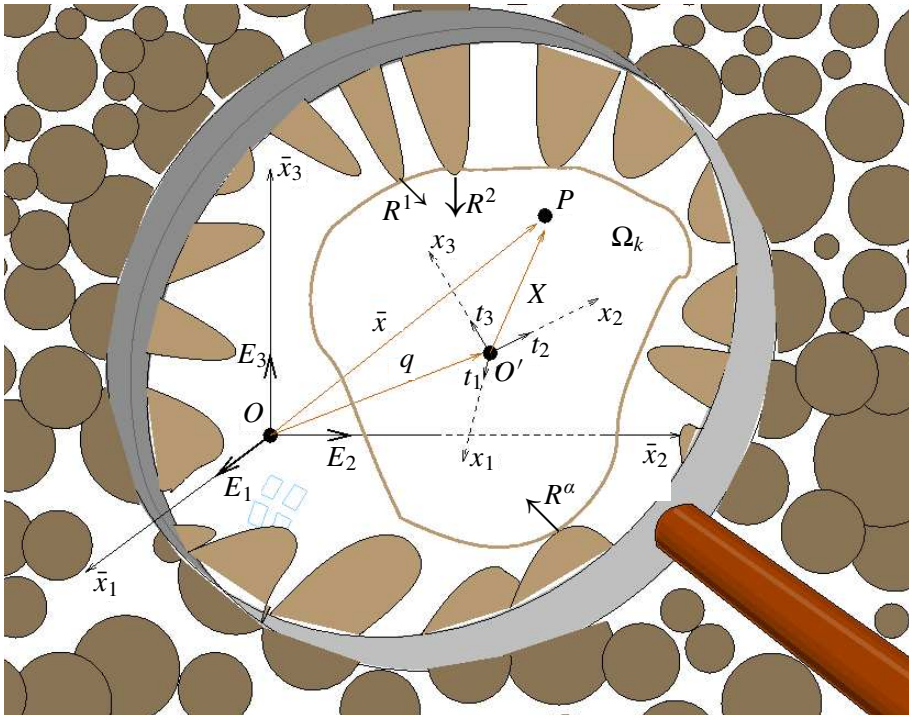


Figure 1. Granular media and its kinematic parameters.

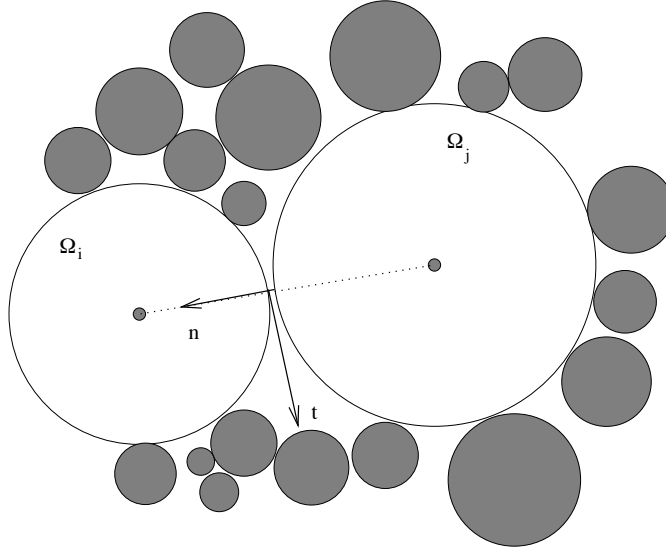


Figure 2. Detection of contact.

In a system composed of p heterogeneous particles (see Figure 1), the critical parameter for the modelling time is the maximum number of interactions between particles. The more the interaction range is important the more we have to test the possible interactions between particles. MULTICOR uses the partitioning method coupled to a connectivity table [Fortin and Coorevits 2004]. This technique allows us to reduce the computational time considerably. In this case, the computational time no longer increases as $O(p^2)$ but only as $O(p)$, which is almost optimal. To each pair of particles Ω_i and Ω_j which may enter in contact, we associated a local reference whose axes are oriented according to the two unit vectors n and t , respectively the normal and tangential vectors in the contact plan (see Figure 2). The normal n is directed from Ω_j to Ω_i . The variables put in duality are \dot{u}^{ij} , the relative local velocity of Ω_i with respect to Ω_j , and the contact reaction r^{ij} of Ω_j on Ω_i . In the local base, they are written as

$$\dot{u}^{ij} = \dot{u}_t^{ij}.t + \dot{u}_n^{ij}.n, \quad r^{ij} = r_t^{ij}.t + r_n^{ij}.n,$$

where \dot{u}_n^{ij} is the normal separation velocity, \dot{u}_t^{ij} the sliding velocity, r_n^{ij} the contact pressure, and r_t^{ij} the friction force.

The introduction of Coulomb's friction μ leads to a nonlinear problem which cannot be solved by a linear programming method. Unlike the usual approach, the bipotential method leads to a single variational principle and an inequality [Fortin and de Saxcé 1999]. Using Usawa's algorithm, we obtain a resolution algorithm of the constitutive law based on the predictive-corrective scheme expressed by

$$\text{predictor} : \tau^{ij} = r^{ij} - \gamma [\dot{u}_t^{ij} + (\dot{u}_n^{ij} + \mu || - \dot{u}_t^{ij} ||).n], \quad \text{corrector} : r^{ij} = \text{proj}(\tau^{ij}, K_\mu),$$

where γ is a numerical parameter, and τ^{ij} the projection of Coulomb's cone K_μ leads, according to the value of τ^{ij} , to one of the following states: noncontact, contact with friction, or sliding contact. Conventionally, at each time step, the contact forces in the system are determined repeatedly by the method of successive balances based on a Gauss–Seidel algorithm for the two-dimensional version. Each

contact force is calculated by adopting temporary values over the other contacts. The convergence is obtained when the force satisfies the unilateral contact law with dry friction.

The calculation cycle is a time-stepped algorithm which requires the repetition of the following resolution scheme.

$$\left[\begin{array}{l} t = t + \Delta t \\ \text{Evaluation of the particle positions } q_n \\ \text{Detection of the contact number } \alpha \text{ in the system} \\ \text{Evaluation of the particle velocity (without contact)} \\ \left[\begin{array}{l} i = i + 1 \quad (\text{iteration of solver bipotential}) \\ \left[\begin{array}{l} \alpha = \alpha + 1 \quad (\text{contact loop : } \alpha \text{ is the current index of contact}) \\ \text{Evaluation of the contact reactions } r^{\alpha, i+1} \end{array} \right. \\ \text{Indicator of error} \end{array} \right. \\ \text{Evaluation of the velocity} \end{array} \right.$$

3. Heat transfer in granular media and thermomechanical formulation

In general, heat transfer in granular media with a stagnant interstitial fluid is assumed to occur by the following physical phenomena:

- Heat conduction through the particles and heat conduction through the fluid between the neighboring particles. Furthermore, in a multicontact system, as considered in this work, we must consider heat conduction through the contact area between two particles Ω_i and Ω_j . Contact conductance refers to the ability to transmit heat across their mutual interface.
- Radiant heat transfer between the fluid within neighboring voids and radiant heat transfer between the surfaces of neighboring particles. For heat transfer by radiation, contact between surfaces is not required. Radiation is linked to the production of electromagnetic waves by a heat surface.
- For fluids, flow heat transfer by interparticle convection can be considered if there is a difference of temperature between the particles and the fluid.

In frictional granular flow regimes, heat transfer occurs from the phenomena presented above. In addition, it is necessary to consider the heat generated by friction between two particles Ω_i and Ω_j . Indeed, sliding contact is an important source of heat generation for the dynamic granular problems considered in this work (see [Figure 3](#), left).

In this paper, we assume that conduction through the solid phase dominates the heat conduction. This assumption is verified when

$$\frac{\lambda_s s}{\lambda_f a} \gg 1,$$

where λ_s and λ_f are respectively the conductivities of the particles and the fluid, a and s are the radius of the particle and the contact area, respectively.

This expression is satisfied for high thermal conductivity solid materials or for solid particles in a vacuum ($\lambda_f \rightarrow 0$) [[Vargas-Escobar and McCarthy 2002b](#)]. Also, under these conditions, the heat transfer between two adjacent particles Ω_i and Ω_j is only controlled by the contact conductance. In this work, radiant and convective heat transfers are neglected. Therefore, we only consider heat transfer in granular flow by contact conductance and frictional effects (see [Figure 3](#), right).

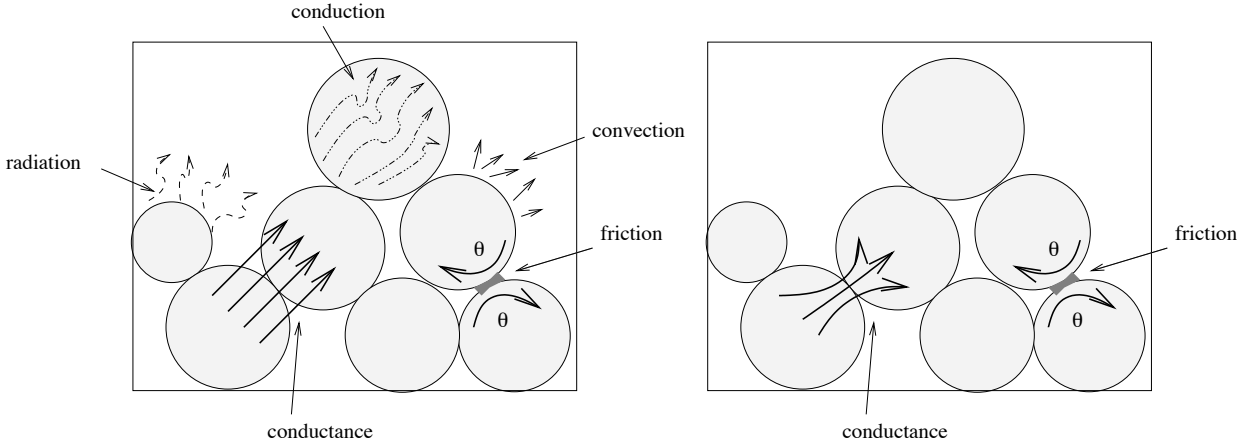


Figure 3. Left: heat transfer mechanisms in granular media. Right: heat transfer in MULTICOR.

3.1. Heat transfer by conductance. Contact conductance is directly linked to the constriction of the heat flow lines in the contact point [Laguerre et al. 2006]. The thermal contact conductance is defined by the ratio of the heat flow across a contact interface and the magnitude of the temperature discontinuity at the interface,

$$\varphi_{ij} = H_c^{ij}(T_j - T_i),$$

where φ_{ij} is the heat flow transferred between the particles Ω_i and Ω_j , $T_j - T_i$ the temperature difference between the mid-planes of the spheres and H_c^{ij} the contact conductance between the particles Ω_i and Ω_j , with j varying from 1 to the contact number α .

The coefficient H_c^{ij} , which is a function of the compression force, refers to the ability of two materials in contact to transfer heat across their mutual interface (see Figure 4). In our work, contact conductance between particles Ω_i and Ω_j is modeled using Hertz theory as

$$H_c^{ij} = 2\lambda_s \left(\frac{3r_n^{ij} a^*}{4E^*} \right)^{1/3},$$

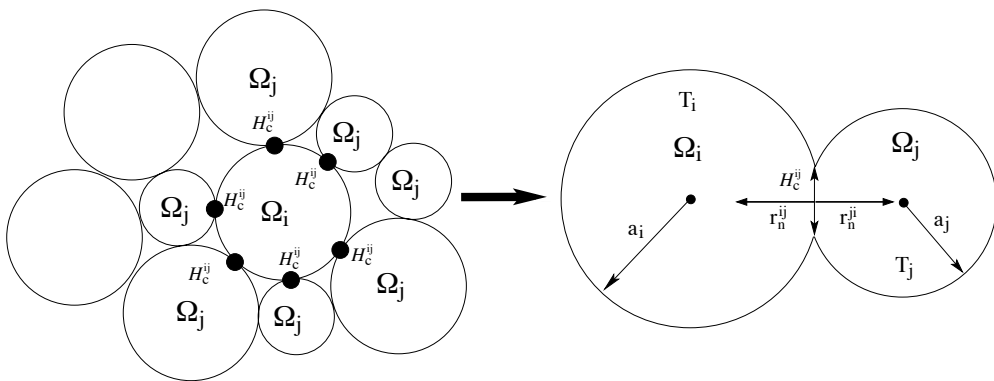


Figure 4. Schematic representation of heat transfer by conductance.

where r_n^{ij} is the normal force, a^* is the equivalent radius, given by $\frac{1}{a^*} = \frac{1}{a_i} + \frac{1}{a_j}$, and E^* is the effective Young's modulus, so that

$$\frac{1}{E^*} = \frac{1 - \nu_i^2}{E_i} + \frac{1 - \nu_j^2}{E_j},$$

ν being Poisson's ratio.

The contact between two adjacent particles is assumed to be smooth and sliding. The contact conductance is calculated dynamically at each time step and for all contacts of a particle Ω_i .

We recall that the considered particles are nondeformable and nonpenetrable. The use of Hertz's theory only enables us to compute the contact conductance coefficient H_c^{ij} . We assume that the particles remain rigid all the time.

3.2. Heat generated by friction. In this case, heat flow is generated by dissipation of energy during friction between particles. The deformation is neglected because the particles are assumed rigid. Friction is understood as a continuous mechanical solicitation between two bodies. The heat energy generated by friction, E_{fij} , at the frictional interface during a time step Δt is

$$E_{fij} = \mu \dot{u}_t^{ij} r_n^{ij} \Delta t,$$

where μ is the friction coefficient, \dot{u}_t^{ij} the sliding velocity, and r_n^{ij} the normal force.

The modelling of the heat generated by friction requires us to share it between particles in sliding contact. Therefore, we define the partition coefficient of generated heat flow β_{ij} . This coefficient depends on different microscopic parameters like the thermal properties, the sliding velocity, heat generation parameters, and the surface roughness if the contact is not perfect [Linck et al. 2006]. Research in this area has proposed different equations to estimate this coefficient. In our study, this coefficient is obtained from the analytical solution of Mokrani and Bourouga [2005],

$$\beta_{ij} = \frac{1}{2} \left(\frac{\rho_i}{\rho_i + \rho_j} + \frac{\lambda_s^i}{\lambda_s^i + \lambda_s^j} \right),$$

where ρ is the electric resistivity ($\Omega \text{ m}$).

We assume that the packed bed is made of only one material. The partition coefficient of generated heat flow β_{ij} is then equal to $\frac{1}{2}$.

3.3. Thermomechanical resolution. Taking into account the various phenomena of heat generation mentioned above, the energy balance and the variation of temperature for a particle during a small time step Δt can be written as

$$m_i C_{Pi} \frac{\Delta T_i}{\Delta t} = \sum_{j=1}^{\alpha} \left(H_c^{ij} (T_j - T_i) + \frac{1}{2} \frac{E_{fij}}{\Delta t} \right), \quad (1)$$

where m_i and C_{Pi} are the mass and the heat capacity for Ω_i respectively, and α the contact number.

The temperature evolution between two bodies in contact is governed by Equation (1), representing the balance of the heat energy. For static problems this equation is solved with a time step $\Delta t = 10^{-3}$ s. For dynamical problems (1) is solved with a time step $\Delta t = 10^{-6}$ s, to assume that the temperature of each particle changes slowly so that thermal perturbations do not propagate further than its immediate

neighbors during one time step. The second requirement is that the heat transfer resistance Ω_i (conduction) through a particle is significantly lower than the contact resistance between two particles, Ω_i and Ω_j , provided that

$$Bi = \frac{H_c^{ij}}{\lambda_s a} \ll 1,$$

where Bi is the Biot number.

Equation (1) is discretized in order to compute the temperature T_i at the time step $t + \Delta t$ as

$$T_i^{t+\Delta t} = T_i^t + \frac{\Delta t}{m_i C_{Pi}} \sum_{j=1}^{\alpha} \left(H_c^{ij} (T_j^t - T_i^t) + \frac{1}{2} \frac{E_{fij}}{\Delta t} \right).$$

The general algorithm implemented in MULTICOR is the following:

$$\left[\begin{array}{l} t = t + \Delta t \\ \text{Evaluation of the particle positions } q_n \\ \text{Detection of the contact number } \alpha \text{ in the system} \\ \text{Evaluation of the particle velocity (without contact)} \\ \left[\begin{array}{l} i = i + 1 \quad (\text{iterations of solver bipotential}) \\ \left[\begin{array}{l} \alpha = \alpha + 1 \quad (\text{contact loop}) \\ \text{Evaluation the contact reaction } r^{\alpha, i+1} \end{array} \right] \\ \text{Indicator of error} \end{array} \right] \\ \text{Evaluation of } T_i^t, H_c^{ij}, E_{fij} \\ \text{Evaluation of the temperature } T_i^{t+\Delta t} \\ \text{Evaluation of the velocity} \end{array} \right.$$

The resolution of the heat problem requires us to compute at each time step the contact detection, the determination of forces, and the velocities of particles.

4. Numerical simulations

4.1. Comparison between DEM and FEM. The first application was carried out on a particulate system obtained by a triangular arrangement of 14 particles, with an identical size and a circular shape (1 mm radius). We initially suppose that particle 1 is heated at 100° C and the remaining particles are at a temperature of 20° C (see Figure 5).

From the thermal point of view, this simple modelling allows us to study the diffusion of the heat flow by conductance in granular media. In order to check that this assumption is not too restrictive, we compared MULTICOR's prediction with the results obtained by CAST3M software based on the finite element method (FEM) (see Figure 5). Through this comparison, we could check that the thermal resistance within particles was negligible compared to the thermal contact resistance between particles. This hypothesis could be checked by studying the temperature evolution in particles 1, 5 and 14. The good agreement between MULTICOR and CAST3M seems to validate the assumption about the heat transfer occurring only by conductance.

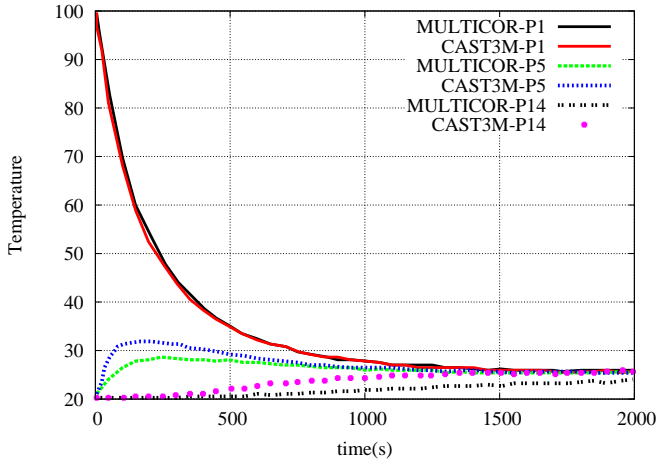


Figure 5. Temperature evolution in particles 1, 5, and 14 obtained by MULTICOR and CAST3M.

4.2. Comparison between experiment and MULTICOR prediction. In order to validate our thermomechanical model implemented in MULTICOR, we have compared the results with experimental data from [Vargas-Escobar and McCarthy 2002b]. In this study, the authors developed an experimental setup to investigate the heat transfer in a quasistatic configuration. The system is composed of dispersed stainless steel spheres forming a two-dimensional packed bed ($30.4 \times 45.7 \text{ cm}^2$). The bottom wall is kept at $T_{\text{wall}} = 50^\circ \text{C}$. The top, left, and right walls are insulated (see Figure 6, left). The initial temperature is $T_0 = 25^\circ \text{C}$.

The DEM model shown in the figure was computed by using the thermomechanical properties of stainless steel based on experiments (see Table 1).

In Figure 6, right, we present a part of thermal map after 30 minutes heating. The temperature in the heated granular bed does not propagate uniformly. The front oscillates as force chains appear and disappear along the bed’s height.

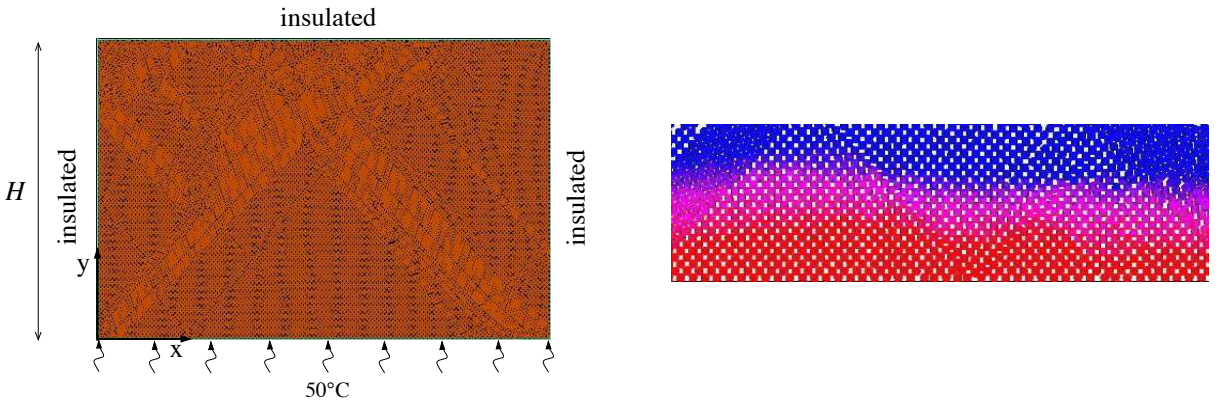


Figure 6. Left: the MULTICOR model. Right: part of the thermal map after 30 minutes heating.

| | | | | |
|-------------------------|-----------------|-----------------|----------------------|-------------|
| Density | Poisson's ratio | Young's modulus | Particle radius | λ_s |
| 7,500 kg/m ³ | 0.29 | 193 GPa | 0.003 m | 15 W/mK |
| Number of particles | Length | Height | Friction coefficient | |
| 15,548 | 0.45 m | 0.31 m | 0.29 | |

Table 1. Parameters used in the simulation.

Figure 7 presents a comparison of the temperature as a function of bed's height given by our predictions and the experimental results obtained by Vargas et al. [2002b] after 30 minutes heating. In this figure, θ represents the dimensionless temperature and η the dimensionless height, given by the equations

$$\theta = \frac{T - T_0}{T_{\text{wall}} - T_0}, \quad \eta = \frac{y}{H}.$$

As shown in Figure 7, the predicted temperature is less accurate for the particles at the bottom of the bed. However, it can be seen that the simulation results matches the experimental curves for dimensionless heights higher than 0.03, which allows us to validate the model prediction in the quasistatic case.

4.3. Simulation of heat generation by friction in granular media. In this part, we begin to investigate the problem of heat generation in the granular material subjected to shearing solicitation at an imposed velocity. Two cases of shearing will be studied: the quasistatic and dynamic regimes.

The granular material consists of p particles having diameters d (1, 2, and 3 mm). The intergranular contact is characterized by the friction coefficient $\mu = 0.5$.

In the geometry of the shearing plane (see Figure 8), the material is compacted by gravity between two rough parallel walls at a distance H . The top wall is set in motion with velocity V in the positive x direction, and the system is compressed by imposing a uniaxial load F . Periodic conditions are imposed at the lateral boundaries. This means that a particle leaving through one of the boundaries reenters the

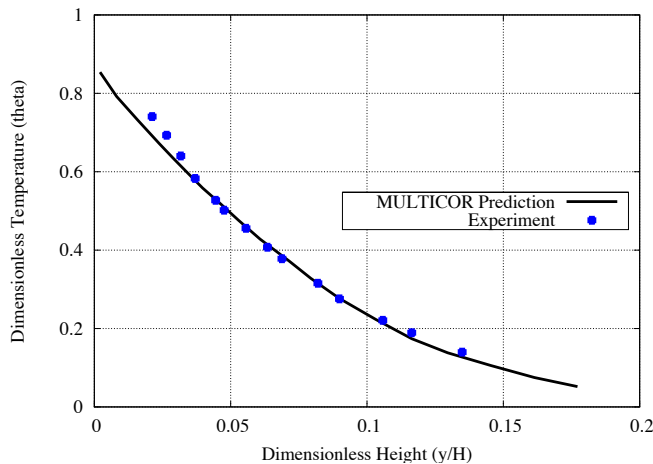


Figure 7. Comparison of experimental data with MULTICOR results after 30 minutes heating.

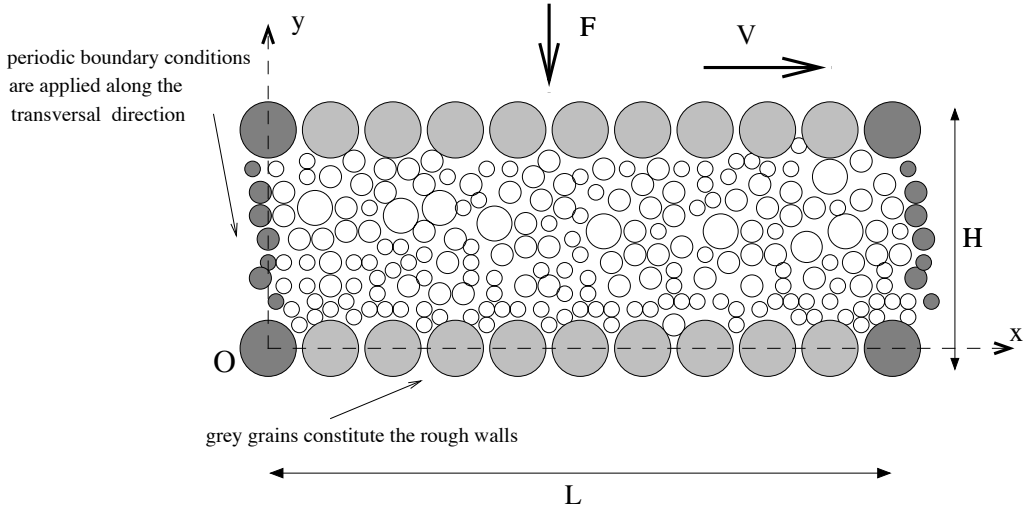


Figure 8. Model of shearing plan with periodic boundaries.

cell at the same vertical location and with the same kinematic and dynamic conditions. L is the length of the simulated flow (equal to 40 particles). This size appears sufficient to neglect the length effects of the simulation box. It is considered here that the roughness of wall surface is modeled by jointed grains having the same characteristics as grains in the flow.

Conventionally, we observe two regimes in granular flows. First we have the quasistatic regime, which usually obeys the constitutive laws in which the effective friction coefficient is constant. Secondly, there is the rapid-flow or dynamic regime, where the particles interact collisionally. In this case, the solid fraction is close to the maximum.

In [Figure 9](#) we present the velocity field of two regimes. On the top two images we note the appearance of small vortices which characterize the quasistatic regime. These disappear in the dynamic regime, observed in the bottom images. In this case, the granular flow becomes stationary. This phenomenon gives us a better understanding of the behavior of granular flow [[da Cruz 2004](#)].

[Figure 10](#) shows the contact force distribution across the granular material at the beginning of the simulation and when the stationary regime is established.

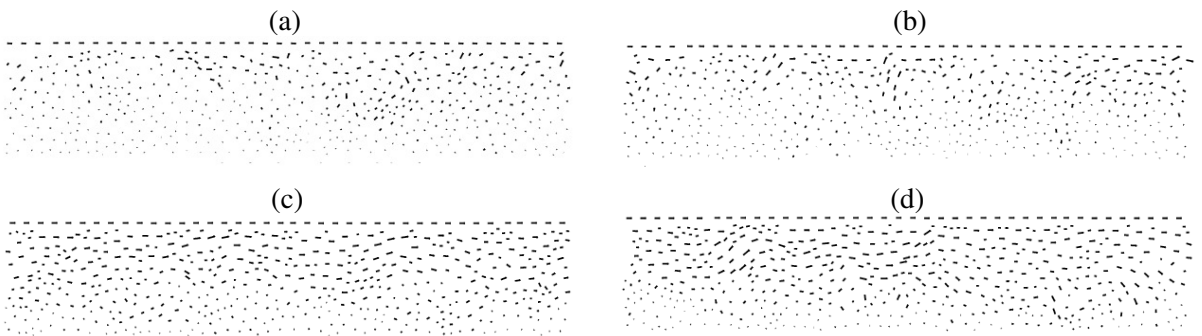


Figure 9. The velocity field in the quasistatic regime with a small vortex (top row), and in the dynamic regime (bottom).

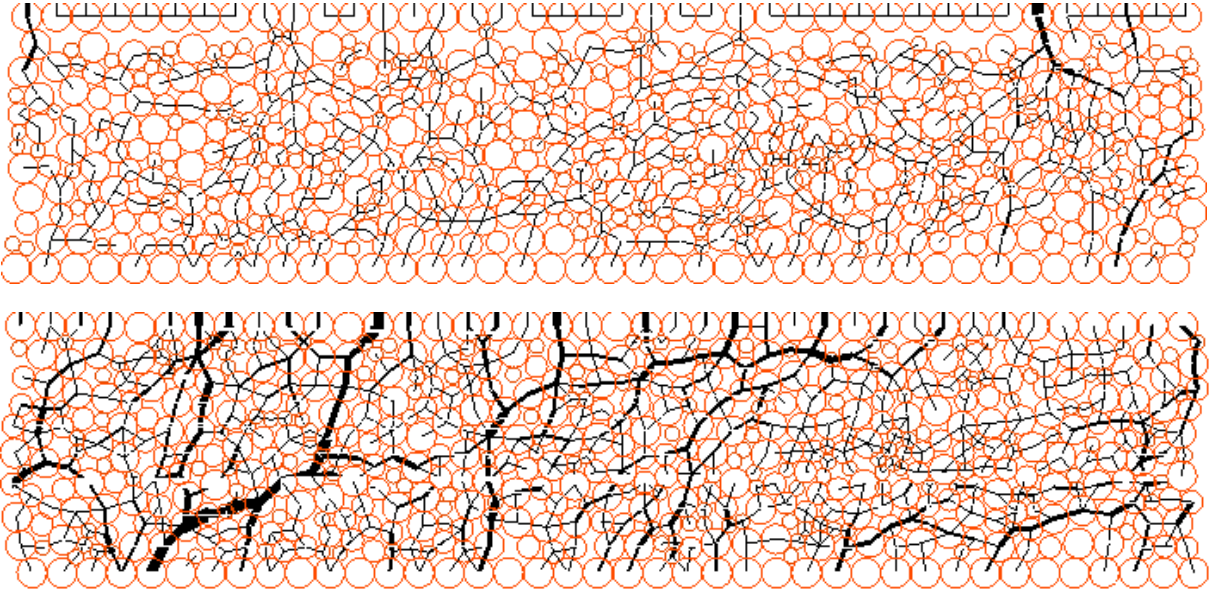


Figure 10. Contact forces field at the beginning of the simulation (top) and when stationary flow is established (bottom).

Another result is concerned with the variation of velocity as a function of the height. We see in Figure 11 that the average velocity agrees well with the curves proposed by da Cruz [2004], [?], and Vargas-Escobar [2002]. There is no turbulence close to the mobile wall.

The major result of this work is the heat generated by friction between particles and the shearing wall (see Figure 12). Figure 13, left, shows how heat varies with the granular layer depth. The hottest zone is located in the middle of the flow. These observations are confirmed by Vargas-Escobar [2002], who has shown that the granular temperature profiles illustrate the high-shear zones in the center of the cell.

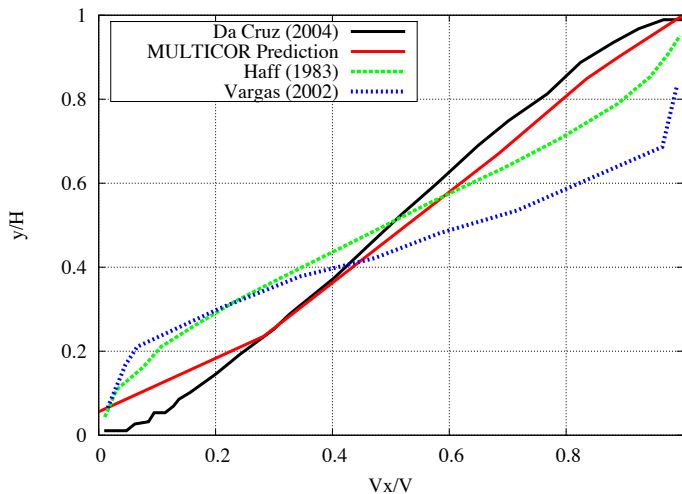


Figure 11. Typical velocity normalized profile.

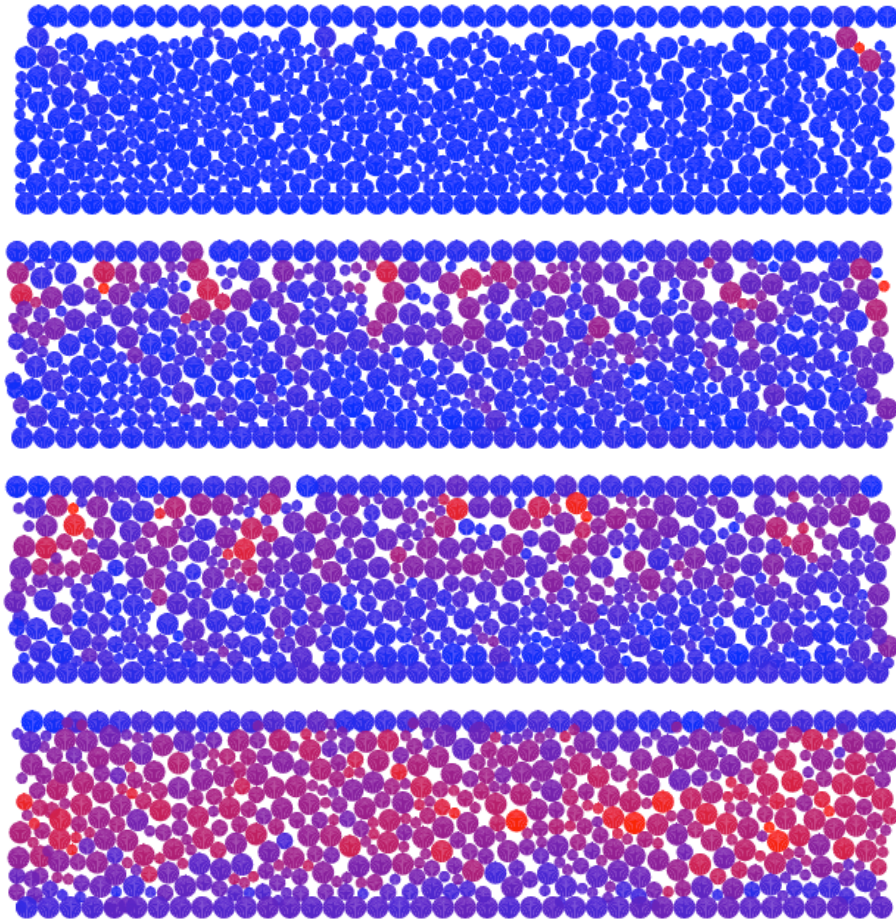


Figure 12. Thermal maps of heat evolution in the quasistatic regime.

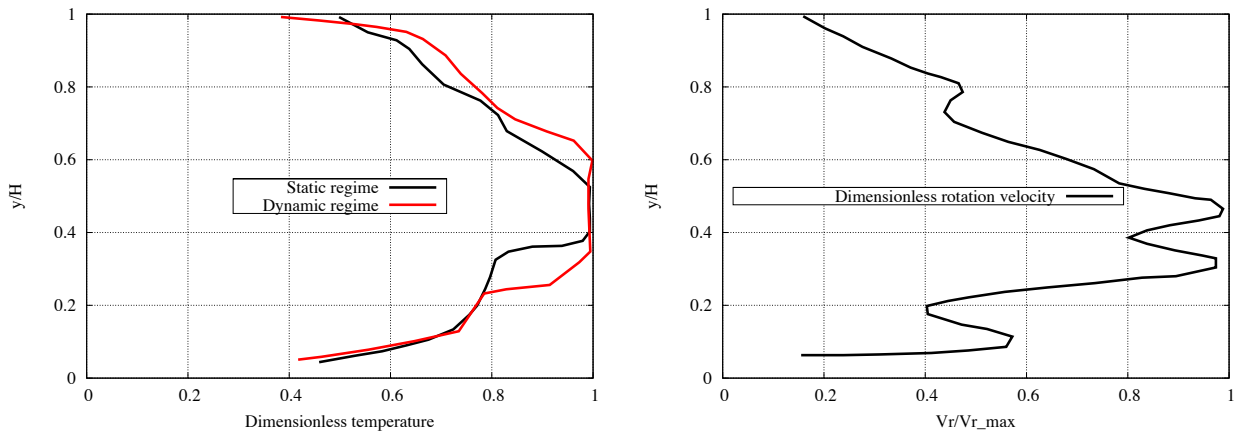


Figure 13. Left: heat profiles. Right: normalized angular velocity.

Here we note that thermal symmetry of the heat profile means that the bottom wall plays an important role in the heat generation. We also note that these profiles correspond with the angular velocity curves, seen on the right in [Figure 13](#).

5. Conclusions

The present work focuses on the modelling of heat transfer and heat generation by friction in a granular material by using DEM. The proposed model has been implemented in MULTICOR software and performed for some examples to check its validity. The numerical predictions obtained with MULTICOR agree with the experimental results and numerical predictions from the literature. Therefore, the proposed assumptions about the predominance of heat transfer by conductance in comparison with other thermal effects has negligible influence on the numerical predictions.

The next step will consist in the incorporation of the other heat transfer phenomena like convection, radiation, and impact effect. Moreover, further studies will focus on the dynamic friction coefficient and its changes with the temperature and the wear in the contact area. An experimental campaign is also planned with our industrial partner [Proust et al. \[2007\]](#).

Acknowledgement

We thank Mrs. Boutinaud-Morel for her advice in the translation of this article.

References

- [Bahrami et al. 2006] M. Bahrami, M. M. Yovanovich, and J. R. Culham, “Effective thermal conductivity of rough spherical packed beds”, *Int. J. Heat Mass Tran.* **49**:19-20 (2006), 3691–3701.
- [da Cruz 2004] F. da Cruz, *Écoulement de grains secs: frottement et blocage*, Ph.D. thesis, École Nationale des Ponts et Chaussées, Paris, 2004.
- [Fortin and Coorevits 2004] J. J. Fortin and P. Coorevits, “Selecting contact particles in dynamics granular mechanics systems”, *J. Comput. Appl. Math.* **168**:1-2 (2004), 207–213.
- [Fortin and de Saxcé 1999] J. Fortin and G. de Saxcé, “Modélisation numérique des milieux granulaires par l’approche du bi-potentiel”, *C. R. Acad. Sci. IIb* **327** (1999), 721–724.
- [Fortin et al. 2005] J. Fortin, O. Millet, and G. de Saxcé, “Numerical simulation of granular materials by an improved discrete element method”, *Int. J. Numer. Meth. Eng.* **62**:5 (2005), 639–663.
- [Haff 1983] P. K. Haff, “Grain flow as a fluid-mechanical phenomenon”, *J. Fluid Mech.* **134** (1983), 401–430.
- [Ketterhagen et al. 2007] W. R. Ketterhagen, J. S. Curtis, C. R. Wassgren, A. Kong, P. J. Narayan, and B. C. Hancock, “Granular segregation in discharging cylindrical hoppers: a discrete element and experimental study”, *Chem. Eng. Sci.* **62**:22 (2007), 6423–6439.
- [Laguerre et al. 2006] O. Laguerre, S. B. Amara, and D. Flick, “Heat transfer between wall and packed bed crossed by low velocity airflow”, *Appl. Therm. Eng.* **26**:16 (2006), 1951–1960.
- [Linck et al. 2006] V. Linck, A. Saulot, and L. Baillet, “Consequence of contact local kinematics of sliding bodies on the surface temperatures generated”, *Tribol. Int.* **39**:12 (2006), 1664–1673.
- [Melka and Bézian 1997] S. Melka and J. J. Bézian, “L’isolation thermique par les matériaux granulaires”, *Rev. Gen. Therm.* **36**:5 (1997), 345–353.
- [Mokrani and Bourouga 2005] H. Mokrani and B. Bourouga, “Modèle de coefficient de partage du flux généré à une interface de contact électrothermique: approche microscopique en régime permanent”, in *12èmes Journées Internationales de Thermique* (Tanger, Morocco, 2005), 2005.

- [Proust et al. 2007] C. Proust, S. Hawksworth, R. Rogers, M. Beyer, D. Lakic, D. Raveau, P. Hervé, V. Pina, C. Petitfrère, and X. Lefebvre, “Development of a method for predicting the ignition of explosive atmospheres by mechanical friction and impacts (MECHEX)”, *J. Loss Prevent. Proc.* **20**:4-6 (2007), 349–365.
- [Slavin et al. 2002] A. J. Slavin, V. Arcas, C. A. Greenhalgh, E. R. Irvine, and D. B. Marshall, “Theoretical model for the thermal conductivity of a packed bed of solid spheroids in the presence of a static gas, with no adjustable parameters except at low pressure and temperature”, *Int. J. Heat Mass Tran.* **45**:20 (2002), 4151–4161.
- [Vargas-Escobar 2002] W. L. Vargas-Escobar, *Discrete modeling of heat conduction in granular media*, Ph.D. thesis, University of Pittsburgh, Pittsburgh, 2002, Available at <http://etd.library.pitt.edu/ETD/available/etd-02012002-193942/unrestricted/FinalBW.pdf>.
- [Vargas-Escobar and McCarthy 2002a] W. L. Vargas and J. J. McCarthy, “Conductivity of granular media with stagnant interstitial fluids via thermal particle dynamics simulation”, *Int. J. Heat Mass Tran.* **45**:24 (2002), 4847–4856.
- [Vargas-Escobar and McCarthy 2002b] W. L. Vargas-Escobar and J. J. McCarthy, “Stress effects on the conductivity of particulate beds”, *Chem. Eng. Sci.* **57**:15 (2002), 3119–3131.

Received 17 Dec 2007. Accepted 7 Mar 2008.

VIET DUNG NGUYEN: vietdung.nguyen@u-picardie.fr

Laboratoire des Technologies Innovantes (EA 3899), Université de Picardie Jules Verne, IUT de l’Aisne, 02100 Saint Quentin, France

JÉRÔME FORTIN: jerome.fortin@insset.u-picardie.fr

Laboratoire des Technologies Innovantes (EA 3899), Université de Picardie Jules Verne, IUT de l’Aisne, 02100 Saint Quentin, France

MOHAMED GUESSASMA: mohamed.guessasma@u-picardie.fr

Laboratoire des Technologies Innovantes (EA 3899), Université de Picardie Jules Verne, IUT de l’Aisne, 02100 Saint Quentin, France

EMMANUEL BELLENGER: emmanuel.bellenger@u-picardie.fr

Laboratoire des Technologies Innovantes (EA 3899), Université de Picardie Jules Verne, IUT de l’Aisne, 02100 Saint Quentin, France

CLAUDIA COGNÉ: claudia.cogne@u-picardie.fr

Laboratoire des Technologies Innovantes (EA 3899), Université de Picardie Jules Verne, IUT de l’Aisne, 02100 Saint Quentin, France

SEISMIC BEARING CAPACITY OF CIRCULAR FOOTINGS: A YIELD DESIGN APPROACH

JEAN SALENÇON, CHARISIS THEODOROU CHATZIGOGOS AND ALAIN PECKER

As developed during the past decades, the yield design theory provides an approach to the stability analysis of civil engineering structures under seismic conditions which has been often used, explicitly or implicitly. New results related to circular footings resting on a purely cohesive soil, taking into account the horizontal inertia forces, are presented in this paper for practical applications to the safety coefficient to be applied to the vertical load when designing seismic foundations.

An outline of yield design theory

Just to clarify the terminology and to fix the notations, a brief outline of the yield design theory [Salençon 1983; 1990] is recalled here within the three dimensional continuum mechanics framework. It aims at estimating the *extreme loads* that can be supported by a structure from the knowledge of its geometry, of the loading process it undergoes and of the strength criteria of its constituent materials, whatever the physical phenomena they are related to. Since they do not refer to any data about the constitutive law of its materials before and at failure, the results obtained are but upper bound estimates for the *actual* ultimate loads and no information can be obtained regarding the displacements. With the generic notations Ω and $\partial\Omega$ for the volume and the boundary of the system, the quasistatic loading mode of the system is described through a multiparameter loading vector \mathbf{Q} with components Q_i and the associated dual kinematic parameters \hat{q}_i defining the virtual kinematic vector $\hat{\mathbf{q}}$. The principle of virtual rates of work [Salençon 2001] thus takes the form: *for all σ statically admissible with \mathbf{Q} , and all $\hat{\mathbf{U}}$ kinematically admissible with $\hat{\mathbf{q}}$,*

$$\int_{\Omega} \sigma : \hat{\mathbf{d}} \, d\Omega + \int_{\Sigma_{\hat{\mathbf{U}}}} \mathbf{n} \cdot \sigma \cdot \llbracket \hat{\mathbf{U}} \rrbracket \, d\Sigma_{\hat{\mathbf{U}}} = \mathbf{Q} \cdot \hat{\mathbf{q}}, \quad (1)$$

with

$$\sigma \mapsto \mathbf{Q} \text{ and } \hat{\mathbf{U}} \mapsto \hat{\mathbf{q}} \text{ linear}, \quad (2)$$

where the symbol “:” denotes the double contracted product and “·” the dot product. σ stands for the Cauchy stress tensor field, and $\hat{\mathbf{d}}$ for the virtual strain rate tensor field derived from the virtual velocity field $\hat{\mathbf{U}}$. Since such a field may exhibit velocity jumps $\llbracket \hat{\mathbf{U}} \rrbracket$ across velocity jump surfaces $\Sigma_{\hat{\mathbf{U}}}$ which are part of its definition, the second term in Equation (1) accounts for the corresponding contribution.

The same description is adopted in the case of a dynamic loading treated as a quasistatic phenomenon by incorporating the corresponding given inertia forces within the applied external forces.

Homogeneity is not assumed and the term *constituent material* will be used generically from now on to describe all the materials constituting the system, including the interfaces between different elements.

Keywords: seismic bearing capacity, circular footings, yield design, external approach, interaction diagrams, Tresca criterion.

The resistance of the constituent material is defined at any point of the system through a convex strength criterion to be satisfied by the stress state.

When trying to determine the loads \mathbf{Q} that can be supported by the system under the specified strength conditions, it is clear that mathematical compatibility at any point of the system between the equilibrium equations and the material resistance conditions is necessary for a load to be supported. Such loads generate the convex domain of the *potentially safe* loads in the \mathbf{Q} vector space. The boundary of K defines the *extreme loads* of the system. Any stress field in equilibrium with a load \mathbf{Q} that complies with the strength criterion is sufficient to prove that $\mathbf{Q} \in K$. This is the basis of the *internal approach* or lower bound approach to the extreme loads. Although the extreme loads are but upper bound estimates of the actual ultimate loads of the system, it must be emphasized that since they are independent of the material behaviour characteristics, other than the strength criteria, and of the loading paths and loading history, they are valid regardless of these data. Thus, after being assessed and through the introduction of convenient safety factors, they provide a reliable theoretical benchmark for practical applications.

The construction of stress fields that satisfy the conditions above often makes the implementation of the internal approach difficult. Dualization through the principle of virtual rates of work (see Equations (1) and (2)) leads to an external approach based upon the construction of kinematically admissible virtual velocity fields. The details of the related reasoning can be found in [Salençon 1990]. The key idea is that the material resistance may be equivalently defined through the strength criteria on the stress tensor, as indicated above, or through the associated π -functions of the strain rate $\hat{\mathbf{d}}$: denoting generically by G the domain of resistance on $\boldsymbol{\sigma}$ defined at a point of the system, the corresponding π -function is just the *support function*

$$\pi(\hat{\mathbf{d}}) = \sup\{\boldsymbol{\sigma}' : \hat{\mathbf{d}} \mid \boldsymbol{\sigma}' \in G\}, \quad (3)$$

from which we derive

$$\pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) = \sup\{\mathbf{n} \cdot \boldsymbol{\sigma}' \cdot \llbracket \hat{\mathbf{U}} \rrbracket \mid \boldsymbol{\sigma}' \in G\}. \quad (4)$$

The π -functions are called the maximum resisting rate of work densities related to G . From the definition of K and through Equations (1) and (2), the fundamental inequality of the external approach is obtained in the form: For every $\mathbf{Q} \in K$, and every kinematically admissible virtual velocity field $\hat{\mathbf{U}}$,

$$\mathbf{Q} \cdot \hat{\mathbf{q}} \leq \int_{\Omega} \pi(\hat{\mathbf{d}}) d\Omega + \int_{\Sigma_{\hat{\mathbf{U}}}} \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) d\Sigma_{\hat{\mathbf{U}}}. \quad (5)$$

The right side of Equation (5) is called the maximum resisting rate of work in the virtual velocity field $\hat{\mathbf{U}}$:

$$P_{\text{rm}}(\hat{\mathbf{U}}) = \int_{\Omega} \pi(\hat{\mathbf{d}}) d\Omega + \int_{\Sigma_{\hat{\mathbf{U}}}} \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) d\Sigma_{\hat{\mathbf{U}}}, \quad (6)$$

while the left side of (5) is just the rate of work of all the external forces $P_e(\mathbf{Q}, \hat{\mathbf{U}})$. It follows from Equations (5) and (6) that the construction of any kinematically admissible virtual velocity field yields an external approximation of the boundary of K :

$$K \subset \{P_e(\mathbf{Q}, \hat{\mathbf{U}}) - P_{\text{rm}}(\hat{\mathbf{U}}) \leq 0\}, \quad \text{for any kinematically admissible } \hat{\mathbf{U}}. \quad (7)$$

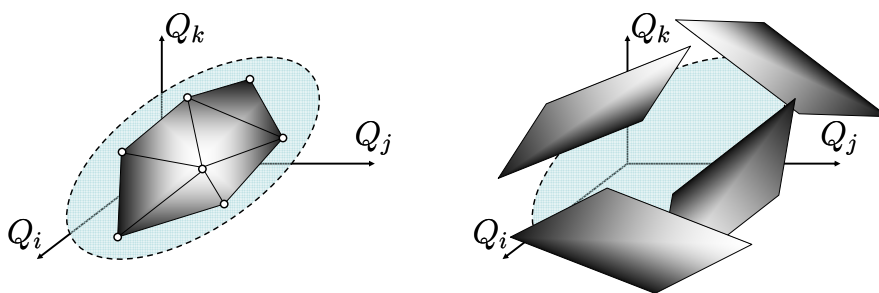


Figure 1. Internal and external approaches to the domain of potentially safe loads.

The two approaches are schematically presented in [Figure 1](#). Important points regarding the external approach are:

- Tables giving the expressions of the π -functions for usually encountered criteria are available [[Salençon 1983; 2002](#)].
- For a given G the values of the π -functions are either finite or infinite depending on the values of the arguments \hat{d} and $(n, \llbracket \hat{U} \rrbracket)$.
- For the approach to be efficient, kinematically admissible virtual velocity fields \hat{U} must be chosen in order that the values of the π -functions remain finite everywhere in O .
- The third condition has no relationship whatsoever with a constitutive law.

One may wonder how the external approach can be of any use in practical applications for the design of structures since it only provides upper bounds of the extreme loads, the practical significance of which has been discussed above. As a matter of fact, this is one reason for the introduction of the *model factor* in ultimate limit state design codes [[Salençon 1994](#)]. Independently of the partial safety factors which define the design loads and design resistances to be introduced in the design procedure of the considered structure, derived from the nominal ones, the model factor is imposed on the resisting rate of work as a whole in order to take into account the method for the yield design analysis that is performed (say, by assessing the *quality* of the considered potential collapse mechanisms).

Seismic bearing capacity of a circular footing on a purely cohesive soil

Problem motivation. The problem under consideration arises from a series of field observations after several major earthquakes within the last twenty-five years, which revealed a particular type of foundation failure without the presence of liquefaction in the supporting soil layers: large permanent rotations were observed at the foundation level together with a zone of detachment at the soil-foundation interface and the development of a failure mechanism within the soil volume [[Mendoza and Auvinet 1988](#)]. The same failure mechanism was also identified experimentally [[Knappett et al. 2006; Zeng and Steedman 1998](#)].

From a theoretical point of view, one initial approach to the problem of the seismic bearing capacity is to work within the classical framework of Terzaghi's bearing capacity formula, modifying the bearing capacity factors in order to account for the effect of the inertia forces within the soil volume during the seismic excitation, while applying appropriate correction factors for the load eccentricity and inclination

[Fishman et al. 2003; Richards et al. 1993; Sarma and Iossifelis 1990]. A second approach represents the seismic bearing capacity of the foundation system as an *ultimate surface* in the space of the loading parameters of the footing as a function of the intensity of the horizontal inertia forces in the soil volume [Paolucci and Pecker 1997a; 1997b; Pecker and Salençon 1991; Salençon and Pecker 1995a; 1995b]. The results so obtained in the case of shallow strip footings were incorporated in the European norms for earthquake-resistant design of civil engineering structures [Eurocode 1998].

This study aims at extending the analysis to the seismic bearing capacity of a shallow circular footing resting on the surface of a purely cohesive soil layer.

Definition of the seismic bearing capacity problem. A rigid circular footing with radius r resting on a purely cohesive soil half-space is considered (Figure 2). The resistance of the constituent soil is described by the Tresca criterion with a cohesion c depending linearly on the depth (8), with c_0 being the surface cohesion and α the vertical cohesion gradient:

$$c = c_0 + \alpha z. \quad (8)$$

In order to assess the importance of the corresponding assumption, two extreme cases are considered, namely, the classical Tresca criterion (9) and the Tresca criterion with zero resistance to tension (10):

$$f(\boldsymbol{\sigma}) = |\sigma_1 - \sigma_3| - 2c \leq 0, \quad (9)$$

$$f(\boldsymbol{\sigma}) = \sup\{|\sigma_1 - \sigma_3| - 2c, \sigma_1\} \leq 0, \quad (10)$$

with σ_1 and σ_3 being the major and minor principal stresses respectively (tensile stresses positive).

The soil-footing interface is also considered to be purely cohesive and its resistance is modelled by the Tresca criterion with no tensile resistance (zero tension *cut-off*). This is a deliberate choice in order to allow for the potential creation of a zone of detachment between the footing and the soil, an essential characteristic of observed seismic bearing capacity failure. The interface cohesion is considered equal to c_0 .

$$f(\sigma, \tau) = \sup\{|\tau| - c_0, \sigma\} \leq 0. \quad (11)$$

The quasistatic loading mode of the system is defined by means of the wrench of external forces acting on the footing due to the weight and to the inertial response of the superstructure (including the footing itself), of the unit weight of the soil, and of the intensity of the inertia forces developing within the soil mass.

Following common practice, this intensity is assumed to be uniform throughout the soil mass with vertical and horizontal components F_v and F_h . The physical validity of this assumption has already been discussed by various authors (for example, Pecker and Salençon [1991] suggested that, denoting by d

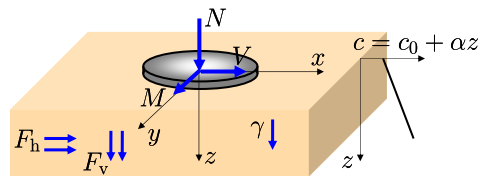


Figure 2. Circular shallow foundation under seismic loading on a purely cohesive soil.

and D the failure mechanism thickness and the depth of the soil layer, the following condition should be satisfied: $d/D < 10$) and will be revisited in a following section. From now on the vertical component will be added to the unit weight of the soil and will give rise to the modified unit weight:

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma} + \mathbf{F}_v = \gamma^* \mathbf{e}_z, \quad (12)$$

with \mathbf{e}_z the unit vector in the descending vertical direction.

Due to the origin of the external loads acting on the superstructure and on the foundation, it is also assumed that the horizontal component V of the resultant force of the wrench is collinear with the horizontal inertia force F_h in the soil along the x -axis, and that the horizontal overturning moment M at the center of the footing is oriented around the y -axis, perpendicular to that direction. As a matter of fact, this assumption is rigorously valid when dealing with the excitation of a single-degree-of-freedom (SDOF) superstructure and under specific conditions in the case of multiple degrees of freedom.

Relevant variables for the seismic bearing capacity problem. The determination of the bearing capacity of the foundation under the conditions specified above is based upon the theory of yield design. Concerning the influence of the modified unit weight on the extreme loads, it has been shown [Salençon 1983] that, for the classical Tresca criterion, the unit weight has no influence on the value of the extreme loads supported by the footing. The result holds also if the modified unit weight is not constant but depends only on the z -coordinate. For the case of the Tresca criterion with no resistance to tension, the result remains true if $\gamma^* \leq 0$, which is true in the usual cases of seismic excitations. Therefore, γ^* will no longer appear in the problem loading parameters.

Dimensionless parameters. The vertical component of the resultant force acting on the footing is denoted by N . The horizontal component F_h is defined from a horizontal acceleration a_h characteristic of the examined earthquake, which can be for instance the peak ground horizontal acceleration (PGHA):

$$F_h = \rho a_h, \quad (13)$$

with ρ being the mass density of the soil.

The vector \mathbf{Q} representing the loading parameters of the system is written as

$$\mathbf{Q} = (N, V, M, F_h), \quad (14)$$

and the strength parameters are c_0 and α .

When presenting the results, dimensionless parameters will be introduced:

$$\tilde{N} = \frac{N}{\pi c_0 r^2}, \quad \tilde{V} = \frac{V}{\pi c_0 r^2}, \quad \tilde{M} = \frac{M}{2\pi c_0 r^3}, \quad \tilde{F}_h = \frac{\rho r a_h}{\pi c_0}, \quad \tilde{k} = \frac{r\alpha}{c_0}. \quad (15)$$

The parameter \tilde{k} expresses the degree of heterogeneity in the system. For a homogeneous soil layer, $\tilde{k} = 0$. Common values of r , α and c_0 give rise to \tilde{k} less than or equal to approximately 2.

Solution procedure. The external approach is implemented in this problem through the construction of kinematically admissible virtual velocity fields in the whole system that are relevant to the strength criteria under consideration by referring to the π -functions corresponding to Equations (9)–(11):

- Classical Tresca criterion (Equation (9)):

$$\begin{aligned}
 \pi(\hat{\mathbf{d}}) &= +\infty && \text{if } \text{tr}(\hat{\mathbf{d}}) \neq 0, \\
 \pi(\hat{\mathbf{d}}) &= c(|\hat{d}_1| + |\hat{d}_2| + |\hat{d}_3|) && \text{if } \text{tr}(\hat{\mathbf{d}}) = 0, \\
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= +\infty && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} \neq 0, \\
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= c|\llbracket \hat{\mathbf{U}} \rrbracket| && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} = 0.
 \end{aligned} \tag{16}$$

- Tresca criterion with zero tension cut-off (Equation (10)):

$$\begin{aligned}
 \pi(\hat{\mathbf{d}}) &= +\infty && \text{if } \text{tr}(\hat{\mathbf{d}}) < 0, \\
 \pi(\hat{\mathbf{d}}) &= c(|\hat{d}_1| + |\hat{d}_2| + |\hat{d}_3| - \text{tr}(\hat{\mathbf{d}})) && \text{if } \text{tr}(\hat{\mathbf{d}}) \geq 0, \\
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= +\infty && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} < 0, \\
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= c|\llbracket \hat{\mathbf{U}} \rrbracket - \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n}| && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} \geq 0.
 \end{aligned} \tag{17}$$

- Tresca criterion for the interface without resistance to tension (Equation (11)):

$$\begin{aligned}
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= +\infty && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} < 0, \\
 \pi(\mathbf{n}, \llbracket \hat{\mathbf{U}} \rrbracket) &= c|\llbracket \hat{\mathbf{U}} \rrbracket - (\llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n}) \cdot \mathbf{n}| && \text{if } \llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} \geq 0.
 \end{aligned} \tag{18}$$

It follows obviously that a relevant virtual velocity field for the classical Tresca criterion is also relevant for the Tresca criterion without tensile strength, but not vice versa. Attention must be paid also to Equation (18) as it shows that a nonzero maximum resisting rate of work is obtained even when a detachment between the footing and the soil is induced by the virtual velocity field.

Since the circular footing is assumed to be perfectly rigid, any kinematically admissible virtual velocity $\hat{\mathbf{U}}$ field must comply with a rigid body motion of the footing as a boundary condition. For the planar velocity fields that will be considered hereafter, assuming $\hat{U}_y = 0$, such a rigid body motion is defined by the virtual rate of rotation $\hat{\omega}$ and the two components $\hat{U}_{O,x}$, $\hat{U}_{O,z}$ of the virtual velocity of the center O of the footing. Consequently the rate of work of the external forces is written:

$$\mathbf{Q} \cdot \hat{\mathbf{q}} = N\hat{U}_{O,z} + V\hat{U}_{O,x} + M\hat{\omega} + F_h \int_{\Omega} \hat{\mathbf{U}} \cdot \mathbf{e}_x dO. \tag{19}$$

Three classes of kinematically admissible virtual velocity fields $\hat{\mathbf{U}}$ have been examined, which are derived from plane strain potential failure mechanisms used to determine the seismic bearing capacity of strip footings. Completely described in [Chatzigogos 2007; Chatzigogos et al. 2007], these planar and nonplane strain virtual velocity fields are parallel to Oxz and depend on the three coordinates. They are relevant to the strength criteria (Equations (9)–(11)) since they are isochoric everywhere in Ω : $\text{tr}(\hat{\mathbf{d}}) = 0$, $\llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} = 0$, but for the interface where virtual uplift of the footing with respect to the soil surface may occur $\llbracket \hat{\mathbf{U}} \rrbracket \cdot \mathbf{n} \geq 0$. In order to implement the external approach through Equation (7) $P_{\text{rm}}(\hat{\mathbf{U}})$ must be computed, which implies deriving $\hat{\mathbf{d}}$ from $\hat{\mathbf{U}}$, a tedious task until it was drastically simplified by Puzrin and Randolph [2003a; 2003b] in a method based upon the use of wisely chosen curvilinear coordinates.

- *Translational* virtual failure mechanisms were originally proposed in plane strain [Green 1954] for the indentation by a rigid punch submitted to an inclined load; the extension to a rigid circular footing was given by Puzrin and Randolph [2003b] by considering that the width of the mechanism in a cross-section by a vertical plane is proportional to the width of the footing in the same cross-section (Figure 3). The footing translates with a virtual velocity \hat{U}_0 which propagates with a constant magnitude along the streamlines of the mechanism. The shape of each virtual failure mechanism is defined by two angles: $0 < \delta < \pi/2$, $0 < \epsilon < \pi/2$. The mechanisms exhibit three zones within the soil mass, presented in Figure 3. Zones 1 and 3 translate rigidly while zone 2 is a region where nonplane shear strain rate is developed. Contributions to the maximum resisting rate of work for this class are developed within the volume of zone 2 and along the velocity jump surface in the soil. These virtual mechanisms involve no rotation of the footing: the upper bound estimates they provide through (7) do not involve the moment M .
- *Purely rotational* virtual failure mechanisms are adapted from the plane strain version studied in [Salençon and Pecker 1995a; 1995b; Sekiguchi and Kobayashi 1997]. The rigid circular footing is considered to rotate rigidly around an axis parallel to Oy and induces rigid rotational failure of the soil below with a virtual angular velocity $\hat{\omega}$ (Figure 4). Each mechanism is defined by the geometrical parameters κ and λ . For $1 < \lambda < 2$, there is no uplift of the footing with respect to the soil surface and the maximum resisting rate of work is only produced along the velocity jump surface within the soil volume. For $0 < \lambda < 1$, uplift of the footing with respect to the soil surface takes place in the zone of soil-footing detachment where consequently a fraction of the maximum resisting rate of work is developed.
- *Shear-rotational* virtual failure mechanisms follow a pattern derived from the plane-strain virtual velocity field originally proposed by Hansen [1953] for the study of active earth pressures (see Figure 5).

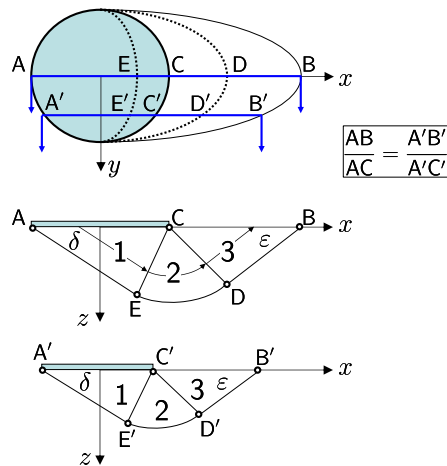


Figure 3. Translational virtual failure mechanism.

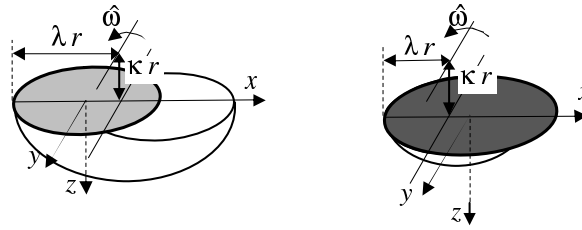


Figure 4. Purely rotational virtual failure mechanism.

Rigid body rotation of the footing around an axis of rotation parallel to Oy located within the soil mass induces the development of nonplanar shear strain rate in the zones 2 and 3 within the soil volume. The mechanisms depend on the three geometrical parameters κ, λ, μ and three distinct configurations are obtained depending on the position of the axis of rotation with respect to the footing: without uplift of the footing with respect to the soil surface or with a small or large zone of detachment; contributions to the maximum resisting rate of work are developed within the soil volume in zones 2 and 3, along the velocity jump surface in the soil mass and on the zone of soil-footing detachment, if any.

Results

Implementing the external approach through Equation (7), the rate of work of the external forces is given by (19) and the maximum resisting rate of work $P_{rm}(\hat{U})$ is computed through (6) with the relevant expressions for $\pi(\hat{d})$ and $\pi(\mathbf{n}, \|\hat{U}\|)$. Looking for optimal upper bounds for the ultimate loads supported by the system requires the implementation of an optimization procedure over each separate geometrical configuration of the three considered classes of virtual mechanisms: in each case, the mathematical problem involved reduces to minimizing a nonlinear objective function with respect to the parameters of the geometric configuration. The algorithm used for this purpose is described in [Coleman and Li 1996]. The following results were obtained.

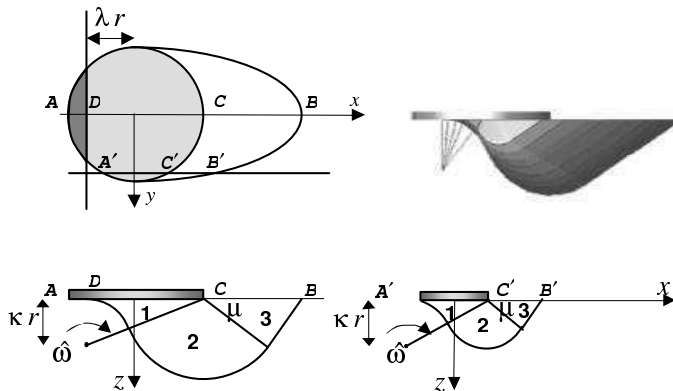


Figure 5. Shear-rotational virtual failure mechanism.

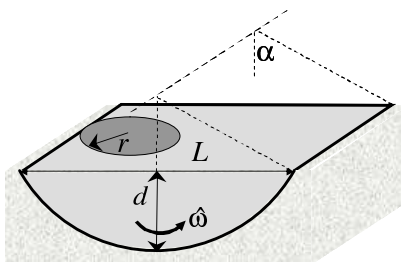


Figure 6. Plane strain circular rigid body virtual collapse mechanism.

Critical value of F_h . The assumption of uniformly distributed horizontal inertia forces F_h would lead to a pathological conclusion if, in addition to the limitation on the ratio d/D which has already been introduced for physical relevance, the virtual failure mechanisms to be taken in consideration were given the possibility of infinite extension.

As an illustrative example, the plane strain circular rigid body virtual collapse mechanism presented in Figure 6 can be considered. In the case of a homogeneous soil layer, denoting by L and d its horizontal and vertical extension respectively, the minimization procedure on this class of mechanisms proves the instability of the system just due to the action of the soil inertia forces if $\rho a_h > 8.3c_0/L = 2.74c_0/d$, with the other loading parameters being zero: $\tilde{N} = 0$, $\tilde{V} = 0$, $\tilde{M} = 0$. As a consequence, for any circular footing with radius $r \leq L/2$, the value $\tilde{F}_h = 2.64r/L = 0.87d/L$ would appear as the critical value for the considered mechanism. This critical value tends to zero if the extension of the mechanism is unlimited. The value $\tilde{F}_h = 0$ would therefore appear as the critical value of \tilde{F}_h , a conclusion that is obviously unrealistic! This brings us to the condition that the virtual failure mechanisms implemented in the yield design theory must be physically relevant, apart from the fact that horizontal uniformity of \tilde{F}_h could also be questioned. In other words, pathological circumstances are put aside through realistic constraints on the minimization procedure on the virtual failure mechanisms.

Table 1 summarizes the calculated values of critical \tilde{F}_h as a function of \tilde{k} for the three classes of examined three-dimensional virtual failure mechanisms. As could be anticipated, it shows that the critical value increases with \tilde{k} . The minimum value is obtained for the homogeneous soil ($\tilde{k} = 0$) through a shear-rotational virtual failure mechanism. For usual values of the other parameters such as $r = 4$ m, $c_0 = 20$ kPa, $\rho = 2 \times 10^3$ kg/m³ it corresponds to a very strong earthquake: $a_h = 5$ m/sec² $\cong 0.5g$.

| Class of mechanisms | Critical value of \tilde{F}_h | | | |
|---------------------|---------------------------------|-------------------|-----------------|-----------------|
| | $\tilde{k} = 0$ | $\tilde{k} = 0.5$ | $\tilde{k} = 1$ | $\tilde{k} = 3$ |
| Translational | 1.32 | 1.80 | 2.29 | 4.05 |
| Purely rotational | 0.99 | 1.28 | 1.54 | 2.45 |
| Shear rotational | 0.66 | 0.90 | 1.15 | 2.03 |
| MINIMUM | 0.66 | 0.90 | 1.15 | 2.03 |

Table 1. Critical values of \tilde{F}_h as a function of \tilde{k} .

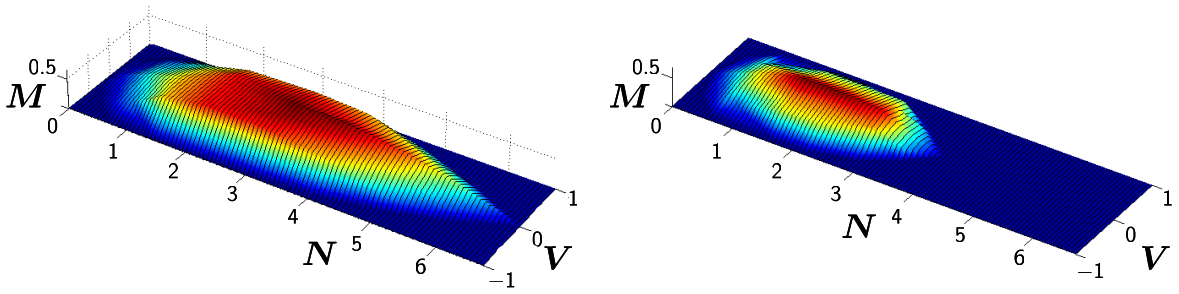


Figure 7. External approach of the ultimate load surface for a homogeneous soil ($\tilde{k} = 0$) when $\tilde{F}_h = 0$ (left) and $\tilde{F}_h = 0.5$ (right).

Presentation of the results. The upper bounds for the ultimate loads supported by the foundation are represented as surfaces in the space of the loading parameters (\tilde{N} , \tilde{V} , \tilde{M}) for different values of \tilde{F}_h and \tilde{k} (Figure 7).

Since these surfaces are obtained through the minimization procedure described earlier over three different classes of virtual collapse mechanisms, they are just piecewise regular, each smooth part being the result of one class of mechanisms as it will appear in the interaction diagrams (Figures 8–9).

For practical applications the most convenient presentation of the results is by means of sections of those surfaces that may be called classically *interaction diagrams*: these curves represent the upper bounds for the ultimate combinations of the loading parameters and indicate the class of mechanisms from which each bound is obtained. In the following, the results refer to the Tresca criterion with a zero tension cut-off for the soil strength which seems the most realistic criterion to be considered. For brevity's sake, only some significant results will be reported here; a comprehensive and commented report may be found in [Chatzigogos 2007; Chatzigogos et al. 2007].

Interaction diagram (\tilde{N} , \tilde{V} , $\tilde{M} = 0$, \tilde{F}_h). The diagrams in Figure 8 present the relation between the ultimate horizontal and vertical force for $\tilde{k} = 0$ and $\tilde{k} = 1$, for three different values of \tilde{F}_h . The maximum value for \tilde{V} is 1, corresponding to a translational mechanism of failure by pure sliding along the soil-footing interface when $\tilde{F}_h = 0$. It is also interesting to note that for $\tilde{F}_h > 0$ the purely rotational virtual mechanism gives an upper bound that is slightly better than the pure sliding one: although the depth of this mechanism is relatively small, so that it is very close to a pure sliding, it incorporates a contribution of \tilde{F}_h to the rate of work of the external forces. This phenomenon is less pronounced for larger values of \tilde{k} as revealed by the diagram for $\tilde{k} = 1$. For practical applications it is worth noting that, both for $\tilde{k} = 0$ and for $\tilde{k} = 1$, the effect of \tilde{F}_h remains negligible as long as $\tilde{N}_{\max}^0 / \tilde{N} > 2.5$, where \tilde{N}_{\max}^0 denotes the known exact value of the maximal vertical force supported by the footing with $\tilde{F}_h = 0$ [Salençon and Matar 1982]. As \tilde{N} increases so does the negative effect of a high value of \tilde{F}_h , especially when $\tilde{N}_{\max}^0 / \tilde{N} > 2i$ but it is observed that it is less pronounced for $\tilde{k} = 0$ than for $\tilde{k} = 1$: a favourable effect of the vertical cohesion gradient.

Interaction diagram (\tilde{N} , \tilde{V} , $\tilde{M} = 0$, \tilde{F}_h). Figure 10 presents the optimal upper bounds for the ultimate combinations of \tilde{M} and \tilde{N} obtained with $\tilde{V} = 0$, for $\tilde{k} = 0$ and $\tilde{k} = 1$. Experimental data in the case $\tilde{k} = 0$, $\tilde{F}_h = 0$ in [Houlsby and Martin 1993] are plotted for comparison.

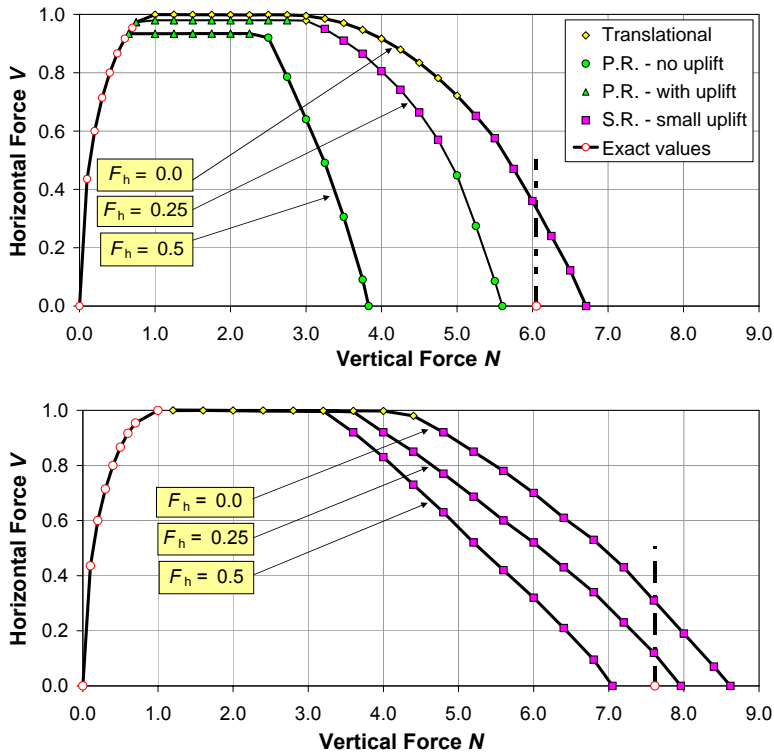


Figure 8. Interaction diagram ($\tilde{N}, \tilde{V}, \tilde{M} = 0, \tilde{F}_h$) for $\tilde{k} = 0$ (top) and $\tilde{k} = 1$ (bottom). Tresca criterion with zero tension cut-off.

It comes out that the upper bounds are satisfactory, from a practical point of view, and that the difference is larger for the larger values of \tilde{N} that correspond to quasiasymmetric loading configurations, to which the considered unilateral virtual failure mechanisms are not well suited. For small values of \tilde{N} , the optimal upper bounds are obtained by mechanisms with a significant zone of detachment of the footing, which is not the case as \tilde{N} increases. The effect of \tilde{F}_h follows the same behaviour as in Figure 8, which, from a practical point of view, enforces the conclusion from this observation that a factor of safety against permanent loads $\tilde{N}_{max}^0/\tilde{N} > 2.5$ can guarantee that the effect of soil inertia forces is negligible, even for very strong earthquakes. Such conclusions are in agreement with observations of real foundation bearing capacity failures, mainly after the Guerrero–Michoacán earthquake (Mexico, 1985) as presented in [Mendoza and Auvinet 1988].

Interaction diagram ($\tilde{N} = \text{const}, \tilde{V}, \tilde{M}, \tilde{F}_h$). The interaction diagrams between the ultimate values of \tilde{V} and \tilde{M} for fixed values of the vertical force \tilde{N} are shown in Figure 9 for $\tilde{k} = 0$ in the two cases $\tilde{N}_{max}^0/\tilde{N} = 3$ corresponding to a proper foundation design and $\tilde{N}_{max}^0/\tilde{N} = 1.5$ corresponding to a nonconservative design.

These diagrams can be used for practical applications when a relationship between the resultant moment and the horizontal force (base shear force) on the footing is known from the geometrical and

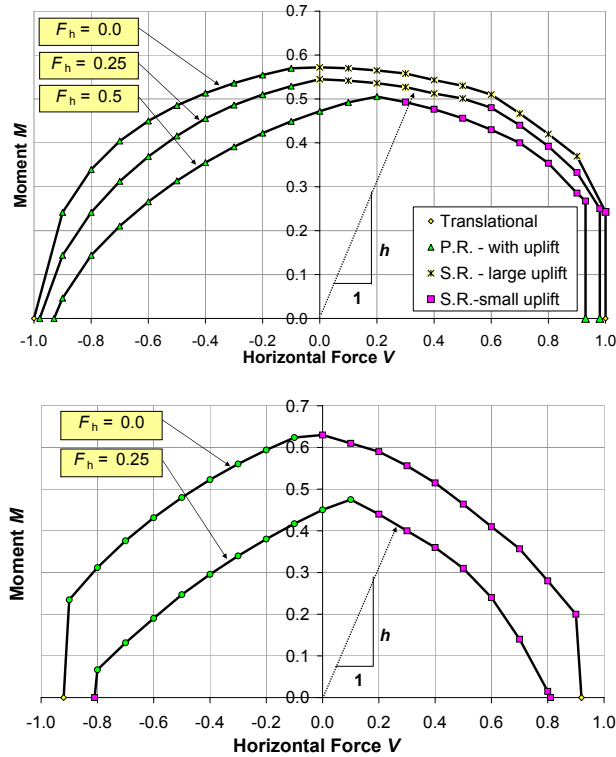


Figure 9. Tresca criterion with zero tension cut-off: interaction diagram ($\tilde{N} = \text{const}$, \tilde{V} , \tilde{M} , \tilde{F}_h) for $\tilde{k} = 0$, $\tilde{N}_{\max}^0/\tilde{N} > 3$ (top) and $\tilde{N}_{\max}^0/\tilde{N} > 1.5$ (bottom); ultimate loading paths for $\tilde{F}_h = 0.25$ in a SDOF superstructure.

rigidity characteristics of the superstructure. Such a relationship defines a loading path in the (\tilde{V}, \tilde{M}) -plane allowing for the determination of the ultimate combination of \tilde{V} and \tilde{M} for given \tilde{N} and \tilde{F}_h as presented in Figure 10, it being observed that only the right halves of the diagrams, $\tilde{V} \geq 0$, correspond to realistic loading paths. The two diagrams highlight the significant decrease of the bearing capacity with increasing \tilde{F}_h for $\tilde{N}_{\max}^0/\tilde{N} = 1.5$. Even for $\tilde{V} = 0$, $\tilde{M} = 0$, a value of $\tilde{F}_h = 0.5$ causes the collapse of the footing.

Practical implementation. As explained earlier, the Eurocode [1998] expression for the seismic bearing capacity of shallow foundations is only valid for strip footings resting on homogeneous soils, either purely cohesive or purely frictional. This study makes it possible to propose a modified version of those rules for shallow circular footings on a purely cohesive soil with a vertical cohesion gradient. The corresponding expression is presented and discussed in [Chatzigogos 2007; Chatzigogos et al. 2007]. As far as the design principles are concerned, it should be remembered that the effect of the horizontal inertia forces in the soil volume is, in general, negligible as long as the vertical force on the footing remains smaller than one third of its static bearing capacity.

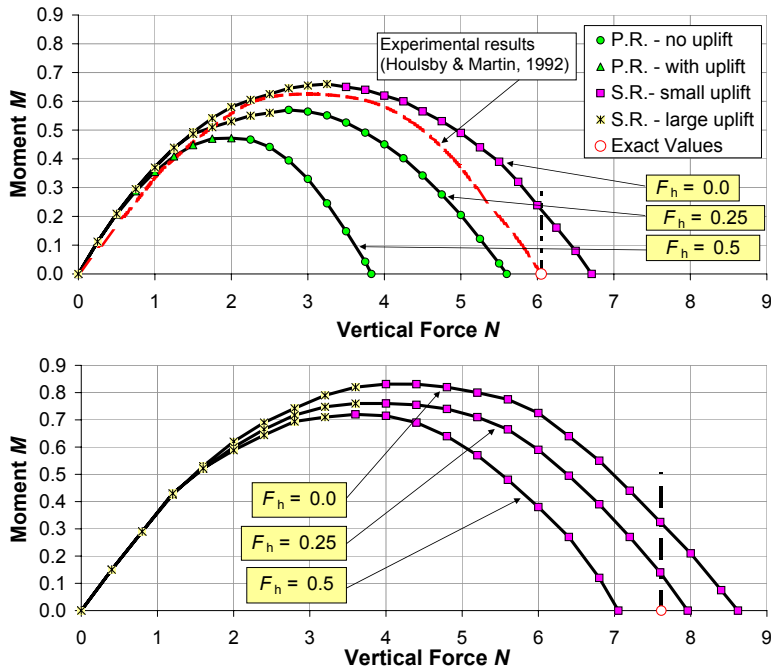


Figure 10. Interaction diagram (\tilde{N} , $\tilde{V} = 0$, \tilde{M} , \tilde{F}_h) for $\tilde{k} = 0$ (top) and $\tilde{k} = 1$ (bottom), and experimental results by Houlsby and Martin [1993]. Tresca criterion with zero tension cut-off.

Conclusion

Seismic actions must now often be taken into account when designing private or industrial buildings or structures such as bridges, dams, nuclear plants, etc. We have briefly outlined how such a problem can be thoroughly studied from the theoretical point of view through the yield design approach up to the writing of new international design codes. For this purpose it is necessary that the rationale of the yield design theory be thoroughly understood since most (not to say all) equations that are used in such codes are obtained through the external approach, either explicitly or implicitly. To this extent the yield design theory is a corner stone of ultimate limit state design (ULSD).

References

- [Chatzigogos 2007] C. T. Chatzigogos, *Comportement sismique des fondations superficielles: vers la prise en compte d'un critère de performance dans la conception*, Ph.D. thesis, École Polytechnique Palaiseau, Palaiseau, France, 2007, Available at <http://www.imprimerie.polytechnique.fr/Theses/Files/Chatzigogos.pdf>.
- [Chatzigogos et al. 2007] C. T. Chatzigogos, A. Pecker, and J. Salençon, “Seismic bearing capacity of a circular footing on a heterogeneous cohesive soil”, *Soils Found.* **47**:4 (2007), 783–787.
- [Coleman and Li 1996] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds”, *SIAM J. Optim.* **6**:2 (1996), 418–445.
- [Eurocode 1998] “Calcul des structures pour leur résistance aux séismes”, EUROCODE 8, Commission Européenne de Normalisation, 1998. Partie 5.
- [Fishman et al. 2003] K. L. Fishman, R. Richards Jr., and D. Yao, “Inclination factors for seismic bearing capacity”, *J. Geotech. Geoenviron.* **129**:9 (2003), 861–865.

- [Green 1954] A. P. Green, “The plastic yielding of metal junctions due to combined shear and pressure”, *J. Mech. Phys. Solids* **2**:3 (1954), 197–211.
- [Hansen 1953] J. B. Hansen, *Earth pressure calculation: application of a new theory of rupture to the calculation and design of retaining walls, anchor slabs, free sheet walls, anchored sheet walls, fixed sheet walls, braced walls, double sheet walls and cellular cofferdams*, Danish Technical Press — Institution of Danish Civil Engineers, Copenhagen, 1953.
- [Houlsby and Martin 1993] G. T. Houlsby and C. M. Martin, “Modelling of the behaviour of foundations of jack-up units on clay”, pp. 339–358 in *Predictive soil mechanics*, edited by G. T. Houlsby and A. N. Schofield, Thomas Telford, London, 1993.
- [Knappett et al. 2006] J. A. Knappett, S. K. Haigh, and S. P. G. Madabhushi, “Mechanisms of failure for shallow foundations under earthquake loading”, *Soil Dyn. Earthq. Eng.* **26**:2-4 (2006), 91–102.
- [Mendoza and Auvinet 1988] M. J. Mendoza and G. Auvinet, “The Mexico earthquake of September 19, 1985 — behavior of building foundations in Mexico City”, *Earthq. Spect.* **4**:4 (1988), 835–852.
- [Paolucci and Pecker 1997a] R. Paolucci and A. Pecker, “Soil inertia effects on the bearing capacity of rectangular foundations on cohesive soils”, *Eng. Struct.* **19**:8 (1997), 637–643.
- [Paolucci and Pecker 1997b] R. Paolucci and A. Pecker, “Seismic bearing capacity of shallow strip foundations on dry soils”, *Soils Found.* **37**:3 (1997), 95–105.
- [Pecker and Salençon 1991] A. Pecker and J. Salençon, “Seismic bearing capacity of shallow strip foundations on clay soils”, pp. 287–304 in *Proceedings of International Workshop on Seismology and Earthquake Engineering (Mexico City)*, CENAPRED Mexico City, 1991.
- [Puzrin and Randolph 2003a] A. M. Puzrin and M. F. Randolph, “Generalized framework for three-dimensional upper bound limit analysis in a tresca material”, *J. Appl. Mech. (Trans. ASME)* **70**:1 (2003), 91–100.
- [Puzrin and Randolph 2003b] A. M. Puzrin and M. F. Randolph, “New planar velocity fields for upper bound limit analysis”, *Int. J. Solids Struct.* **40**:13-14 (2003), 3603–3619.
- [Richards et al. 1993] R. J. Richards, D. G. Elms, and M. Budhu, “Seismic bearing capacity and settlements of foundations”, *J. Geotech. Eng. ASCE* **119**:4 (1993), 662–674.
- [Salençon 1983] J. Salençon, *Calcul à la rupture et analyse limite*, Presses de l’E.N.P.C., Paris, 1983.
- [Salençon 1990] J. Salençon, “An introduction to the yield design theory and its applications to soil mechanics”, *Eur. J. Mech. A Solid.* **9**:5 (1990), 477–500.
- [Salençon 1994] J. Salençon, “Approche théorique du calcul aux états limites ultimes”, pp. 701–722 in *Les grands systèmes des sciences et de la technologie*, edited by J.-L. L. J. Horowitz and, Masson, Paris, 1994.
- [Salençon 2001] J. Salençon, *Handbook of continuum mechanics: general concepts — thermoelasticity*, Springer, Berlin, 2001.
- [Salençon 2002] J. Salençon, *De l’élastoplasticité au calcul à la rupture*, Éditions de l’École Polytechnique, Palaiseau, 2002.
- [Salençon and Matar 1982] J. Salençon and M. Matar, “Capacité portante des fondations superficielles circulaires”, *J. Mec. Theor. Appl.* **1**:2 (1982), 237–267.
- [Salençon and Pecker 1995a] J. Salençon and A. Pecker, “Ultimate bearing capacity of shallow foundations under inclined and eccentric loads, I: purely cohesive soil”, *Eur. J. Mech. A Solid.* **14**:3 (1995), 349–375.
- [Salençon and Pecker 1995b] J. Salençon and A. Pecker, “Ultimate bearing capacity of shallow foundations under inclined and eccentric loads, II: purely cohesive soil without tensile strength”, *Eur. J. Mech. A Solid.* **14**:3 (1995), 377–396.
- [Sarma and Iossifelis 1990] S. K. Sarma and I. S. Iossifelis, “Seismic bearing capacity factors of shallow strip footings”, *Géotechnique* **40** (1990), 265–273.
- [Sekiguchi and Kobayashi 1997] H. Sekiguchi and S. Kobayashi, “Limit analysis of bearing capacity for a circular footing subjected to eccentric loadings”, pp. 1029–1032 in *Proceedings of the 14th International Conference on Soil Mechanics and Foundation Engineering - ICSMFE (Hambourg, 1997)*, vol. 2, Balkema, Rotterdam, 1997.
- [Zeng and Steedman 1998] X. Zeng and R. S. Steedman, “Bearing capacity failure of shallow foundations in earthquakes”, *Géotechnique* **48**:2 (1998), 235–256.

Received 11 Jan 2008. Revised 29 Jun 2008. Accepted 29 Jun 2008.

JEAN SALENÇON: salencon@lms.polytechnique.fr

Laboratoire de Mécanique des Solides, UMR 7649, École Polytechnique, 91128 Palaiseau Cedex, France

CHARISIS THEODOROU CHATZIGOGOS: charisis.chatzigogos@wanadoo.fr

Laboratoire de Mécanique des Solides, UMR 7649, École Polytechnique, 91128 Palaiseau Cedex, France

ALAIN PECKER: pecker@lms.polytechnique.fr

Geodynamique et Structure, 157 rue des Blains, 92220 Bagneux, France

SUBMISSION GUIDELINES

ORIGINALITY

Authors may submit manuscripts in PDF format on-line. Submission of a manuscript acknowledges that the manuscript is *original and has neither previously, nor simultaneously, in whole or in part, been submitted elsewhere*. Information regarding the preparation of manuscripts is provided below. Correspondence by email is requested for convenience and speed. For further information, write to

jomms.steele@stanford.edu

LANGUAGE

Manuscripts must be in English. A brief abstract of about 150 words or less must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate), and email address if available. A home-page URL is optional.

FORMAT

Authors are encouraged to use L^AT_EX and the standard article class, but submissions in other varieties of T_EX, and, exceptionally in other formats, are acceptable. Electronic submissions are strongly encouraged in PDF format only; after the refereeing process we will ask you to submit all source material.

REFERENCES

Bibliographical references should be listed alphabetically at the end of the paper and include the title of the article. All references in the bibliography should be cited in the text. The use of B^IB_TE_X is preferred but not required. Tags will be converted to the house format (see a current issue for examples), however, in the manuscript, the citation should be by first author's last name and year of publication, e.g. "as shown by Kramer, et al. (1994)". Links will be provided to all literature with known web locations and authors are encouraged to provide their own links on top of the ones provided by the editorial process.

FIGURES

Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple, or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to:

production@mathscipub.org

Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

WHITE SPACE

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

PROOFS

Page proofs will be made available to authors (or to the designated corresponding author) at a web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Journal of Mechanics of Materials and Structures

Volume 4, N° 2 February 2009

Special issue

New trends in the thermomechanical modeling of solids

| | | |
|--|--|------------|
| Preface | ANDREI CONSTANTINESCU and NGUYEN QUOC SON | 187 |
| Corrugation models and the roaring rails enigma: a simple analytical contact mechanics model based on a perturbation of Carter's solution | LUCIANO AFFERRANTE and MICHELE CIAVARELLA | 191 |
| Stability and memory effects in a homogenized model governing the electrical conduction in biological tissues | MICOL AMAR, DANIELE ANDREUCCI, PAOLO BISEGNA and ROBERTO GIANNI | 211 |
| Thermal effects of collisions: Does rain turn into ice when it falls on frozen ground? | ANNA MARIA CAUCCI and MICHEL FRÉMOND | 225 |
| Fields of stored energy associated with localized necking of steel | A. CHRYSOCHOOS, B. WATTRISSE, J.-M. MURACCIOLE and Y. EL KAÏM | 245 |
| Prediction of drying shrinkage beyond the pore isodeformation assumption | OLIVIER P. COUSSY and SÉBASTIEN BRISARD | 263 |
| On the method of virtual power in continuum mechanics | GIANPIETRO DEL PIERO | 281 |
| Numerical exploration of the Dang Van high cycle fatigue criterion: application to gradient effects | F. HOFMANN, G. BERTOLINO, A. CONSTANTINESCU and M. FERJANI | 293 |
| Cyclic approximation of the heat equation in finite strains for the heat build-up problem of rubber | Y. LE CHENADEC, I. RAOULT, C. STOLZ and M.-L. NGUYEN-TAJAN | 309 |
| Inelastic heat fraction evaluation for engineering problems involving dynamic plastic localization phenomena | PATRICE LONGÈRE and ANDRÉ DRAGON | 319 |
| A contribution to a critical review of friction stir welding numerical simulation | O. LORRAIN, J. SERRI, V. FAVIER, H. ZAHROUNI and M. EL HADROUZ | 351 |
| Anisotropic thin-walled beam models: A rational deduction from three-dimensional elasticity | FRANCO MACERI and GIUSEPPE VAIRO | 371 |
| Cyclic behavior and energy approach to the fatigue of shape memory alloys | ZIAD MOUMNI, WAEL ZAKI and HABIBOU MAITOURNAM | 395 |
| Thermomechanical modelling of friction effects in granular flows using the discrete element method | V. D. NGUYEN, J. F., M. GUESSASMA, E. BELLENGER and C. COGNÉ | 413 |
| Seismic bearing capacity of circular footings: a yield design approach | JEAN SALENÇON, CHARISIS THEODOROU CHATZIGOGOS and ALAIN PECKER | 427 |