Davide Bigoni · Angelo Carini
Massimiliano Gei · Alberto Salvadori *Editors*

Fracture Phenomena in Nature and Technology

Proceedings of the IUTAM Symposium on Fracture Phenomena in Nature and Technology held in Brescia, Italy, 1–5 July 2012

Springer

Fracture Phenomena in Nature and Technology

Davide Bigoni • Angelo Carini
Massimiliano Gei • Alberto Salvadori
Editors

Fracture Phenomena in Nature and Technology

Proceedings of the IUTAM Symposium on Fracture Phenomena in Nature and Technology held in Brescia, Italy, 1–5 July 2012

Reprinted from *International Journal of Fracture*, Volume 184, Nos. 1–2 (2013)

Springer

Editors
Davide Bigoni
Massimiliano Gei
Dipartimento di Ingegneria Civile,
Ambientale e Meccanica
Università di Trento
Trento
Italy

Angelo Carini
Alberto Salvadori
Dipartimento di Ingegneria Civile,
Architettura, Territorio,
Ambiente e di Matematica
Università di Brescia
Brescia
Italy

ISBN 978-3-319-04396-8 ISBN 978-3-319-04397-5 (eBook)
DOI 10.1007/978-3-319-04397-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014936174

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

DOI 10.1007/978-3-319-04397-5_1
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9896-2

Foreword

D. Bigoni · A. Carini · M. Gei · A. Salvadori

Published online: 25 October 2013

This Special Issue of the International Journal of Fracture contains selected papers presented at the IUTAM Symposium *Fracture Phenomena in Nature and Technology* that was held at the School of Engineering, University of Brescia, Italy, during the week of July 1–5, 2012. The symposium focused on innovative contributions in fracture research, interpreted broadly to include new engineering and structural mechanics treatments of damage development and crack growth, large-scale failure processes as exemplified by earthquake or landslide failures, ice shelf break-up, and hydraulic fracturing (natural, or for resource extraction or CO_2 sequestration), small-scale rupture phenomena in materials physics including inception of shear banding, void growth, adhesion and decohesion in contact and friction, crystal dislocation processes, and atomic/electronic scale treatment of brittle crack tips and fundamental cohesive properties. The Special Issue manuscripts were limited to original work and were reviewed following the standard procedures of the Journal. The Organizing Committee is grateful to the IUTAM that fostered and supported the symposium, to the financial support from the EU (PIAP-GA-2011-286110-INTERCER2), to the University of Brescia (Italy) and to the University of Trento (Italy) that contributed greatly to the success of the Conference, to the Municipality of Brescia that welcomed the event and the participants.

Guest Editors

D. Bigoni (✉) · A. Carini · M. Gei · A. Salvadori
Trento, Italy
e-mail: bigoni@ing.unitn.it

DOI 10.1007/978-3-319-04397-5_2
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-012-9788-x

ORIGINAL PAPER

Modeling fracture by material-point erosion

A. Pandolfi · B. Li · M. Ortiz

Received: 3 August 2012 / Accepted: 6 November 2012 / Published online: 20 November 2012

Abstract The present work is concerned with the verification and validation of an implementation of the eigenfracture scheme of Schmidt et al. (SIAM J Multiscale Model Simul 7:1237–1266, 2009) based on material-point erosion, which we refer to as *eigenerosion*. Eigenerosion is derived from the general eigenfracture scheme by restricting the eigendeformations in a binary sense: they can be either zero, in which case the local behavior is elastic; or they can be equal to the local displacement gradient, in which case the corresponding material neighborhood is failed, or *eroded*. When combined with a material-point spatial discretization, this scheme gives rise to *material-point erosion*, i. e., each material point can be either intact, in which case its behavior is elastic, or be completely failed—or eroded—and has no load bearing capacity. We verify the eigenerosion scheme through convergence studies for mode I fracture propagation in three-dimensional problems. By way of validation we apply the eigenerosion scheme to the simulation of combined torsion-traction experiments in aluminum-oxide bars.

A. Pandolfi (✉)
Dipartimento di Ingegneria Strutturale, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: pandolfi@stru.polimi.it

B. Li · M. Ortiz
Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: libo@caltech.edu

M. Ortiz
e-mail: ortiz@aero.caltech.edu

Keywords Meshfree approaches · Material point erosion · Eigen fracture · Max-Ent shape functions · Brittle Fracture

1 Introduction

Lagrangian meshfree methods are well-suited to a number of areas of application, such as terminal ballistics, machining, fluid-structure interaction. Lag-rangian meshfree methods offer significant advantages over competing approaches, such as purely Eulerian formulations, particle methods, purely Lagrangian formulations with continuous adaptive remeshing, arbitrary-Lagrangian–Eulerian (ALE). These competing approaches may suffer from a variety of shortcomings, for example: the introduction of large numerical diffusion errors; large discretization errors at fluid-solid interfaces; difficulties in maintaining monotonicity, positivity and in tracking state variables; spurious modes and tensile instabilities; mesh entanglement; the need to remesh or rezone arbitrary three-dimensional domains and the attendant remapping of state variables; ad-hoc transition or blending regions; difficulties in defining numerical integration rules and satisfying essential boundary conditions; unknown convergence and stability properties; and others.

A representative example of Lagrangian meshfree schemes is furnished by the Optimal-Transportation Meshfree (OTM) method of Li et al. (2010).

In the quasistatic setting of interest here, the OTM method combines: (i) Maximum-entropy (max-ent) meshfree interpolation Arroyo and Ortiz (2006) from a nodal-point set; and (ii) material-point sampling (cf., e. g., Sulsky et al. 1994) in order to track the local state of material points, carry out complex constitutive updates and perform spatial integrals. Max-ent interpolation Arroyo and Ortiz (2006) offers the advantage of being meshfree and entirely defined—essentially explicitly—by the current nodal-set positions, thus effectively sidestepping the need for continuous remeshing in simulations of unconstrained flows. In addition, max-ent interpolation satisfies a Kronecker-delta property at the boundary, which greatly facilitates the enforcement of essential boundary conditions, and has good accuracy convergence and monotonicity conditions. Because of interpolatory nature, OTM is free from the tensile numerical instabilities that plague particle methods. In dynamic problems, the OTM method additionally draws on optimal transportation concept, such as the Wasserstein distance between successive mass densities, in order to discretize the action integral in time. The optimal-transportation approach to time discretization leads to geometrically-exact updates of the local volumes and mass densities, and exact conservation properties including symplecticity, linear and angular momentum.

Many of the applications where Lagrangian meshfree schemes, such as the OTM method, are attractive involve material failure and fracture of some kind. However, there is limited experience at present concerning the simulation of fracture and fragmentation processes within the framework of meshfree interpolation schemes. Notable exceptions are the contributions in the meshfree Galerkin approximation (Belytschko et al. 1993; Lu et al. 1995; Belytschko et al. 1996), and in the smooth particle hydrodynamics method (Rabczuk and Eibl 2003; Rabczuk et al. 2004; Karekal et al. 2011). In this paper we assess the performance of a recently proposed approach to fracture, termed *eigenfracture* (Schmidt et al. 2009), within a meshfree framework. The eigenfracture scheme resorts to the classical device of *eigendeformations* (Mura 1987; Colonnetti 1917) in order to account for material fracture. To this end, the energy functional depends on two fields: the displacement field u and an eigendeformation field ε^* that describes such cracks as may be present in the body. Specifically, eigendeformations allow the displacement field to develop jumps at no cost in local elastic energy. In addition, in the eigenfracture scheme the fracture energy is set to be proportional to the volume of the ϵ-neighborhood of the support of the eigendeformation field, suitably scaled by $1/\epsilon$. The optimal crack set is obtained by minimizing the resulting energy functional with respect to both the displacement and the eigendeformation fields, subject to irreversibility constraints. We note that other two-field approximation schemes for brittle fracture, most notably the Ambrosio-Tortorelli scheme (Ambrosio and Tortorelli 1992; Braides and Defranceschi 1998), have been proposed in the past and used as a basis for numerical approximations (Bourdin and Chambolle 2000; Bourdin et al. 2000; Bourdin 2007), but the use of eigendeformations to describe brittle fracture in a variational framework does not appear to have been pursued prior to Schmidt et al. (2009). We also note that other damage regularizations of brittle fracture (Braides and Dal Maso 1997; Braides 2002; Braides and Defranceschi 1998; Negri 2005) have been proposed in the past and shown to be convergent.

In the present work, we specifically consider a meshfree approach based on maximum-entropy (max-ent) interpolation (Arroyo and Ortiz 2006) combined with material-point sampling and integration (Li et al. 2010). We confine our attention throughout to quasistatic problems. Extensions of the max-ent meshfree approach to dynamics, based on optimal-transportation theory, may be found in Li et al. (2010). In the max-ent/material-point scheme considered here, the spatial discretization is based on two sets of points: the nodal points and the material points. Specifically, the nodal points carry the information concerning the displacements, whereas the material points carry the material state, including eigendeformations. The link between material points and nodes is established through max-ent interpolation. The max-ent shape functions exhibit rapid decay and their support can be restricted to a finite range, with the result that every material point is connected to a limited number of nodes within its immediate environment.

When combined with the max-ent/material-point scheme, eigenfracture may be implemented as *material-point erosion*, i. e., the material-points can be either intact, in which case their behavior is elastic, or be completely failed—or eroded—and have no load bearing capacity. The implementation of the method, included the all-important ϵ-neighborhood construction, is exceedingly simple and applies to general

situations, possibly involving complex three-dimensional fracture patterns such as branching and fragmentation. The accuracy and convergence of the eigenerosion approach is comparable—at a much reduced implementation cost and complexity—to that of other numerical fracture schemes. We note that element erosion has been extensively used to simulate fracture in a number of areas of application, including terminal ballistics (Johnson and Stryk 1987; Belytschko and Lin 1987; Ortiz and Giannakopoulos 1990; Johnson and Stryk 1990; Whirley and Hallquist 1991; Borvik et al. 2008). However, some of these methods fail to converge or converge to the wrong limit (Negri 2003). By contrast, the eigenfracture scheme is known to properly converge to Griffith fracture (Griffith 1920) in the limit of vanishingly small mesh sizes (Schmidt et al. 2009). In particular, the local-neighborhood averaging of the energy which underlies the calculation of the effective energy-release has the effect of eliminating spurious mesh-dependencies.

We base our assessment of the method on selected verification and validation test cases. We verify the approach by means of convergence studies for mode I fracture propagation in three-dimensional plates. We additionally present a validation of the method through simulations of combined traction-torsion experiments on aluminum oxide bars (Suresh and Tschegg 1987). We find that the eigenerosion scheme indeed results in convergent approximations, both as regards crack paths as well as the attendant deformation fields and structural response. We also find that the scheme enables the simulation of exceedingly complex three-dimensional fracture patterns. The range and versatility afforded by the approach is all the more remarkable given the simplicity of its implementation.

The paper is organized as follows. In Sect. 2, we begin by formulating the problem to be approximated, namely, the problem of quasi-static crack-growth in an otherwise linear-elastic solid. We continue with brief review to the max-ent/material-point approach in Sect. 3. Then we recall briefly the eigenerosion method in Sect. 4. In Sect. 5.1 we verify the approach by means of convergence studies for mode I fracture propagation in three-dimensional plates. In Sect. 5.2 we present a validation of the method through simulations of combined traction-torsion experiments on aluminum oxide bars, taken from Suresh and Tschegg (1987). We conclude with some comments on the actual results and possible extensions in Sect. 6.

2 Variational formulation of fracture mechanics

In this section we succinctly summarize the formulation of fracture mechanics that we take as the basis for subsequent developments. We specifically follow Larsen et al. (2009) and Pandolfi and Ortiz 2012, which may be consulted for additional mathematical detail.

We consider an elastic body occupying a domain $\Omega \subset \mathbb{R}^n$, $n \geq 2$. The boundary $\partial\Omega$ of the body consists of an exterior boundary Γ, corresponding to the boundary of the uncracked body, and a collection of cracks jointly defining a crack set C. In addition, Γ is partitioned into a displacement boundary Γ_1 and a traction boundary Γ_2. The body undergoes deformations under the action of body forces, displacements prescribed over Γ_1 and tractions applied over Γ_2, subject to a contact condition on C. Under these conditions, the potential energy of the body is

$$E(u, C, t) = \begin{cases} \int_\Omega W(x, u, \nabla u)\, dx + \int_{\Gamma_2} V(x, u)\, d\mathcal{H}^{n-1}, & \text{if } [\![u]\!] \cdot \nu \geq 0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{1}$$

where, here and subsequently, $\mathcal{H}^d$ is the d-dimensional Hausdorff measure,[1] ν is a unit normal to C, $[\![u]\!]$ is the displacement jump,

$$[\![u]\!] \cdot \nu \geq 0 \tag{2}$$

defines the contact constraint, W is the elastic strain energy density of the body—possibly including distributed body forces— and V is the potential of the applied tractions. The explicit dependence of E on t in (1) is meant to reflect the time dependence of the forcing, namely, the applied forces and prescribed displacements. Suppose now that the applied loads and prescribed displacements are incremented over the time interval $[t, t + \Delta t]$ and that, in response to this incremental loading, the crack set extends from $C(t)$ to $C(t + \Delta t)$. Owing to the irreversibility of fracture we must necessarily have that

$$C(t) \subset C(t + \Delta t), \tag{3}$$

i. e., the crack set must be monotonically increasing in time. Let the elastic energy increment recorded during the time increment be ΔE. Then, a classical calculation

[1] cf., e. g., Dal Maso and Toader (2002); on smooth curves, $d\mathcal{H}^1$ is the element of length; on smooth surfaces, $d\mathcal{H}^2$ is the element of area.

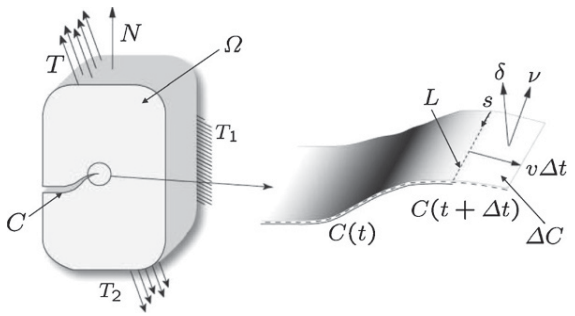

Fig. 1 Crack advancing in a body occupying a domain Ω and zoom of the crack-front region showing the crack set $C(t)$ at time t, contained in the extended crack set $C(t+\Delta t)$ at time $t+\Delta t$. During the time interval Δt the crack front L sweeps an area ΔC of unit normal ν, and propagates in the direction of the crack front velocity v

(Knees and Mielke 2008; Larsen et al. 2009) gives the rate of energy release as

$$-\dot{E} = -\lim_{\Delta t\to 0}\frac{\Delta E}{\Delta t} = \int_L G v\, d\mathcal{H}^{n-2} \tag{4}$$

where L is the crack front, Fig. 1b, v is the crack-front velocity, and

$$G = \lim_{\Delta t\to 0} \sigma\, \nu\cdot[\![u_{t+\Delta t}]\!], \quad \sigma = \partial_{\nabla u} W\,(x, u, \nabla u) \tag{5}$$

is the energetic force acting on the crack front. The identity (4) gives the rate at which energy flows to the crack front.[2] In continuum thermodynamics, the duality-pairing structure of (4) is conventionally taken to mean that the energetic force G does power, or *drives* on the crack-front velocity v. On this basis, within Osanger's general framework for inelastic processes, we may postulate the existence of a *crack-tip equation of motion* of the form

$$G = \partial\psi(v), \tag{6}$$

where ψ is a dissipation potential density per unit crack-front length. The total dissipation potential for the entire crack front finally follows by additivity as

$$\Psi(v) = \int_L \psi(v)\, d\mathcal{H}^{n-2}. \tag{7}$$

We note that the dissipation attendant to crack growth is localized to the crack front L. Under the assumption of rate-independence the dissipation potential is of the form

$$\psi(v) = G_c|v|, \tag{8}$$

which is subject to the monotonicity constraint (3). In (8), G_c is the critical energy release rate, or specific fracture energy, of the material. The assumption of rate-independence is characteristic of ideally brittle behavior and forms the basis of Griffith's theory of fracture Griffith (1920). Since the rate-independent dissipation potential $\psi(v)$, Eq. (8), is not differentiable at the origin, the equation of motion (6) must be understood in the sense of subdifferentials, namely,

$$G - G_c \le 0, \tag{9a}$$

$$v \ge 0, \tag{9b}$$

$$(G - G_c)v = 0, \tag{9c}$$

which embody Griffith's crack propagation and arrest criteria.

Because of the rate-independent nature of Griffith's criterion, the crack tracking problem can be reduced, in the spirit of the so-called *deformation theory*, to the minimization of the energy-dissipation functional Mielke and Ortiz (2007)

$$F(u, C, t) = E(u, C, t) + G_c|C| \tag{10}$$

at every time, subject to the monotonicity constraint (3), i. e.,

$$\big(u(t), C(t)\big) \in \operatorname{argmin} F(\cdot, \cdot, t), \tag{11a}$$

$$\text{subject to: } C(t_1) \subset C(t_2), \text{ whenever } t_1 < t_2. \tag{11b}$$

In (10), $|C|$ denotes the area of the crack set. Thus, the geometry of a growing crack and the corresponding equilibrium elastic field in a perfectly brittle material is obtained by jointly minimizing $F(u, C, t)$ at all times with respect to both the displacement field u and the crack set C subject to the constraint (3). In particular, the crack path results from a competition between: the elastic energy, which promotes fracture as an energy-release mechanism; the specific fracture energy, which penalizes fracture proportionally to the crack area; and the monotonicity and contact constraints, which introduce irreversibility, path dependency, hysteresis and tension-compression asymmetry.

A rigorous derivation of the deformation-theoretical formulation (11) of the crack-tracking problem may be

[2] cf. Larsen et al. (2009) for a rigorous mathematical definition of the crack front and attendant crack-front velocity.

based on energy-dissipation functionals (Mielke and Ortiz 2007), which in the present context supply a minimum principle that characterizes entire crack paths Larsen et al. (2009). Problem (11) then follows simply from by noting that, for monotonically growing brittle cracks, the dissipation (7) is an exact time-differential of $G_c|C|$, i. e.,

$$\Psi(v) = \frac{d}{dt}\left(G_c|C|\right). \tag{12}$$

Conditions for the existence of solutions of the crack-tracking problem (11), which is not guaranteed in general, may be found in Dal Maso and Toader (2002), Francfort and Larsen (2003), and Dal Maso et al. (2005).

In order to obviate the need for minimizing the energy-dissipation functional $F(u, C, t)$ with respect to the crack set C, which may be numerically cumbersome, in the framework of linearized elasticity Schmidt et al. (2009) have proposed a reformulation of the crack-tracking problem (10) in terms of *eigendeformations* ε^*, with the aid of a small parameter ϵ with units of length. In particular, the crack set is approximated as $C = \{\varepsilon^* \neq 0\}$, namely the support of the eigendeformation field, i. e., the domain over which the eigendeformation field is nonzero; and C_ϵ is the *ϵ-neighborhood* of C, i. e., the set of points that are at a distance less or equal to ϵ from C. The regularized energy-dissipation functional proposed by Schmidt et al. (2009) is

$$F_\epsilon(u, \varepsilon^*, t) = \int_\Omega W(\varepsilon(u) - \varepsilon^*)\, dV + \int_{\Gamma_2} V(x, u)\, d\mathcal{H}^{n-1} + G_c \frac{|C_\epsilon|}{2\epsilon}. \tag{13}$$

In this expression $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$ is the the strain operator of linear elasticity and $|C_\epsilon|$ denotes the volume of the ϵ-neighborhood C_ϵ. We note that the regularized energy-dissipation functional $F_\epsilon(u, \varepsilon^*, t)$ now allows for eigendeformation fields that are spread over a volume, and thus represent a damaged volume of material. As before, the eigendeformations allow the material to relax its energy locally. The centerpiece of the approach of Schmidt et al. (2009) concerns the proper evaluation of the fracture-energy cost attendant to a distribution of eigendeformations, which is given by the last term in (13). Indeed, Schmidt et al. (2009) have shown that the regularized energy-dissipation functional F_ϵ, Eq. (13), Γ-converges to the Griffith functional F, Eq. (10), as $\epsilon \to 0$. We recall that Γ-convergence is a notion of variational convergence that implies convergence of minimizers. In particular, the scaling of the volume of the ϵ-neighborhood C_ϵ by ϵ^{-1} in the energy-dissipation functional (13) in turn penalizes the volume of the approximate crack set C, which in the limit converges to a surface.

The regularized crack tracking problem consists of minimizing the regularized energy-dissipation functional (13) for every time, subject to the monotonicity constraint (3), which now requires that the approximate crack sets grow monotonically, and to a suitable contact constraint in lieu of (2). For instance, the constraint

$$\varepsilon^* \geq 0, \tag{14}$$

first proposed by Ortiz (1985) and widely used since, can be used to enforce the contact constraint within an eigendeformation framework. Constraint (14) specifically requires that all the eigenvalues of ε^* be nonnegative, which effectively satisfy the crack closure constraint.

It bears emphasis that the net effect of the regularization of the fracture energy in (13) is to eliminate the spurious mesh-dependencies that afflict naive erosion schemes and ensure convergence of the approximations. For instance, a typical scheme consists of introducing C^0 finite-element interpolation for the displacements and piecewise constant interpolation for the eigendeformations, i. e., restricting the eigendeformations to be constant over the elements, see, e. g., Ortiz and Giannakopoulos (1990). These schemes indeed converge pointwise as the mesh size goes to zero, provided that the crack set is aligned with the mesh, as in the case of a structured mesh in a rectangular plate subjected to mode I loading, but may fail to converge otherwise Negri (2003). The reason for the lack of convergence is a geometrical one: as the crack zig-zags in accordance with the mesh in order to match the limiting crack path, it overestimates the amount of fracture energy by a geometrical factor. Negri developed converging schemes that overcome this difficulty by recourse to mesh adaption (Negri 2003, 2005) or to nonlocal averaging schemes (Negri 2005; Lussardi and Negri 2007).

3 Max-ent shape functions

The remainder of the paper is devoted to a verification and validation analysis of mesh-free approximation schemes based on the regularized crack-tracking problem just described. The spatial discretization of the energy-dissipation functional (13) considered in this work is the quasistatic version of the optimal-transportation meshfree (OTM) method developed in Li et al. (2010) for particular applications to flow of fluids and solids. The discretization is based on two sets of points: the nodal points and the material points. Thus, the nodal points x_a carry the information concerning the displacements, whereas the material points x_p carry the material state, including eigendeformations. The link between material points and nodes is established through the displacement interpolation rule:

$$u_p = \sum_{a=1}^{n} u_a N_a\left(x_p\right), \tag{15}$$

where $N_a(x)$ are conforming shape functions defined over Ω. The support of the shape functions $N_a(x)$ is expected to have a finite range, so that every material point is connected to a limited number of nodes within its immediate environment. Following Li et al. (2010), in this work we specifically use the local max-ent interpolation introduced in Arroyo and Ortiz (2006). Thus, the shape function for the node a is

$$\begin{aligned} N_a(x) &= \frac{1}{Z(x, \lambda^*(x))} \\ &\quad \times \exp\left[-\beta|x - x_a|^2 + \lambda^*(x) \cdot (x - x_a)\right], \\ a &= 1, \ldots, n, \end{aligned} \tag{16}$$

where

$$\lambda^*(x) = \arg\min_{\lambda \in \mathbb{R}^d} \log Z(x, \lambda). \tag{17}$$

The function $Z : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the partition function associated with the node set X, i. e.:

$$Z(x, \lambda) \equiv \sum_{a=1}^{n} \exp\left[-\beta|x - x_a|^2 + \lambda \cdot (x - x_a)\right]. \tag{18}$$

The scalar β can be chosen to be dependent on the position, e. g., related to the size of the material-point neighborhood h_p as

$$\beta_p = \gamma \frac{1}{h_p^2}, \tag{19}$$

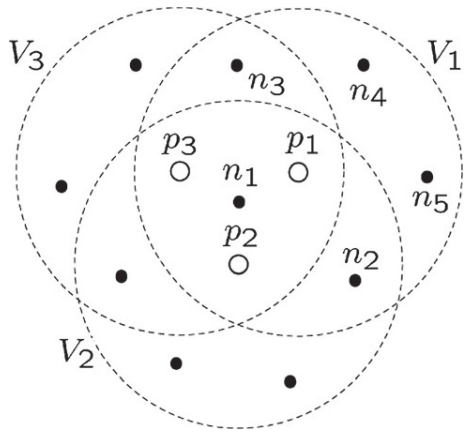

Fig. 2 Sketch of the neighborhoods of three material points, p_1, p_2 and p_3, labeled V_1, V_2 and V_3. The nodes are denoted n_i. Material point p_1 has connections to five nodes lying in V_1. Node n_1 lies inside the three neighborhoods V_k and, therefore, is connected to the three material points

where γ is a dimensionless constant that assumes positive values close to the unity. In the present applications we assume $\gamma = 0.1$. The neighborhood size h_p measures the radius of a spherical volume V_p, centered at the material point and including n_p nodal points, where the max-ent shape functions of the material point are defined. The size h_p—and therefore the number of support nodes—is not necessarily constant for all the material points. For example, in the present implementation h_p is chosen to scale with the local distance between material points. Fig. 2 illustrates the concept of h_p-neighborhood in two dimensions.

The discretization of the regularized energy-dissipation potential (13) requires the use of the shape-function derivatives, which can be found in Arroyo and Ortiz (2006). The computation of the nodal forces due to the tractions on Γ_2 may be simplified by introducing a balanced stress field τ satisfying the identities:

$$\nabla \cdot \tau = 0 \ \text{ in } \Omega, \quad \tau n = q \ \text{ on } \Gamma_2, \tag{20}$$

see Li et al. (2010), and applying the divergence theorem. By imposing the stationarity of (13), we obtain the standard nonlinear equilibrium equations. In static applications involving nonlinear material behaviors and kinematics a consistent linearization of the internal forces may be necessary. In the calculations presented here we solve the nonlinear equilibrium equations by means of an explicit dynamic relaxation algorithm Oakley and Knight (1995), which does not require computation of the tangent stiffness.

4 The eigenerosion criterion

Schmidt et al. (2009) have proved that approximations of the regularized problem converge to exact solutions of the Griffith crack-tracking problem when the mesh size and the regularization parameter ϵ tend to zero in the right order. One specific scheme that was shown to be convergent consists of approximating the displacement field by means of conventional C^0 finite-element interpolations and taking the eigendeformation field to be constant over elements but otherwise unconstrained, by using a local regularization Schmidt et al. (2009). Since the local element eigendeformations are allowed to take arbitrary values, they either are zero, in order to minimize the attendant fracture energy, or completely negate the local deformation of the element, thus rendering its elastic energy zero. Thus, in that finite element approximation scheme, the discrete crack-tracking problem is reduced to successively *failing* or *eroding* elements when the attendant elastic energy release exceeds the attendant cost in fracture energy.

A detailed algorithm for finite element erosion has been provided in Pandolfi and Ortiz (2012). The same algorithm applies to the present material-point discretization, with elements replaced by material points *mutatis mutandis*. In particular, the energy release rate attendant to the erosion of one material point can be computed explicitly, as the difference of the energies of the body before and after the erosion of the material point, or, more conveniently, it can be approximated using first-order asymptotic formulae for notches Ortiz and Giannakopoulos (1990). In addition, the fracture-energy cost is computed by recourse to the ϵ-neighborhood construction, where a small parameter ϵ with the dimension of a length is used to define a volume-like neighborhood that approximates the crack surface Pandolfi and Ortiz (2012). Finally, the contact constraint is imposed by restricting erosion to material points in a state of volumetric expansion, i. e., material points whose volume in the deformed configuration is larger than the undeformed volume.

The simulation of crack propagation requires the sequential solution of equilibrium step and material point-erosion steps. If material points are eroded, the equilibrium step needs to be repeated under the same boundary conditions in order to restore mechanical equilibrium. However, in some cases unstable crack growth may result in several material-point erosion steps at fixed external load, or on runaway material-point erosion if the problem has no solution, in the sense of existence of joint minimizers (u, C) of the energy (10). Evidently, the intermediate configurations resulting from multiple material-point erosions at fixed applied load are the result of ancillary constraints on the growth of the crack, namely, an ordering of the material points by energy-release rate and the sequential failing of material points according to that ordering. A rigorous mathematical framework for appending such ancillary quasistatic crack-growth constraints has been put forth by Larsen et al. (2010b) and Larsen (2010). An alternative regularization of the problem consists of replacing the rate-independent kinetics (12) characteristic of Griffith fracture by kinetics defined by dissipation potentials Eq. (12), with super-linear growth. This type of crack-growth kinetics has been investigated in Larsen et al. (2009) by means of energy-dissipation functionals Mielke and Ortiz (2007). Yet another—physically based— regularization of the problem consists of accounting for inertia and dynamic crack growth Bourdin et al. (2011); Li et al. (2012).

The implementation of the eigenerosion scheme into the meshfree material-point code is particularly simple. In addition to the standard arrays required by any static solver, the eigenerosion scheme requires the introduction of an array to keep track of the material points that are progressively included in the evolving ϵ-neighborhood of the crack. We recall that, according to the Γ-convergence analysis of Schmidt et al. (2009), the size ϵ of the neighborhood can be chosen freely, as long as it tends to zero more slowly than the mesh size. In the calculations presented here we simply take, based on calibration studies, $\epsilon = 2.5h_{\min}$, where $h_{\min}$ is the minimum distance between nodes.

5 Numerical examples

In this section we collect verification tests aimed at assessing the convergence characteristics of the eigenerosion scheme. We additionally present a validation example concerned with the simulation of the mixed mode I-III tests on aluminum oxide bars presented in Suresh and Tschegg (1987). This validation example showcases the ability of the eigenerosion scheme to simulate complex three-dimensional crack geometries.

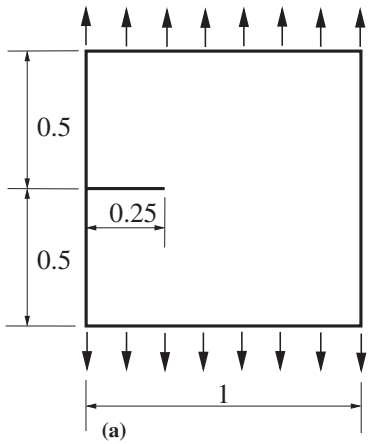

Fig. 3 Displacement boundary conditions for the square plate loaded in mode I. Edge $H = 1$, precrack $a = 0.25$, thickness $t = h_{\min}$

5.1 Edge-cracked square panel in mode I

We begin by assessing the performance of the eigenerosion scheme by means of standard numerical tests concerned with plane-strain crack growth in mode I. For purposes of comparison, we replicate the dimensionless conventions used in Pandolfi and Ortiz (2012). Specifically, we consider a square plate of size $H = 1$ containing an initial edge crack of length $a = 0.25H$ loaded in pure mode I by displacement control on the outer flanks of the plate, Fig. 3. The calculations are carried out in finite deformations for a compressible neo-Hookean material with Young's modulus $E = 1.06$, Poisson's ratio $\nu = 0.333$ and critical energy-release rate $G_c = 0.0001$. Numerical simulations are performed using a full three-dimensional code. In order to simulate two-dimensional plane-strain geometries, the boundary displacements in the third direction are constrained. In addition, we place two material points across the thickness, which is taken equal to the mesh size.

We consider three discretizations defined by the minimum distance between nodes $h_{\min}$, and labeled M1, M2 and M3, see Table 1. The distance $h_{\min}$ is taken as a basis for the definition of the ϵ-neighborhood size, leading from 3 to 4 material points in each material point ϵ-neighborhood.

The cracks for the three discretizations at the final –or at an advanced– stage of the fracture propagation are compared in Fig. 4. The crack surfaces are generated by the erosion of a single layer of material points and appear smooth and straight. The effective crack, modeled by the set of eroded material points, converges to a flat surface for $\epsilon \to 0$ following $h_{\min} \to 0$. Contour levels in Fig. 4 refer to the normal component of the Cauchy stress in the vertical direction.

Table 1 Data of the discretizations considered for the verification analysis through convergence tests in the squared prenotched plate

Discretization	Material points	Nodes	$h_{\min}$
M1	9, 600	3, 382	0.0029
M2	38, 400	13, 162	0.0014
M3	153, 600	51, 922	0.0007

Plots of global quantities per unit of thickness versus the prescribed boundary displacement demonstrating the expected mesh independency are shown in Fig. 5. In particular, Fig. 5a illustrates the global vertical reaction and Fig. 5b describes the total displacement norm $|u|_{L1}$, defined as:

$$|u|_{L1} = \int_{\Omega} |u(x)| d\Omega. \tag{21}$$

Finally, Fig. 5c shows the dependence of the strain energy E of the body on crack extension Δa. Under a prescribed boundary displacement δ the initial crack does not propagate as long as the strain energy E is less than the expenditure of fracture energy $G_c l$ necessary to break the initial ligament of length $l = 0.75$. In such situations, a null crack length $\Delta_a = 0$ is the minimizer of the functional (13). When δ equals the critical prescribed boundary displacement δ_c at which $E = G_c l$, the crack extends through the entire initial ligament with a *discrete jump* of length $\Delta a = l$, showing a sudden change of stability. At this point, the minimizer of the functional (13) is $\Delta_a = l$. We note that intermediate crack extensions $0 < \Delta a < l$ are not energy-dissipation minimizers for any prescribed boundary displacement δ and, therefore, are devoid of special meaning within the variational framework Larsen (2010).

5.2 Mixed tension-torsion experiment simulations

As a selected example of application showcasing the range and scope of the eigenerosion scheme, we proceed to simulate the combined tension-torsion

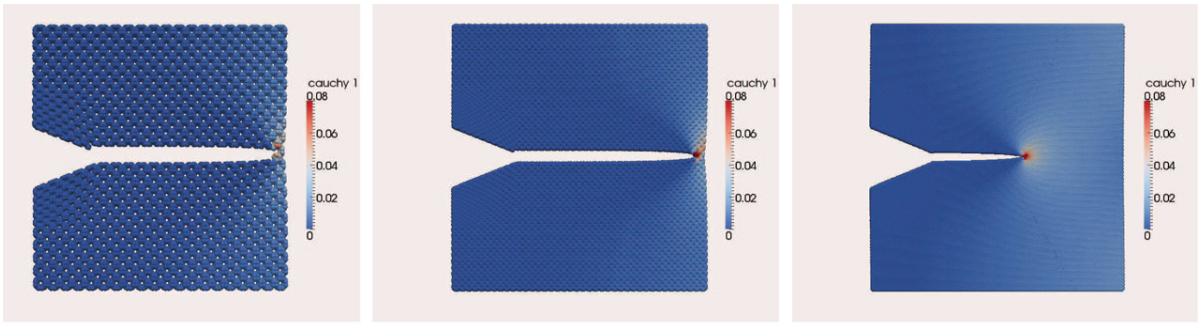

Fig. 4 Edge-crack square panel. Predicted crack paths for three meshes of increasing fineness. Contour levels refer to the normal component of the Cauchy stress in the vertical direction. **a** M1, **b** M2, **c** M3

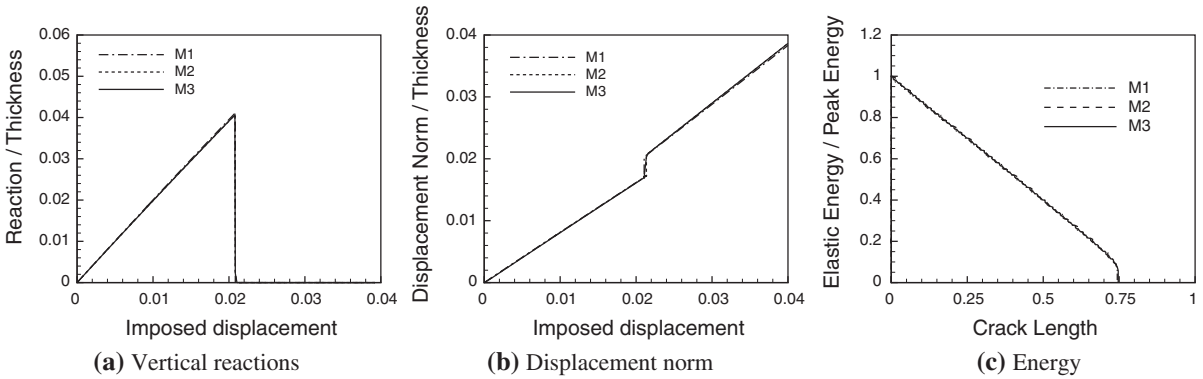

Fig. 5 Edge-cracked square panel. Dependence of: **a** global reaction force, and **b** displacement norm, on prescribed boundary displacement. **c** Dependence of the strain energy on crack length at fixed prescribed boundary displacement

experiments on aluminum oxide bar specimens with a circular notch reported by Suresh and Tschegg (1987). Fig. 6 shows the geometry of the specimen used in the experiments, with the detail of the annular pre-crack smooth notch. In the experiments the pre-crack has been sharpened by fatigue through a mode I cyclic loading in tension. To limit the computational effort, in our numerical discretization we model the central part of the specimen only, while describing accurately the notch and the fatigue crack, and apply displacement boundary conditions consistent with the experimental tension-torsion configuration, see Fig. 6. Specifically, we fix one end of the computational domain and apply a uniform axial displacement and rigid rotation about the axis at the other end. The resultants of the nodal reactions on the fixed base provide the numerical axial force and torque to be compared with the experimental data. Displacements are increased monotonically up to the onset of crack growth from notch. Thereafter, the displacements are kept constant, as in all cases the cracks grow unstably.

The experimental paper provided the material properties used in the numerical analyses, i. e., the elastic modulus $E = 345\,\text{GPa}$ and the mode I toughness $K_{Ic} = 3.35\,\text{MPam}^{1/2}$. Assuming a Poisson's coefficient $\nu = 0.3$, we derive the shear modulus $\mu = 138\,\text{GPa}$ for the neo-Hookean model. The critical energy release rate is computed through the relation

$$G_c = \frac{K_{Ic}^2}{E} \tag{22}$$

which gives $G_c = 32\,\text{N/m}$.

The three-dimensional numerical model comprises 603,996 material points and 221,191 nodes. The spatial distribution of the material points is not uniform and the discretization coarsens away from the notch,

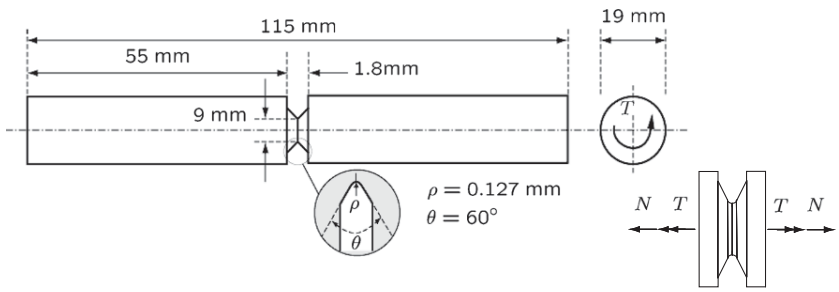

Fig. 6 Geometry of the specimen used in the experimental tests of Suresh and Tschegg (1987). The computational domain and loading used for the simulations is inset in the lower-right corner

see Fig. 7. The minimum discretization size at the notch is $h_{\min} = 0.0032$ mm. In all calculations, the max-ent interpolation parameter is set to $\gamma = 0.1$. By virtue of this choice, the range of the max-ent shape functions extends beyond nearest-neighboring nodes and, therefore, the attendant interpolation departs significantly from finite-element interpolation. An initial triangulation Fig. 7a is use in order to define the material-point set Fig. 7b and assign volumes to each material point. In order to calibrate the eigenerosion ϵ-neighborhood size we begin by performing preliminary numerical analyses in pure tension. The experimentally observed maximum axial force at failure is approximately 3,750 N Suresh and Tschegg (1987). The minimum discrepancy between the numerical and experimental axial force is found to be 5 % and to occur for $\epsilon = 2.5h_{\min}$. This value of ϵ is then used in all the subsequent calculations.

We consider three loading cases leading to different stress states at the notch: (i) pure tension or mode I loading; (ii) mixed tension—torsion or mixed-mode I–III loading, and (iii) pure torsion or mode III loading. In all cases, failure occurs catastrophically once the limit load is reached, in agreement with experiment. Suresh and Tschegg (1987) provide the pictures of the post-mortem crack surfaces for different specimens, included the tension—pure mode I—and torsion—pure mode III—cases. In the pure-tension case, the fracture surface appears flat and smooth. In all other cases, the fracture surfaces exhibit a striking saw-toothed pattern forming inclined surfaces akin to *petals*. This orderly *petal structure* breaks down in the central part of the specimen. The sharpness of the petal structure, as well as the failure load, increase with the magnitude of applied torque.

Figures 8, 9, and 10 show the aspect of the numerically computed crack surfaces for the three simulations, at one of the last stages of the fracture process. On the left side of each figure, the fracture surface is visualized by the collection of the eroded material points. In this representation, the eroded material points are located in proximity of—and *trace*—the crack surface. By contrast, on the right side of the figure we show the structure of the crack surface as obtained by means of a rendering procedure based on averaging and smoothing the positions of the eroded material points and on spline approximation (cf. Alliez et al. 2007; Mullen et al. 2010). This *postprocessing construction* aims to determine a smooth surface, which may then be identified with the crack set, that is as close as possible to the failed material-point set while costing the same amount of fracture energy. Thus, whereas eigenerosion supplies a volume approximation of the crack set, the postprocessing construction effectively reverses that approximation and provides a sharp surface representation of the crack set. A projection of the crack surfaces is additionally shown at the bottom of the figures in order to visualize the complex crack front.

It may be seen from the figures that the pure-traction case is predicted to produce a flat crack surface, whereas the pure-torsion case is predicted to result in a periodic petal structure that breaks down in the central part, in agreement with experiment. In addition, the mixed-mode case generates less sharp petals that converge smoothly towards the center of the specimen, also in agreement with experiment. The precise fractography predicted by the calculations cannot be compared quantitatively with experiment, since the original publication does not report quantitative fractographic measurements. This limitation notwithstanding, a qualitative comparison can be performed in terms of the observed number of petals, to wit: eight petals in pure torsion and seven larger petals in the mixed-mode case. Remarkably, the calculations predict exactly the same number of petals in both cases, cf. Figs. 9 and 10.

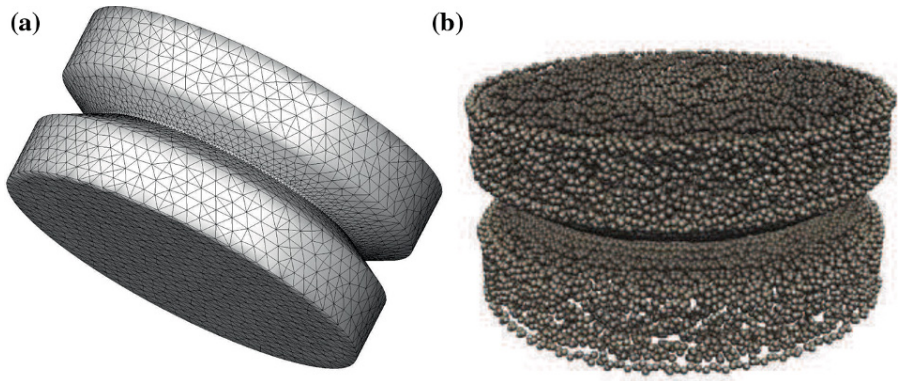

Fig. 7 Mixed-Mode tests of Suresh and Tschegg (1987). Computational model consisting of 603,996 material points and 221,191 nodes. Note the strong refinement at the notch. **a** Initial triangulation of the nodal set used to define the material-point set. **b** Visualization of the resulting material-point set

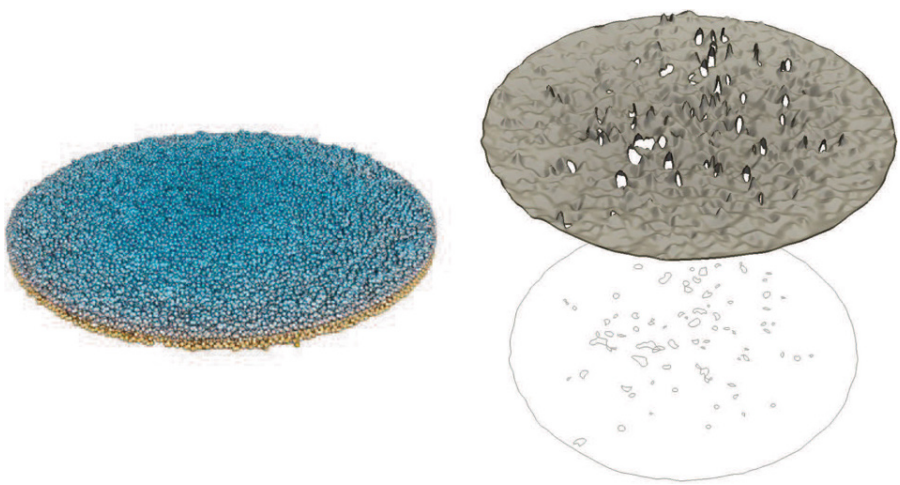

Fig. 8 Simulation of the mixed-mode tests of Suresh and Tschegg (1987), pure-tension case. *Left*: failed material-point set. *Right*: reconstructed crack surface

Further insight into the formation of complex fractographies in the pure torsion and mixed-mode experiments and simulations may be derived from stability analyses. Thus, in keeping with observation, our simulations describe the segmentation of the initially flat annular crack into daughter cracks that rotate progressively towards the direction of maximum tensile stress. Subsequently, the axial symmetry of the test forces the cracks to merge in the central part of the specimen. As observed experimentally in other geometries, mode I-III cracks often propagate unstably by the formation of inclined facets, steps and self-similar branching patterns. Stability analyses of mode I-III crack growth that shed light into such observations may be found, e. g., in (Xu et al. 1994; Movchan 1998; Lazarus et al. 2001; Lin et al. 2010; Leblond et al. 2011). These analyses reveal the existence of a critical stress intensity factor ratio K_{III}/K_I, dependent on Poisson's ratio, that separates stable and unstable planar crack growth (Xu et al. 1994; Movchan 1998; Leblond et al. 2011). Thus, for sufficiently small, respectively large, K_{III}/K_I ratio planar crack growth is stable, respectively unstable. The stress-intensity ratio corresponding to the mixed-mode test under consideration here is $K_{III}/K_I \approx 3$, which is greatly in excess of the stability limit for $\nu = 0.333$. Under these conditions, planar crack growth is unstable and the crack may indeed be expected to grow out of the plane and form complex patterns, as predicted by our calculations.

6 Summary and concluding remarks

We have described a meshfree material-point approximation method for the numerical simulation of brittle fracture propagation. The approach is based on the combination of: meshfree max-ent interpolation (Arroyo and Ortiz 2006); material-point sampling and integration (Li et al. 2010); and a convergent material-point erosion method based on the concept of eigendeformations (Schmidt et al. 2009; Pandolfi and Ortiz 2012). Specifically, the crack set is approximated by

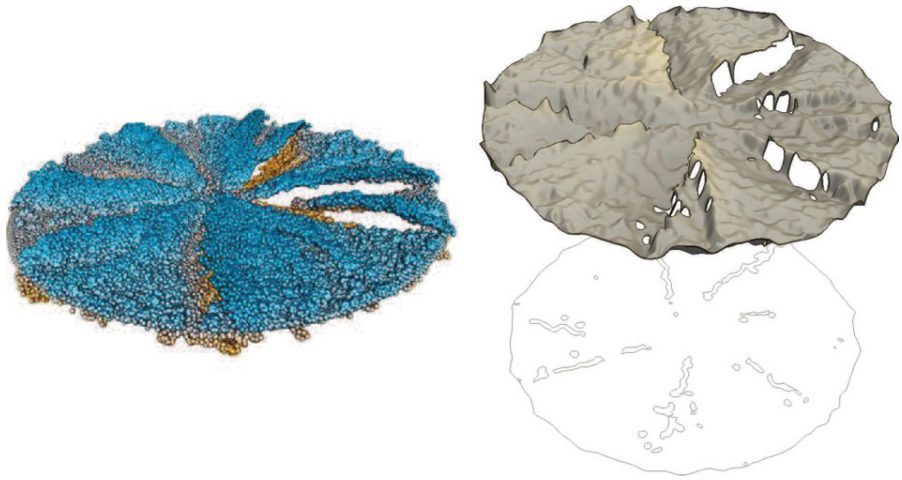

Fig. 9 Simulation of the mixed-mode tests of Suresh and Tschegg (1987), mixed tension-torsion case. *Left*: failed material-point set. *Right*: reconstructed crack surface

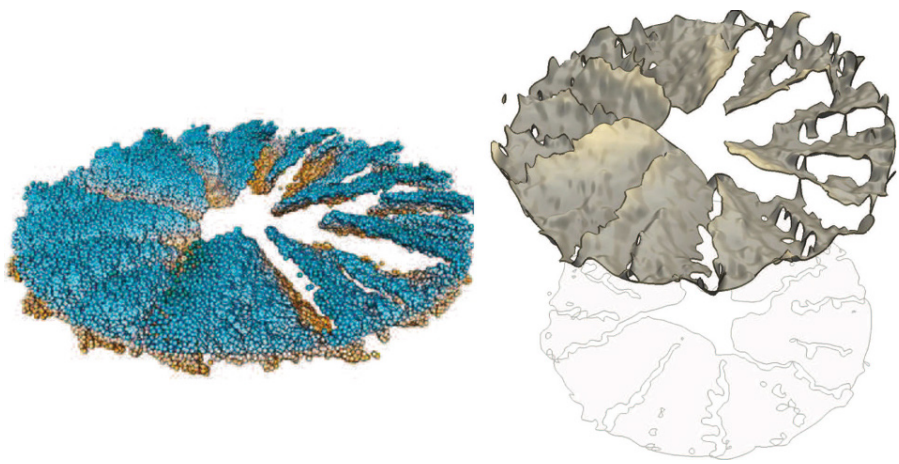

Fig. 10 Simulation of mixed-mode tests of Suresh and Tschegg (1987), pure-torsion case. *Left*: failed material-point set. *Right*: reconstructed crack surface

means of eigendeformations, which enable the material to develop displacement jumps at no cost of local elastic energy. In the implementation developed in this work, which we term *eigenerosion*, we compute the energy-release rate attendant to the failure of one material point by means of an ϵ-neighborhood construction. This construction averages the elastic energy over a length scale ϵ intermediate between the mesh size and the size of the body. In this manner, the details of the mesh are averaged over and the scheme results in mesh-insensitive—and ultimately convergent—crack paths and fracture energies. The overall convergence of the method is clearly apparent in benchmark tests such as crack initiation and growth in an edge-crack panel. The range and scope of the method has been demonstrated through the simulation of the combined torsion-traction experiments of Suresh and Tschegg (1987). The ability of eigenerosion to predict the salient features of the complex crack patterns that arise in those experiments is remarkable.

As already noted, the algorithm presented here is capable of tracking the propagation of both stable and unstable cracks. For unstable cracks, by ordering the material points by energy-release rate and failing them sequentially in accordance with that ordering, the algorithm provides crack-growth paths joining two consecutive stable crack configurations or representing runaway unstable crack growth. There exist at present rigorous mathematical approaches for understanding possibly-unstable quasistatic crack-growth in brittle solids (Larsen et al. 2009, 2010b; Larsen 2010). Whereas the present approach appears to be predictive on the basis of direct comparisons with experiment, a firm mathematical grounding of numerical crack-tracking algorithms remains to be established and is greatly to be desired.

In closing, we additionally remark that the present approach may be made extensive to inelastic behavior, including plasticity, by recourse to variational constitutive updates (Ortiz and Stainier 1999; Yang et al. 2006). Thus, variational updates provide well-defined incremental energies that combine both internal energy and dissipation. The incremental energies can in turn be used in order to define the driving force for fracture in the presence of inelasticity (Li et al. 2012).

Acknowledgments The renderings of crack surfaces presented in this paper were performed by Dr. Santiago Lombeyda and are gratefully acknowledged. The authors gratefully acknowledge the support of the Department of Energy National Nuclear Security Administration under Award Number DE-FC52-08NA28613 through Caltech's ASC/PSAAP Center for the Predictive Modeling and Simulation of High Energy Density Dynamic Response of Materials.

References

Alliez P, Cohen-Steiner D, Tong Y, Desbrun M (2007) Voronoi-based variational reconstruction of unoriented point sets. In: Belyaev A, Garland M, (eds) Eurographics symposium on geometry processing

Ambrosio L, Tortorelli VM (1992) On the approximation of free discontinuity problems. Boll Un Mat Ital B 7:105–123

Arroyo M, Ortiz M (2006) Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods. Int J Numer Methods Eng 65:2167–2202

Belytschko T, Krongauz T, Organ D, Fleming M, Krysl P (1996) Meshless methods: an overview and recent developments. Comput Methods Appl Mech Eng 139(3):3–47

Belytschko T, Lin J (1987) A three-dimensional impact-penetration algorithm with erosion. Comput Struct 25:95–104

Belytschko T, Lu YY, Gu L (1993) Crack propagation by element free galerkin methods. In: Advanced computational methods for material modeling, AMD-ASME, vol 180. p 191

Borvik T, Hopperstad OS, Pedersen KO (2008) Quasi-brittle fracture during structural impact of AA7075-T651 aluminum plates. Int J Impact Eng 37:537–551

Bourdin B (2007) Numerical implementation of the variational formulation of brittle fracture. In: IUTAM symposium on discretisation methods for evolving discontinuities. IUTAM Bookseries, Lyon, pp 381–393

Bourdin B, Chambolle A (2000) Implementation of an adaptive finite element approximation of the Mumford-Shah functional. Numer Math 85:608–646

Bourdin B, Francfort GA, Marigo JJ (2000) Numerical experiments in revisited brittle fracture. J Mech Phys Solids 48:797–826

Bourdin B, Larsen CJ, Richardson CL (2011) A time-discrete model for dynamic fracture based on crack regularization. Int J Fract 168(2):133–143

Braides A (2002) Γ-convergence for beginners. Oxford University Press, Oxford

Braides A, Dal Maso G (1997) Nonlocal approximation of the mumford-shah functional. Calc Var Partial Differ Equ 5:293–322

Braides A, Defranceschi A (1998) Homogenization of multiple integrals. Claredon Press, Oxford

Colonnetti G (1917) Su certi stati di coazione elastica che non dipendono da azioni esterne. Rendiconti dell'Accademia dei Lincei, Giugno, pp 43–47

Dal Maso G, Francfort GA, Toader R (2005) Quasistatic crack growth in nonlinear elasticity. Arch Ration Mech Anal 176(2):165–225

Dal Maso G, Toader R (2002) A model for the quasi-static growth of brittle fractures: Existence and approximation results. Arch Ration Mech Anal 162(2):101–135

Francfort GA, Larsen CJ (2003) Existence and convergence for quasi-static evolution in brittle fracture. Commun Pure Appl Math 56:1465–1500

Griffith AA (1920) The phenomena of rupture and flow in solids. Philos Trans R Soc Lond 18:16–98

Johnson GR, Stryk RA (1987) Eroding interface and improved tetrahedral element algorithms for high velocity impacts in three dimensions. Int J Impact Eng 5:414–427

Johnson GR, Stryk RA (1990) User instruction for the 1990 version of the combined (1D, 2D, 3D) EPIC code. Technical report DE-AC04-87AL-42550, DARPA

Karekal S, Das R, Mosse L, Cleary PW (2011) Application of a mesh-free continuum method for simulation of rock caving processes. Int J Rock Mech Min Sci 5:339–361

Knees D, Mielke A (2008) Energy release rate for cracks in finite-strain elasticity. Adv Math Sci Appl 31:501–528

Larsen CJ (2010) Epsilon-stable quasi-static brittle fracture evolution. Commun Pure Appl Math 63(5):630–654

Larsen CJ (2010) Epsilon-stable quasi-static brittle fracture evolution. Commun Pure Appl Math 63:630–654

Larsen CJ, Ortiz M, Richardson CL (2009) Fracture paths from front kinetics: Relaxation and rate-independence. Arch Ration Mech Anal 193:539–583

Larsen CJ, Ortner C, Suli E (2010b) Existence of solutions to a regularized model of dynamic fracture. Math Models Methods Appl Sci 20(7):1021–1048

Lazarus V, Leblond J-B, Mouchrif S-E (2001) Crack front rotation and segmentation in mixed mode I+III or I+II+III. Part II: Comparison with experiments. J Mech Phys Solids 49:1421–1443

Leblond J-B, Karma A, Lazarus V (2011) Theoretical analysis of crack front instability in mode I+III. J Mech Phys Solids 59:1872–1887

Li B, Habbal F, Ortiz M (2010) Optimal transportation meshfree approximations schemes for fluid and plastic flows. Int J Numer Methods Eng 83(12):1541–1579

Li B, Kadane A, Ravichandran G, Ortiz M (2012) Verification and validation of the optimal-transportation meshfree (OTM) simulation of terminal ballistics. Int J Impact Eng 42:25–36

Lin B, Mear ME, Ravi-Chandar K (2010) Criterion for initiation of cracks under mixed-mode I+III loading. Int J Fract 165:175–188

Lu YY, Belytschko T, Tabbara M (1995) Element-free Galerkin method for wave propagation and dynamic fracture. Comput Methods Appl Mech Eng 126:131–153

Lussardi L, Negri M (2007) Convergence of non-local finite element energies for fracture mechanics. Numer Funct Anal Optim 28:83–109

Mielke A, Ortiz M (2007) A class of minimum principles for characterizing the trajectories and the relaxation of dissipative systems. ESAIM COCV, (in press)

Movchan AB, Gao H, Willis JR (1998) On perturbations of plane cracks. Int J Solids Struct 35:3419–3453

Mullen P, de Goes F, Desbrun M, Cohen-Steiner D, Alliez P (2010) Signing the unsigned: Robust surface reconstruction from raw pointsets. In: Sorkine O, Lévy B (eds) Eurographics symposium on geometry processing 2010

Mura T (1987) Micromechanics of defects in solids. Kluwer, Dordrecht

Negri M (2003) Finite element approximation of the Griffith's model in fracture mechanics. Numer Math 95(4):653–687

Negri M (2005) A discontinuous finite element approximation of free discontinuity probems. Adv Math Sci Appl 15: 283–306

Negri M (2005) A non-local approximation of free discontinuity problems in SBV and SBD. Calc Var Partial Differ Equ 25:33–62

Oakley DR, Knight NFJr. (1995) Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures. Comput Methods Appl Mech Eng 126:67–89

Ortiz M (1985) A constitutive theory for the inelastic behavior of concrete. Mech Mater 4:67–93

Ortiz M, Giannakopoulos AE (1990) Crack propagation in monolithic ceramics under mixed mode loading. Int J Fract 44:233–258

Ortiz M, Stainier L (1999) The variational formulation of viscoplastic constitutive updates. Comput Methods Appl Mech Eng 171: 419–444

Pandolfi A, Ortiz M (2012) An eigenerosion approach to brittle fracture. Int J Numer Methods Eng Online First: 1–21

Rabczuk T, Eibl J (2003) Simulation of high velocity concrete fragmentation using SPH/MLSPH. Int J Numer Methods Eng 56:1421–1444

Rabczuk T, Eibl J, Stempniewski L (2004) Numerical analysis of high speed concrete fragmentation using a meshfree lagrangian method. Eng Fract Mech 71:547–556

Schmidt B, Fraternali F, Ortiz M (2009) Eigenfracture: an eigendeformation approach to variational fracture. SIAM J Multiscale Model Simul 7:1237–1266

Sulsky D, Chen Z, Schreyer HL (1994) A particle method for history–dependent materials. Comput Methods Appl Mech Eng 118:179–196

Suresh S, Tschegg EK (1987) Combined Mode I-Mode I11 fracture of fatigue-precracked Alumina. J Am Ceram Soc 70:726–733

Whirley RG, Hallquist JO (1991) DYNA-3D User manual. Technical Report Technical Report UCRL-MA-107254, Lawrence Livermore National Laboratory

Xu G, Bower AF, Ortiz M (1994) An analysis of non-planar crack growth under mixed mode loading. Int J Solids Struct 31:2167–2193

Yang Q, Stainier L, Ortiz M (2006) A variational formulation of the coupled thermo-mechanical boundary-value problem for general dissipative solids. J Mech Phys Solids 54(2):401–424

DOI 10.1007/978-3-319-04397-5_3
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-012-9795-y

ORIGINAL PAPER

Crack front perturbations revisited

J. R. Willis

Received: 5 October 2012 / Accepted: 27 November 2012 / Published online: 11 December 2012

Abstract The problem of the in-plane dynamic perturbation of a crack propagating with a front that is nominally straight is solved, to second order in the perturbation. The method of approach is a streamlined and generalized version of that previously applied to first order by the author and co-workers. It emerges, however, that the analysis at second order requires for its consistency the introduction of a new singular term, of a type not present at first order. The analysis is restricted to the case of Mode I loading, for clarity of exposition. It is carried out at a level of generality that incorporates viscoelastic response as well as propagation in a "vertically stratified" medium including, as a special case, propagation in a slab of finite thickness. For illustration, the general solution is specialized to the case of a stationary crack in an infinite elastic medium and agreement with a solution recently developed by methodology that is specific to the static case is confirmed.

Keywords Crack dynamics · Crack front perturbation · Viscoelastic response

J. R. Willis (✉)
Centre for Mathematical Sciences, Wilberforce Road, CB3 0WA Cambridge, United Kingdom
e-mail: J.R.Willis@damtp.cam.ac.uk

1 Introduction

The author and co-workers over a period of years solved a variety of problems involving the dynamic perturbation of a propagating crack (see e.g. Movchan and Willis 1995; Willis and Movchan 1995, 1997, 2007; further references are given later). During the course of this extended work, the methodology became increasingly refined and the purpose now is to summarise the approach that at the present time appears to be optimal. Virtually all of the previous work was devoted just to the first-order perturbation. Perturbation to higher order in fact requires the introduction of terms different in character (and not needed) at first order. This is exposed by developing the perturbation solution explicitly to second order, showing the way, in principle, to obtaining the solution to any order. The exposition is kept as simple as possible, by restricting attention to in-plane perturbation of a crack, propagating under Mode I loading. In general, the cracked medium can be a slab, occupying the domain

$$D = \{\mathbf{x} : -\infty < x_1, x_2 < \infty, -h < x_3 < h\} \qquad (1.1)$$

subjected to loading which in the absence of the crack would generate the stress field $\sigma_{ij}^A(\mathbf{x}, t)$ and corresponding displacement field $u_i^A(\mathbf{x}, t)$. The crack occupies the surface

$$S_\varepsilon = \{\mathbf{x} : -\infty < x_1 < Vt + \varepsilon\phi(x_2, t), \quad -\infty < x_2 < \infty,\ x_3 = 0\}. \qquad (1.2)$$

The more general case, involving out-of plane perturbation, would define the perturbed crack to lie on the surface $x_3 = \varepsilon\psi(x_1, x_2)$ but here $\psi \equiv 0$ is assumed. Out-of-plane perturbation was first treated by Willis and Movchan (1997) but see also Willis (1999) for a correction. Propagation in an elastic slab was addressed by Movchan et al. (2005). The stress and displacement fields in the presence of the crack are denoted σ_{ij}, u_i. They respond to the same external loading as σ^A_{ij}, u^A_i but, in addition, correspond to zero tractions on the surfaces of the crack and display a discontinuity in displacement across the crack. Thus, the difference fields $\sigma_{ij} - \sigma^A_{ij}$, $u_i - u^A_i$ satisfy the equations of motion with zero body force, together with homogeneous boundary conditions on the boundary of D (excluding S_ε), while on S_ε,

$$\sigma_{i3} - \sigma^A_{i3} = -\sigma^A_{i3},\ -\infty < x_1 < Vt + \varepsilon\phi(x_2, t),\quad x_3 = \pm 0. \tag{1.3}$$

The medium is assumed to be homogeneous or, more generally, to be vertically stratified, with properties only varying with x_3, and it can be linearly viscoelastic, with time-harmonic moduli $C_{ijkl}(\omega)$ at radian frequency ω. First-order perturbation of a crack in an infinite viscoelastic medium has been treated by Woolfries and Willis (1999), Woolfries et al. (2002), Movchan and Willis (2001, 2002).

Now introduce the Green's function G^+_{ij} defined for the medium occupying $x_3 > 0$, with traction boundary conditions applying on $x_3 = 0$ and the same types of conditions as those that define σ^A_{ij} and u^A_i, and σ_{ij} and u_i, on $x_3 = h$.[1] It follows that the difference fields satisfy, on $x_3 = 0$, the relation

$$u_i - u^A_i = -G^+_{ij} * (\sigma_{3j} - \sigma^A_{3j}), \tag{1.4}$$

the convolution being with respect to x_1, x_2 and t. Similarly, for the region $x_3 < 0$, on the boundary $x_3 = 0$,

$$u_i - u^A_i = G^-_{ij} * (\sigma_{3j} - \sigma^A_{3j}). \tag{1.5}$$

From Eqs. (1.4) and (1.5), therefore,

$$[u_i] = -2\langle G_{ij}\rangle * (\sigma_{3j} - \sigma^A_{3j}), \tag{1.6}$$

where $[u_i]$ represents the jump in displacement across $x_3 = 0$ and $\langle G_{ij}\rangle$ is the average of G^+_{ij} and G^-_{ij}.

Specialising to Mode I loading (which implies both symmetry of the loading and of the properties of the medium), the only non-zero jump in displacement is $[u_3]$ and the only non-zero traction is σ_{33}. Calling these, respectively, $[u]_-$ and σ_+, since $[u]_- = 0$ ahead of the crack and $\sigma_+ = 0$ on the crack, it follows that these satisfy the single equation

$$[u]_- + 2G * (\sigma_+ - \sigma^A) = 0, \tag{1.7}$$

where $\langle G_{33}\rangle$ is denoted G and σ^A represents σ^A_{33}. The desired aspect of the solution is the stress intensity factor. Its deduction from relation (1.7) constitutes the main task of this work.

2 Method of solution

Before proceeding further, it is useful to change coordinates to (X, x_2, t) where

$$X = x_1 - Vt. \tag{2.1}$$

It is easy to check that, with the functions re-defined as functions of (X, x_2, t), Eq. (1.7) remains exactly the same, with the convolution now interpreted relative to the new coordinates. It is also relevant to note that, if the Fourier transform of any one of the functions is known relative to the original coordinates—for instance

$$\mathscr{F}G(\xi_1, \xi_2, \omega) = \iiint G(x_1, x_2, t)e^{i(\xi_1 x_1 + \xi_2 x_2 + \omega t)}\, dx_1 dx_2 dt, \tag{2.2}$$

then its Fourier transform relative to the new coordinates is

$$\tilde{G}(\xi_1, \xi_2, \omega) = \iiint G(X + Vt, x_2, t)e^{i(\xi_1 X + \xi_2 x_2 + \omega t)}\, dX dx_2 dt = \mathscr{F}G(\xi_1, \xi_2, \omega - V\xi_1). \tag{2.3}$$

Interpreted relative to the new coordinates, at least if $\varepsilon = 0$, Eq. (1.7) defines a problem of Wiener–Hopf type. It is helpful, therefore, to factorize G as follows:

$$G = G_- * G_+, \tag{2.4}$$

where G_- is zero for $X > 0$ and G_+ is zero for $X < 0$. Correspondingly $\tilde{G}_-$ is analytic in ξ_1 for $\mathrm{Im}(\xi_1) < 0$ and $\tilde{G}_+$ is analytic in ξ_1 for $\mathrm{Im}(\xi_1) > 0$.

The basic relation (1.7) is now expressed in the form

$$\tfrac{1}{2}(G_-)^{-1} * [u]_- + G_+ * \sigma_+ = G_+ * \sigma^A. \tag{2.5}$$

In preparation for describing the general strategy, consider first the case $\varepsilon = 0$. The convolution $G_+ * \sigma_+$ is

[1] In the case $h \to \infty$, G^+_{ij} satisfies a radiation condition so that it is composed from waves travelling away from the surface $x_3 = 0$.

a "+" function while $(G_-)^{-1} * [u]_-$ is a "−" function. It follows from (2.5) that

$$G_+ * \sigma_+ = \{G_+ * \sigma^A\}_+, \tag{2.6}$$

meaning that the right side is defined to be zero when $X < 0$. Hence, formally,

$$\sigma_+ = (G_+)^{-1} * \{G_+ * \sigma^A\}_+. \tag{2.7}$$

Similarly,

$$[u]_- = 2G_- * \{G_+ * \sigma^A\}_-. \tag{2.8}$$

Evidently, the type of singularity displayed by σ_+, as $X \to 0$, is determined by the singularity of $(G_+)^{-1}$.[2] Furthermore, the coefficient of this singularity is proportional to $\{G_+ * \sigma^A\}_+$, evaluated as $X \to 0$. In the case of viscoelasticity and subsonic V, the singularity in σ_+ is of square-root type,

$$\sigma_+ \sim K_0/(2\pi X)^{1/2} \tag{2.9}$$

and, with suitable normalization of G_+,

$$K_0 = \lim_{X\to 0} \{G_+ * \sigma^A\}_+, \tag{2.10}$$

the suffix on K_0 indicating that $\varepsilon = 0$. For "intersonic" V, a similar conclusion holds but the singularity is no longer of square-root type. The relevant analysis has been performed by Obrezanova and Willis (2003, 2008). Only subsonic V is considered from now on.

The general strategy is now outlined. The factors G_+ and G_- are hard to find but it is a routine matter to obtain their Fourier transforms; hence, working is mostly performed within the Fourier domain.

First, relation (2.5) is considered as $X \to \pm 0$. Explicitly,

$$\begin{aligned}\sigma_+ \sim\ & K(\varepsilon)/(2\pi(X-\varepsilon\phi))^{1/2} + A(\varepsilon)(X-\varepsilon\phi)^{1/2} \\ & + B(\varepsilon)(X-\varepsilon\phi)^{3/2} + \cdots + \sigma_+^*(X-\varepsilon\phi, x_2, t),\end{aligned} \tag{2.11}$$

where σ_+^* is non-singular and vanishes together with all its X-derivatives as $X \to \varepsilon\phi$. Correspondingly,

$$\begin{aligned}[u]_- \sim\ & K^u(\varepsilon)(\varepsilon\phi - X)^{1/2} + A^u(\varepsilon\phi - X)^{3/2} \\ & + B^u(\varepsilon\phi - X)^{5/2} + \cdots + u_-^*(X-\varepsilon\phi, x_2, t).\end{aligned} \tag{2.12}$$

Also,

$$\begin{aligned}\{G_+ * \sigma^A\}(X) \sim\ & \{G_+ * \sigma^A\}(0) + X\{G_+ * (\sigma^A)'\}(0) \\ & + \tfrac{1}{2}X^2\{G_+ * (\sigma^A)''\}(0) + \cdots,\end{aligned} \tag{2.13}$$

having left implicit the dependence on x_2 and t. The prime denotes differentiation with respect to X.

Next, evaluate the Fourier transform of each side, with respect to X; this transformation will be indicated with a hat symbol. Considering $X \to 0$ is equivalent to considering $\xi_1 \to \infty$. The transform of the left side of (2.5) contains the products of the transforms of the participating functions and hence involves, as $X \to 0$, the asymptotic forms of $\hat{G}_+$ and $(\hat{G}_-)^{-1}$ as $\xi_1 \to \infty$. These are easiest obtained from the corresponding asymptotic forms of the transforms with respect to all arguments:

$$\tilde{G}_+ \sim \frac{(2i)^{1/2}}{(\xi_1 + 0i)^{1/2}} \left\{1 + i\frac{\overline{Q}_1}{\xi_1 + 0i} - \frac{\overline{Q}_2}{(\xi_1 + 0i)^2}\right\} \tag{2.14}$$

$$(\tilde{G}_-)^{-1} \sim \frac{2(2i)^{1/2}(\xi_1 - 0i)^{1/2}}{\mathscr{A}(V)} \left\{1 + i\frac{\overline{R}_1}{\xi_1 - 0i} - \frac{\overline{R}_2}{(\xi_1 - 0i)^2}\right\}, \tag{2.15}$$

having chosen what will be the right normalization for $\tilde{G}_+$ and employed the asymptotic form

$$\tilde{G} \sim \frac{\mathscr{A}(V)}{2|\xi_1|} \left\{1 + i\frac{\overline{G}_1}{\xi_1} - \frac{\overline{G}_2}{\xi_1^2}\right\} \tag{2.16}$$

for $\tilde{G}$. The functions $\overline{Q}_1$ etc. depend on ξ_2 and ω. Correspondingly, Q_1 etc. depend on x_2 and t. They conform to the relations

$$\overline{G}_1 = \overline{Q}_1 - \overline{R}_1, \quad \overline{G}_2 = \overline{Q}_2 - \overline{R}_2 - \overline{R}_1(\overline{Q}_1 - \overline{R}_1). \tag{2.17}$$

Transforming (2.11) and (2.12) with respect to X gives

$$\begin{aligned}\hat{\sigma}_+ \sim\ & \left\{\frac{(i/2)^{1/2}K(\varepsilon)}{(\xi_1 + 0i)^{1/2}} + \frac{\pi^{1/2}(i)^{3/2}A(\varepsilon)}{2(\xi_1 + 0i)^{3/2}}\right. \\ & \left. - \frac{3\pi^{1/2}(i)^{1/2}B(\varepsilon)}{4(\xi_1 + 0i)^{5/2}}\right\} e^{i\varepsilon\xi_1\phi} + \hat{\sigma}_+^*,\end{aligned} \tag{2.18}$$

$$\begin{aligned}[\hat{u}]_- \sim\ & \left\{\frac{K^u(\varepsilon)\pi^{1/2}(-i)^{3/2}}{2(\xi_1 - 0i)^{3/2}} - \frac{3A^u(\varepsilon)\pi^{1/2}(-i)^{1/2}}{4(\xi_1 - 0i)^{5/2}}\right. \\ & \left. - \frac{15B^u(\varepsilon)\pi^{1/2}(-i)^{3/2}}{8(\xi_1 - 0i)^{7/2}}\right\} e^{i\varepsilon\xi_1\phi} + \hat{u}_-^*.\end{aligned} \tag{2.19}$$

Transforming (2.13) with respect to X gives

$$\begin{aligned}\widehat{\{G_+ * \sigma^A\}} \sim\ & \left\{\frac{i}{\xi_1 + 0i} - \frac{i}{\xi_1 - 0i}\right\} \{G_+ * \sigma^A\}(0) \\ & - \left\{\frac{1}{(\xi_1 + 0i)^2} - \frac{1}{(\xi_1 - 0i)^2}\right\} \{G_+ * (\sigma^A)'\}(0) \\ & - \left\{\frac{i}{(\xi_1 + 0i)^3} - \frac{i}{(\xi_1 - 0i)^3}\right\} \{G_+ * (\sigma^A)''\}(0).\end{aligned} \tag{2.20}$$

[2] If $(G_+)^{-1}$ has a singularity like X^{-s} then σ_+ has a singularity like $X^{(1-s)}$.

The formulae just presented allow the development of what amounts to the "inner limit" of relation (2.5), when $X \to 0$ (if preferred, when $X = O(\varepsilon)$). The next step is to make an "outer expansion" of this inner limit. This is done by now regarding X as fixed (and so finite) and letting $\varepsilon \to 0$. This is equivalent to taking ξ_1 fixed and letting $\varepsilon \to 0$, which means adopting the expansion (to second order in ε)

$$e^{i\xi_1\varepsilon\phi} \sim 1 + i\xi_1\varepsilon\phi - \tfrac{1}{2}(\xi_1\varepsilon\phi)^2. \tag{2.21}$$

This, evidently, is equivalent to letting $\xi_1 \to 0$, or $X \to \infty$. Relation (2.5) is an identity; the result of taking these two limits remains an identity from which the desired information can be deduced. The approach is conceptually simple but the new identity contains many terms, summarised in the "Appendix", Section 5.

3 Deductions

Now set

$$K(\varepsilon) \sim K_0 + \varepsilon K_1 + \varepsilon^2 K_2 \tag{3.1}$$

with similar expansions for $A(\varepsilon)$, $B(\varepsilon)$ and the parameters with superscript u.

Implications of the identity (2.5) will be explored, using the expressions (2.20), (5.1) and (5.2). The terms $\widehat{G_+ * \sigma_+^*}$ and $\widehat{(G_-)^{-1} * u_-^*}$ are disregarded in the first instance; they will be introduced only when strictly necessary and this will first occur at order ε^2. At order ε^0, the identity (2.5) delivers three relations, the first of which is

$$K_0\frac{i}{\xi_1+0i} - \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}K_0^u\frac{i}{\xi_1-0i} = \{G_+ * \sigma^A\}(0)\left\{\frac{i}{\xi_1+0i} - \frac{i}{\xi_1-0i}\right\}, \tag{3.2}$$

from which it follows that

$$K_0 = \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}K_0^u = \{G_+ * \sigma^A\}(0). \tag{3.3}$$

The next, considering terms containing $1/\xi_1^2$, yields

$$(\pi/2)^{1/2}A_0 + Q_1 * K_0 = -\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}[\tfrac{3}{2}A_0^u - R_1 * K_0^u] = \{G_+ * (\sigma^A)'\}(0). \tag{3.4}$$

Finally, from the terms containing $1/\xi_1^3$,

$$\tfrac{3}{2}(\pi/2)^{1/2}B_0 + (\pi/2)^{1/2}Q_1 * A_0 + Q_2 * K_0 = \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}[\tfrac{15}{4}B_0^u - \tfrac{3}{2}R_1 * A_0^u + R_2 * K_0^u] = \{G_+ * (\sigma^A)''\}(0). \tag{3.5}$$

It may be noted that (3.4), taken with (3.3) and (2.17), implies the relation

$$A_0^u = -\tfrac{2}{3}\mathscr{A}(V)[A_0 + (2/\pi)^{1/2}G_1 * K_0]. \tag{3.6}$$

Similarly,

$$B_0^u = \tfrac{2}{5}\mathscr{A}(V)[B_0 + \tfrac{2}{3}(G_1 * A_0 + (2/\pi)^{1/2}G_2 * K_0)]. \tag{3.7}$$

This provides some check on the algebra because σ_+ and $[u]_-$ satisfy the relation (1.7) which involves only the Green's function G.

Next, consider terms of order ε. The first, which is independent of ξ_1, gives

$$-\phi K_0 + \phi\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}K_0^u = 0, \tag{3.8}$$

which is true on account of (3.3). Now for the terms containing $1/\xi_1$,

$$K_1 - (\pi/2)^{1/2}\phi A_0 - Q_1 * (\phi K_0) = \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}[K_1^u + \tfrac{3}{2}\phi A_0^u - R_1 * (\phi K_0^u)] = 0. \tag{3.9}$$

The terms containing $1/\xi_1^2$ give

$$\begin{aligned}&(\pi/2)^{1/2}(A_1 - \tfrac{3}{2}\phi B_0) + Q_1 * (K_1 - (\pi/2)^{1/2}\phi A_0) - Q_2 * (\phi K_0)\\ &= -\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}[\tfrac{3}{2}A_1^u + \tfrac{15}{4}\phi B_0^u - R_1 * (K_1^u + \tfrac{3}{2}\phi A_0^u) + R_2 * (\phi K_0^u)]\\ &= 0.\end{aligned} \tag{3.10}$$

The terms at order ε that contain $1/\xi_1^3$ are incomplete through truncation of the basic expansions at just three terms. Thus, B_1 is undetermined.

Using results already obtained, the first equality in (3.9) can be expressed

$$\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}K_1^u = K_1 + \phi G_1 * K_0 - G_1 * (\phi K_0) \tag{3.11}$$

and the first equality in (3.10) gives

$$-\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}(\tfrac{3}{2}A_1^u + \tfrac{15}{4}\phi B_0^u) = (\pi/2)^{1/2}(A_1 - \tfrac{3}{2}\phi B_0) + G_1 * (K_1 - (\pi/2)^{1/2}\phi A_0) - G_2 * (\phi K_0). \tag{3.12}$$

Further reduction is possible but the relations as displayed are obtainable directly from expansion of (1.7).

Consider, finally, terms of order ε^2. The terms that contain $i\xi_1$ give

$$-\tfrac{1}{2}\phi^2 K_0 + \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}(\tfrac{1}{2}\phi^2 K_0^u) = 0, \tag{3.13}$$

which is already known to be true. Now, however, consider the terms that are independent of ξ_1. They give

$$\begin{aligned}&-\phi K_1 + \tfrac{1}{2}(\pi/2)^{1/2}\phi^2 A_0 + \tfrac{1}{2}Q_1 * (\phi^2 K_0)\\ &\quad + \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}[\phi K_1^u + \tfrac{3}{4}\phi^2 A_0^u - \tfrac{1}{2}R_1 * (\phi^2 K_0^u)]\\ &\quad + \cdots = 0,\end{aligned} \tag{3.14}$$

where the terms not shown explicitly are associated with $\{\widehat{G_+ * \sigma_+^*} + \tfrac{1}{2}\widehat{(G_-)^{-1} * u_-^*}\}$.

Equation (3.14) simplifies to

$$\tfrac{1}{2}[G_1 * (\phi^2 K_0) + \phi^2 G_1 * K_0] - \phi G_1 * (\phi K_0) + \cdots = 0. \tag{3.15}$$

It can be satisfied by assuming that σ_+^* and u_-^* have "outer" expansions (i.e. as $\varepsilon \to 0$ with X fixed or, equivalently, as $X \to \infty$ with ε fixed)

$$\sigma_+^* \sim \varepsilon^2 C^* X^{-5/2}, \quad u_-^* \sim \varepsilon^2 C^{*u}(-X)^{-3/2}. \tag{3.16}$$

Correspondingly, as $\xi_1 \to 0$,

$$\begin{aligned}&\hat{\sigma}_+^* \sim \tfrac{4}{3}\varepsilon^2 C^* \pi^{1/2}(i)^{-3/2}(\xi_1 + 0i)^{3/2},\\ &\hat{u}_-^* \sim -2\varepsilon^2 C^{*u}\pi^{1/2}(i)^{1/2}(\xi_1 - 0i)^{1/2}.\end{aligned} \tag{3.17}$$

These must be multiplied[3], respectively, by the asymptotic forms of $\hat{G}_+$, $(\hat{G}_-)^{-1}$, obtainable from (2.14), (2.15). The leading-order terms in the resulting expressions are proportional to $i\xi_1$; they do not spoil Eq. (3.13) if

$$C^{*u} = -\tfrac{2}{3}\mathscr{A}(V)C^*. \tag{3.18}$$

The complete version of Eq. (3.15) now becomes

$$\begin{aligned}&\tfrac{1}{2}[G_1 * (\phi^2 K_0) + \phi^2 G_1 * K_0]\\ &\quad -\phi G_1 * (\phi K_0) + \tfrac{8}{3}(\pi/2)^{1/2} G_1 * C^* = 0.\end{aligned} \tag{3.19}$$

[3] Stated more precisely, convolutions with respect to x_2 and t must be performed.

Thus, C^* and C^{*u} are determined. Finally, the terms of order ε^2 that contain $1/\xi_1$ give

$$\begin{aligned}&K_2 - (\pi/2)^{1/2}\phi A_1 + \tfrac{3}{4}(\pi/2)^{1/2}\phi^2 B_0\\ &\quad -Q_1 * (\phi K_1 - \tfrac{1}{2}(\pi/2)^{1/2}\phi^2 A_0)\\ &\quad +Q_2 * (\tfrac{8}{3}(\pi/2)^{1/2} C^* + \tfrac{1}{2}\phi^2 K_0)\\ &= \frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}\Big\{K_2^u + \tfrac{3}{2}\phi A_1^u + \tfrac{15}{8}\phi^2 B_0^u\\ &\quad -R_1 * (\phi K_1^u + \tfrac{3}{4}\phi^2 A_0^u)\\ &\quad -R_2 * (4(\pi/2)^{1/2} C^{*u} - \tfrac{1}{2}\phi^2 K_0^u)\Big\}\\ &= 0.\end{aligned} \tag{3.20}$$

The terms of order ε^2 that contain $1/\xi_1^2$ or $1/\xi_1^3$ would require for their completion more terms in the original expansions. Thus, A_2 and B_2, as well as B_1, remain undetermined. Using results already established, the first equality in (3.20) can be expressed in the form

$$\begin{aligned}&\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}K_2^u = K_2 + \phi G_1 * K_1 - G_1 * (\phi K_1)\\ &\quad +(\pi/2)^{1/2}\{\tfrac{1}{2}[G_1 * (\phi^2 A_0) + \phi^2 G_1 * A_0] - \phi G_1 * (\phi A_0)\}\\ &\quad +\tfrac{1}{2}[\phi^2 G_2 * K_0 + G_2 * (\phi^2 K_0)] - \phi G_2 * (\phi K_0)\\ &\quad +G_2 * (\tfrac{8}{3}(\pi/2)^{1/2} C^*),\end{aligned} \tag{3.21}$$

again consistent with the fact that surface traction and displacement are related through the Green's function. It follows also that

$$\begin{aligned}K_2 = {}& Q_1 * [\phi Q_1 * (\phi K_0)] - \phi Q_1^2 * (\phi K_0) + \phi Q_2 * (\phi K_0)\\ &-\tfrac{1}{2}Q_2 * (\phi^2 K_0)\\ &+\tfrac{1}{2}(\pi/2)^{1/2} Q_1 * (\phi^2 A_0) + \tfrac{3}{4}(\pi/2)^{1/2}(\phi^2 B_0)\\ &-\tfrac{8}{3}(\pi/2)^{1/2} Q_2 * C^*,\end{aligned} \tag{3.22}$$

having substituted the relevant expressions for K_1 and A_1.

Finally, with the definitions

$$L_0 = \{G_+ * (\sigma^A)'\}|_{X=0}, \quad M_0 = \{G_+ * (\sigma^A)''\}|_{X=0} \tag{3.23}$$

and using $(3.4)_1$ and $(3.5)_1$, the result can be expressed in the form

$$\begin{aligned}K_2 = {}& Q_1 * [\phi Q_1 * (\phi K_0)] - \phi Q_1^2 * (\phi K_0)\\ &+\tfrac{1}{2}[\phi^2 Q_1^2 * K_0 - Q_1 * (\phi^2 Q_1 * K_0)]\\ &+\phi Q_2 * (\phi K_0) - \tfrac{1}{2}[Q_2 * (\phi^2 K_0) + \phi^2 Q_2 * K_0]\\ &+\tfrac{1}{2}[Q_1 * (\phi^2 L_0) - \phi^2 Q_1 * L_0] + \tfrac{1}{2}\phi^2 M_0\\ &-\tfrac{8}{3}(\pi/2)^{1/2} Q_2 * C^*.\end{aligned} \tag{3.24}$$

4 Example: infinite uniform isotropic medium

It can be shown that, for this case, on the surface $x_3 = 0$,

$$G^+_{33}(x_1, x_2, t) = G^-_{33}(x_1, x_2, t) = G(x_1, x_2, t) \quad (4.1)$$

and that, relative to the frame (x_1, x_2, t),

$$\mathscr{F}G(\xi_1, \xi_2, \omega) = \frac{(\omega^2/b^2)(\omega^2/a^2 - |\xi|^2)^{1/2}}{i\mu D(|\xi|, \omega)}, \quad (4.2)$$

where $|\xi| = (\xi_1^2 + \xi_2^2)^{1/2}$ and a, b are the speeds of dilatational and shear waves:

$$a^2 = (\lambda + 2\mu)/\rho, \quad b^2 = \mu/\rho. \quad (4.3)$$

The medium has Lamé moduli λ, μ and density ρ, and

$$D(|\xi|, \omega) = 4|\xi|^2(\omega^2/a^2 - |\xi|^2)^{1/2}(\omega^2/b^2 - |\xi|^2)^{1/2} + (\omega^2/b^2 - 2|\xi|^2)^2. \quad (4.4)$$

Then, relative to the moving frame, $\tilde{G}(\xi_1, \xi_2, \omega)$ is given by replacing ω with $\omega - V\xi_1$, as prescribed in (2.3).

It follows that, as $\xi_1 \to \infty$, $\tilde{G}$ has the expansion (2.16), with

$$\mathscr{A}(V) = \frac{2(V^2/b^2)\alpha}{\mu R(V)}, \quad (4.5)$$

where $R(V)$ is the Rayleigh discriminant

$$R(V) = 4\alpha\beta - (1 + \beta^2)^2, \quad (4.6)$$

with

$$\alpha = (1 - V^2/a^2)^{1/2}, \quad \beta = (1 - V^2/b^2)^{1/2}. \quad (4.7)$$

If the medium is viscoelastic, then a and b become functions of ω relative to the original stationary frame and functions of $\omega - V\xi_1$ relative to the moving frame. It follows immediately that

$$\overline{G}_1(\xi_2, \omega) = -i\omega\frac{\mathscr{A}'(V)}{\mathscr{A}(V)}. \quad (4.8)$$

The corresponding operator in physical space is

$$G_1* = \frac{\mathscr{A}'(V)}{\mathscr{A}(V)}\frac{\partial}{\partial t}. \quad (4.9)$$

The first dependence on x_2 comes in with G_2.

Before proceeding further, consider Eq. (3.19) for C^*: it reduces to

$$\tfrac{8}{3}(\pi/2)^{1/2}\partial C^*/\partial t = -\tfrac{1}{2}[\partial(\phi^2 K_0)/\partial t + \phi^2 \partial K_0/\partial t] + \phi\partial(\phi K_0)/\partial t \equiv 0. \quad (4.10)$$

Thus, remarkably, it is consistent to take $C^* = 0$.[4] Furthermore, this conclusion would persist even if the medium were anisotropic, so long as the basic propagation is in a direction of symmetry. If the coupled Mode II-III problem were considered, even for an isotropic medium, terms analogous to C^* would at least have to be admitted. It is not known at the time of writing whether or not such terms are zero if the medium is isotropic. Detailed analysis is in progress and will be reported separately.

Having established that $C^* = 0$, for the sake of a simple illustration, the problem is pursued just in the static limit, for which there is no distinction between $\mathscr{F}G$ and $\tilde{G}$ and the variable ω is simply absent. Thus,

$$\mathscr{F}G = \tilde{G} = \frac{1 - \nu}{\mu|\xi|}, \quad (4.11)$$

where $\nu = \lambda/2(\lambda + \mu)$ is Poisson's ratio. It follows immediately that

$$\tilde{G}_+ = \frac{(2i)^{1/2}}{(\xi_1 + i|\xi_2|)^{1/2}},$$
$$(\tilde{G}_-)^{-1} = \frac{(2i)^{1/2}\mu(\xi_1 - i|\xi_2|)^{1/2}}{1 - \nu} \quad (4.12)$$

so that

$$\overline{Q}_1 = -\tfrac{1}{2}|\xi_2|, \quad \overline{Q}_2 = \tfrac{3}{8}\xi_2^2. \quad (4.13)$$

Also,

$$\overline{G}_1 = 0, \quad \overline{G}_2 = \tfrac{1}{2}\xi_2^2. \quad (4.14)$$

Taken together with (3.22), these relations suffice to complete the solution of the elastostatic problem.

Consider, finally, the particular case that σ^A depends on x_1 only so that K_0, A_0 and B_0 are constants, and take the perturbation to be

$$\varepsilon\phi(x_2) = a\cos(kx_2) \quad (4.15)$$

(so that a is a suitably small length). In this case,

$$Q_1 * \phi = \mathrm{Re}\int Q_1(x_2')e^{ik(x_2 - x_2')}\,dx_2' = -\tfrac{1}{2}|ka|\cos(kx_2), \quad (4.16)$$

since $\overline{Q}_1$ is given by $(4.13)_1$. A repeat of this type of reasoning gives

$$Q_1 * [\phi Q_1 * \phi] = \tfrac{1}{4}(ka)^2[\cos^2(kx_2) - \sin^2(kx_2)]. \quad (4.17)$$

[4] Movchan and Willis (2001) unwisely included (but did not use) some results for second-order perturbation that were derived simply by comparing terms containing what is here called K_2, without checking the consistency of the other second-order terms. Their formula is correct only because, as now established, $C^* = 0$ in the case that they considered.

The other terms in (3.22) are even simpler to obtain, since Q_1^2 and Q_2 are both proportional to $\partial^2/\partial x_2^2$. The final result is that

$$K_2 = \tfrac{1}{8}K_0(ka)^2\sin^2(kx_2) \\ -\tfrac{1}{4}(\pi/2)^{1/2}(A_0a)|ka|\cos(2kx_2) \\ +\tfrac{3}{4}(\pi/2)^{1/2}(B_0a^2)\cos^2(kx_2) \qquad (4.18)$$

or, in terms of the parameters L_0 and M_0 defined by (3.23),

$$K_2 = \tfrac{1}{8}K_0(ka)^2\sin^2(kx_2) - \tfrac{1}{4}(L_0a)|ka|\cos(2kx_2) \\ +\tfrac{1}{2}(M_0a^2)\cos^2(kx_2). \qquad (4.19)$$

It is necessary to make one final observation: the definition of $K(\varepsilon)$, implicit in relation (2.11), was adopted for mathematical convenience. The conventional stress intensity factor, K_I say, expresses the stress directly ahead of the crack, asymptotically, as $K_I/(2\pi r)^{1/2}$, where r is the distance taken in the direction normal to the crack. This makes no difference at the level of first-order perturbation but, to obtain K_I to second order, it is necessary to subtract $\frac{1}{4}K_0(\varepsilon\phi')^2 = \frac{1}{4}K_0(ka)^2\sin^2(kx_2)$ in the present case. Thus, to second order,

$$K_I \sim K_0 - \tfrac{1}{2}K_0|ka|\cos(kx_2) + (L_0a)\cos(kx_2) \\ -\tfrac{1}{8}K_0(ka)^2\sin^2(kx_2) - \tfrac{1}{4}(L_0a)|ka|\cos(2kx_2) \\ +\tfrac{1}{2}(M_0a^2)\cos^2(kx_2). \qquad (4.20)$$

This result was recently derived, by an entirely different method based on work of Rice (1989), by Leblond et al. (2012), in the special case $L_0 = M_0 = 0$. The complete result is consistent with the general formula derived for elastostatics by Vasoya et al. (2013).

5 Appendix: Expansion of the basic identity to second order

The term that must be equated to the right side of (2.20) is composed from the limiting operations applied to $\frac{1}{2}(G_-)^{-1} * [u]_- + G_+ * \sigma_+$. Completing the algebra gives

$$\widehat{G_+ * \sigma_+} \sim -\tfrac{1}{2}i\xi_1\varepsilon^2\phi^2K - \varepsilon\phi K + \tfrac{1}{2}(\pi/2)^{1/2}\varepsilon^2\phi^2A \\ +\tfrac{1}{2}\varepsilon^2Q_1*(\phi^2K) + \Big\{K - \varepsilon(\pi/2)^{1/2}\phi A + \tfrac{3}{4}\varepsilon^2(\pi/2)^{1/2}\phi^2B \\ -Q_1*[\varepsilon\phi K - \tfrac{1}{2}(\pi/2)^{1/2}\varepsilon^2\phi^2A] + \tfrac{1}{2}\varepsilon^2Q_2*(\phi^2K)\Big\}\frac{i}{\xi_1+0i} \\ -\Big\{(\pi/2)^{1/2}(A - \tfrac{3}{2}\varepsilon\phi B) + Q_1*(K - \varepsilon(\pi/2)^{1/2}\phi A \\ +\tfrac{3}{4}\varepsilon^2\phi^2B) - Q_2*[\varepsilon\phi K - \tfrac{1}{2}(\pi/2)^{1/2}\varepsilon^2\phi^2A]\Big\}\frac{1}{(\xi+0i)^2} \\ -\Big\{\tfrac{3}{2}(\pi/2)^{1/2}B + (\pi/2)^{1/2}Q_1*(A - \tfrac{3}{2}\varepsilon\phi B) \\ +Q_2*[K - \varepsilon(\pi/2)^{1/2}\phi A + \tfrac{3}{4}\varepsilon^2(\pi/2)^{1/2}\phi^2B]\Big\}\frac{i}{(\xi_1+0i)^3} \\ +\widehat{G_+ * \sigma_+^*}. \qquad (5.1)$$

Also,

$$\tfrac{1}{2}\widehat{(G_-)^{-1} * [u]_-} \sim -\frac{(\pi/2)^{1/2}}{\mathscr{A}(V)}\Big[-\tfrac{1}{2}i\xi_1\varepsilon^2\phi^2K^u \\ -\varepsilon\phi K^u - \tfrac{3}{4}\varepsilon^2\phi^2A^u + \tfrac{1}{2}\varepsilon^2R_1*(\phi^2K^u) \\ +\Big\{K^u + \tfrac{3}{2}\varepsilon\phi A^u + \tfrac{15}{8}\varepsilon^2\phi^2B^u \\ -R_1*(\varepsilon\phi K^u + \tfrac{3}{4}\varepsilon^2\phi^2A^u) + \tfrac{1}{2}\varepsilon^2R_2*(\phi^2K^u)\Big\}\frac{i}{\xi_1-0i} \\ +\Big\{\tfrac{3}{2}A^u + \tfrac{15}{4}\varepsilon\phi B^u - R_1*(K^u + \tfrac{3}{2}\varepsilon\phi A^u + \tfrac{15}{8}\varepsilon^2\phi^2B^u) \\ +R_2*(\varepsilon\phi K^u + \tfrac{3}{4}\varepsilon^2\phi^2A^u)\Big\}\frac{1}{(\xi_1-0i)^2} \\ -\Big\{\tfrac{15}{4}B^u - R_1*(\tfrac{3}{2}A^u + \tfrac{15}{4}\varepsilon\phi B^u) \\ +R_2*(K^u + \tfrac{3}{2}\varepsilon\phi A^u + \tfrac{15}{8}\varepsilon^2\phi^2B^u)\Big\}\frac{i}{(\xi_1-0i)^3}\Big] \\ +\tfrac{1}{2}\widehat{(G_-)^{-1} * u_-^*}. \qquad (5.2)$$

The exact forms taken for $\widehat{G_+ * \sigma_+^*}$ and $\widehat{(G_-)^{-1} * u_-^*}$ are discussed in the main text.

References

Leblond J-B, Patinet S, Frelat J, Lazarus V (2012) Second-order coplanar perturbation of a semi-infinite crack in an infinite body. Eng Fract Mech 90:129–142

Movchan AB, Willis JR (1995) Dynamic weight functions for a moving crack. II. Shear loading. J Mech Phys Solids 43:1369–1383

Movchan AB, Willis JR (2001) The influence of viscoelasticity on crack front waves. J Mech Phys Solids 49:2177–2189

Movchan AB, Willis JR (2002) Theory of crack front waves. In: Abrahams ID, Martin PA, Simons MJ (eds) Diffraction and scattering in fluid mechanics and elasticity. Kluwer, Dordrecht, pp 235–250

Movchan NV, Movchan AB, Willis JR (2005) Perturbation of a dynamic crack in an infinite strip. Q J Mech Appl Math 58:333–347

Obrezanova O, Willis JR (2003) Stability of intersonic shear crack propagation. J Mech Phys Solids 51:1957–1970

Obrezanova O, Willis JR (2008) Stability of an intersonic crack to a perturbation of its edge. J Mech Phys Solids 56:51–69

Rice JR (1989) Weight function theory for three-dimensional elastic crack analysis. In: Wei RP, Gangloff RP (eds) Fracture mechanics: perspectives and directions (twentieth symposium), ASTM STP 1020. American Society for Testing and Materials, Philadelphia, pp 29–57

Vasoya M, Leblond J-B, Ponson L (2013) A geometrically non-linear analysis of coplanar crack propagation in some heterogeneous medium. Int J Solids Struct 50:371–378

Willis JR, Movchan AB (1995) Dynamic weight functions for a moving crack. I. Mode I loading. Mech Phys Solids 43:319–341

Willis JR, Movchan AB (1997) Three-dimensional dynamic perturbation of a propagating crack. J Mech Phys Solids 45:591–610

Willis JR (1999) Asymptotic analysis in fracture: an update. Int J Fract 100:85–103

Willis JR, Movchan NV (2007) Crack front waves in an anisotropic medium. Wave Motion 44:458–471

Woolfries S, Willis JR (1999) Perturbation of a dynamic planar crack moving in a model elastic solid. J Mech Phys Solids 47:1633–1661

Woolfries S, Movchan AB, Willis JR (2002) Perturbation of a dynamic planar crack moving in a model viscoelastic solid. Int J Solids Struct 39:5409–5426

DOI 10.1007/978-3-319-04397-5_4
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9812-9

ORIGINAL PAPER

Localisation near defects and filtering of flexural waves in structured plates

S. G. Haslinger · R. C. McPhedran · N. V. Movchan · A. B. Movchan

Received: 2 October 2012 / Accepted: 24 January 2013 / Published online: 19 February 2013

Abstract The paper deals with localisation of flexural waves within gratings composed of either pinned points or rigid inclusions of finite radius in a structured plate. We study the filtering and resonant action of such systems. The effect of the finite size of inclusions on the dynamic localisation is analysed for the range of frequencies where only zeroth grating orders propagate. The structure of the resonant modes within gratings of inclusions is of special interest. In particular, we consider the circumstances under which such gratings can deliver for flexural waves a phenomenon similar to Electromagnetically Induced Transparency, where a resonant maximum of transmission is cut in two by a resonant minimum. We identify system designs which yield very high concentration of flexural fields within the interface that may lead to a further structural failure.

Keywords Flexural waves · Elastodynamics · Wave filters · Structured plates · Localisation

Electronic supplementary material The online version of this article (doi:10.1007/s10704-013-9812-9) contains supplementary material, which is available to authorized users.

S. G. Haslinger (✉) · N. V. Movchan · A. B. Movchan
Department of Mathematical Sciences, Mathematical Sciences Building, Peach Street, Liverpool L69 3BX, United Kingdom
e-mail:sgh@liverpool.ac.uk

R. C. McPhedran
CUDOS, School of Physics, University of Sydney 2006, Sydney, NSW, Australia

1 Introduction

In problems of structural mechanics, studies of vibration of plates appear to be important for prediction of the response of elastic systems to earthquakes and other dynamic loads. Structures like aircraft, long bridges and reinforced roads would also involve an arrangement of inclusions, voids or masses as a part of the structural design. Stress concentration often occurs for different dynamic regimes. Special resonance regimes characterised by an enhanced transmission are considered in this paper.

We consider here the interaction of flexural waves in thin elastic plates with a structured interface consisting of a finite number of periodic gratings. Finite-thickness interfaces were analysed by Bigoni and Movchan (2002) for both static and dynamic cases. The term structural interface was introduced for a finite-width material possessing inertia joining continuous media, with continuum and semi-discrete microstructures being considered. The elastostatic theory of structural interfaces was advanced by Bertoldi et al. (Parts I - III, 2007). In the dynamic examples, the effect of inertia was shown to be crucial for the filtering properties of the interface. For low frequency, time-harmonic vibrations and elastic waves of a specific nature, it was shown that exceptionally narrow pass bands were demonstrable for the periodic structures. Further analysis of Bloch-Floquet waves for periodic systems showed that structural interfaces can be designed to illustrate similarly interesting properties of

filtering for elastic waves (Bigoni et al. 2008; Gei et al. 2009 and Brun et al. 2010). Our periodically structured interface supports sharp transmission resonances for low-frequency flexural vibrations for the interaction of a plane wave with the structured plate.

Such interaction problems, together with resonant interaction and trapped waves, have been the subject of a number of papers in recent years, notably by Evans and Porter (2007), Movchan et al. (2007), Movchan et al. (2009), Movchan et al. (2011), Farhat et al. (2010). A related aspect that is also of great interest concerns the design of systems which can cloak objects from detection by flexural wave scattering by placing them within structured "shields" (see, for example, Farhat et al. 2009 and Stenger et al. 2012).

The investigations here were inspired by the striking results obtained in our previous papers Movchan et al. (2009) and Haslinger et al. (2012). In that work the interaction of flexural waves with gratings composed of fixed pins of zero radius was considered. It was shown that very high quality factor (Q) resonances could be easily obtained, with Q of around 3.5×10^4 being exhibited for a system of two gratings (see Haslinger et al. 2012). Furthermore, for systems of three gratings it was shown that one could achieve a filtering action showing a strong resemblance to the quantum mechanical phenomenon of Electromagnetically Induced Transparency (for an authoritative review of this topic see Fleischhauer et al. 2005).

We also mention that there have been related investigations involving transmission properties of electromagnetic waves through doubly periodic grids (Botten et al. 1985; Ulrich and Tacke 1973; Pelton and Munk 1979). In these papers, square symmetric grids illuminated by normally-incident plane waves exhibited a resonant transmission maximum, whereas for slightly off-normal incidence, the transmission maximum was bisected by a zero of transmission. It was shown that this was due to the presence of a mode with the opposite symmetry (odd) to that of the resonant mode for normal incidence (even), the former not being able to couple to the incident wave. However, for off-axis angles, its coupling coefficient became nonzero, and the null of reflectance was then caused by the resonances of the even and odd modes coinciding and cancelling each other. The paper of Botten et al. (1985) gives information on the phase as well as amplitude of reflection and transmission coefficients, and also remarks that the resonance of the even mode can involve the carriage of more than 100 % of the incident energy in the downward direction (i.e. parallel to the incident wave), since the odd mode resonance involves energy being carried upwards (i.e. anti-parallel to the incident wave).

The effect of resonance transmission is illustrated in Fig. 1 where the field plot shows the flexural displacement inside a triple grating stack of rigid pins and the plane wave outside the stack. The plane wave appears to be virtually unperturbed for the chosen frequency. In the present paper we show the effect of varying the arrangement of inclusions within the gratings, and analyse the resonance modes for inclusions of nonzero radius.

In the main text of the paper we use the abbreviation EDIT for the Elasto-Dynamically Inhibited Transparency. The change in the geometry from zero radius pins to inclusions can have important effects on the EDIT phenomenon. In particular, the symmetry of the localised field is affected by the scattering from the inclusions of nonzero radius: it has been noted by Haslinger et al. (2012) that an anti-symmetric resonance mode, where displacements are equal to zero along the neutral line within the triple of rigid pins, is unaltered for the case of a horizontal shift of the middle grating; this is no longer the case for the situation when the middle grating consists of rigid inclusions of finite radius.

In Sect. 2 we give a brief account of the governing equations and the method of solution for the problem of scattering of a flexural wave by a finite set of gratings of rigid inclusions of common period. (For the convenience of the reader, all the key equations are given in the Supplementary Material to the paper.) The studies that follow Sect. 2 address interaction of scatterers of finite radius. We show that the EDIT phenomenon can still be obtained for the configuration in which the central grating has rigid pins and the outer gratings have finite radius scatterers. Parameters of the system which enable the quality factors of transmission and reflection resonances to be controlled are identified and flexural wave patterns within the grating stack are exhibited.

2 Governing equations: scattering by a single grating

This paper presents a considerable breakthrough compared to the two previous publications by Movchan et al. (2009) and Haslinger et al. (2012). The novel

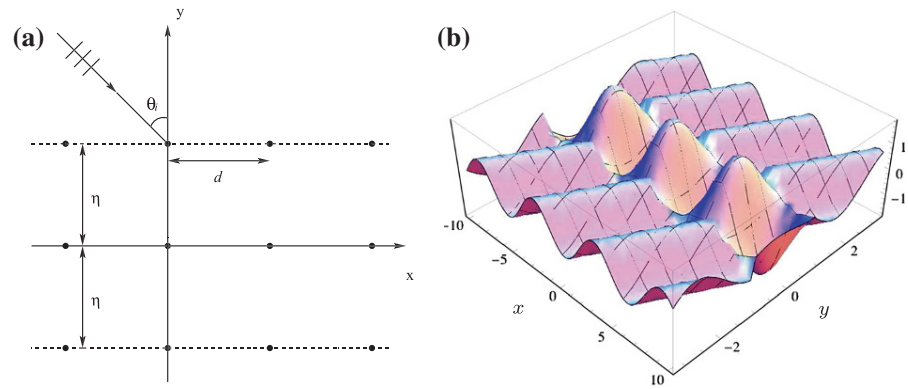

Fig. 1 **a** Triplet of aligned rigid pin gratings with period d and relative grating separation $\eta = d$. The angle of incidence is denoted by θ_i. **b** Flexural displacement plotted as a function of x and y for the transmission resonance associated with a symmetric trapped mode for $\theta_i = 30°$ for the grating stack illustrated in (a)

Elasto-Dynamically Inhibited Transmission phenomenon has never been observed in elasticity problems, but its analogues are common in electro-magnetism. This is essential to emphasise the importance of the defect modes for the case of stacks of gratings containing rigid inclusions of finite radius a rather than fixed pins. This makes it necessary to take into account higher-order multipole terms characterising the scattered field, and the periodicity of the grating will also lead to the use of higher order lattice sums. The account of the theory given below and in the Supplementary Material is kept to the bare minimum and further details can be found in the aforementioned papers. (Note that references in the text to equations from the Supplementary Material have numbers with a prefix S.)

The flexural displacement $w(\mathbf{x}; t) = W(\mathbf{x}) \exp(i\omega t)$ is considered to be time-harmonic of radian frequency ω, and its amplitude W satisfies the following equation

$$\Delta^2 W(\mathbf{x}) - \beta^4 W(\mathbf{x}) = 0, \tag{1}$$

with $\beta^2 = \omega\sqrt{\rho h/D}$. We also note that the flexural rigidity of the plate is denoted by $D = Eh^3/(12(1-\nu^2))$, where h stands for the thickness of the plate, ρ is the mass density, E is the Young modulus and ν is the Poisson ratio.

The solution of Eq. (1) can be divided into two parts W_H and W_M, which satisfy the Helmholtz equation and its counterpart form which we call the modified Helmholtz equation:

$$\left(\Delta + \beta^2\right) W_H = 0 \quad \text{and} \quad \left(\Delta - \beta^2\right) W_M = 0. \tag{2}$$

An important aspect of the physics of the problem is that W_H contains both propagating and evanescent waves, while W_M consists entirely of evanescent waves.

Figure 2 shows a sample configuration, which will be of considerable interest in the paper, and serves to

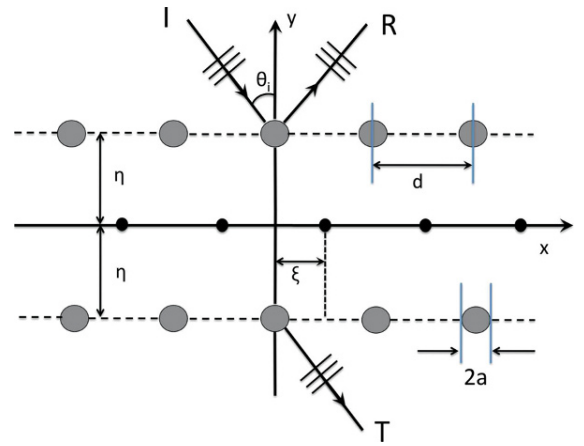

Fig. 2 Stack of gratings consisting of an outer pair of finite nonzero inclusions with radius a and period d and a central grating of rigid pins characterised by the relative lateral shift ξ. The relative grating separation between consecutive gratings is denoted by η

introduce important parameters characterising the grating structure and the incident wave. The stack of gratings comprises two or three elements, each assumed to be periodic, of period d, in the horizontal direction, and typically the gratings are separated by the common distance η in the vertical direction. The successive gratings may have a relative horizontal shift, denoted by ξ in Fig. 2. The stack may include a combination of fixed pin gratings and gratings of rigid cylindrical inclusions of finite radius a.

For each rigid inclusion, the boundary conditions are

$$W\Big|_{r=a} = 0, \quad \frac{\partial W}{\partial r}\Big|_{r=a} = 0, \tag{3}$$

where r represents the distance from the centre of the inclusion. Physically the above boundary conditions

represent clamping on the rigid boundary. In particular, when the radius a tends to zero we retrieve the case of fixed pins discussed by Movchan et al. (2009) and Haslinger et al. (2012).

A plane incident wave is propagating towards the grating in the upper half-plane. We need to consider two cases, the first dealing with an incident wave of the Helmholtz type, with amplitude $\mathcal{A}_H$:

$$W_{i,H}(\mathbf{x}) = \frac{\mathcal{A}_H}{\sqrt{|\chi_0|}} \exp\{\mathrm{i}(\alpha_0 x - \chi_0 y)\}, \tag{4}$$

where $\alpha_0^2 + \chi_0^2 = \beta^2$ and α_0 is the Bloch parameter, $\alpha_0 = \beta \sin\theta_i$, with θ_i being the angle of incidence (see Fig. 2). The second case treats an incident wave of modified Helmholtz type, with amplitude $\mathcal{A}_M$:

$$W_{i,M}(\mathbf{x}) = \frac{\mathcal{A}_M}{\sqrt{|\hat{\chi}_0|}} \exp\{\mathrm{i}(\alpha_0 x - \hat{\chi}_0 y)\}, \tag{5}$$

where this wave is always evanescent so that $\alpha_0^2 + \hat{\chi}_0^2 = -\beta^2$, $\hat{\chi}_0 = i\tau_0$, $\tau_0 > 0$.

The periodicity of the structure in the horizontal direction and the parameters of the incident wave are represented by the quasi-periodicity condition for W along the horizontal axis:

$$W\left(\mathbf{x} + md\mathbf{e}^{(1)}\right) = W(\mathbf{x})\, e^{i\alpha_0 md}, \tag{6}$$

where $m \in \mathbf{Z}$ and d is the period.

We now sketch the method used to treat the scattering of the plane incident waves, either of the Helmholtz type (4) or of the modified Helmholtz type (5), by a grating of inclusions of radius a. The flexural displacement W can be expanded for $y > a$ and $y < -a$ in terms of sums of plane waves (W_H) and modified plane waves (W_M). Above the grating the expansion has a down-going incident wave term and up-going reflected waves with amplitudes R_p and $\hat{R}_p$, for the respective wave types. Below the grating the amplitudes of the down-going transmitted waves are denoted by T_p and $\hat{T}_p$.

In order to connect these two types of expansions we introduce multipole expressions for W in the region $-a \le y \le a$. The multipole expansion for W_H involves cylindrical waves $J_n(\beta r)e^{\mathrm{i}n\theta}$ and $H_n^{(1)}(\beta r)e^{\mathrm{i}n\theta}$ with respective amplitudes A_n and E_n:

$$W_H(\mathbf{x}) = \sum_{n=-\infty}^{\infty} \left\{A_n J_n(\beta r) + E_n H_n^{(1)}(\beta r)\right\} e^{in\theta}. \tag{7}$$

The multipole expansion for W_M involves modified Bessel function terms $I_n(\beta r)e^{\mathrm{i}n\theta}$ and $K_n(\beta r)e^{\mathrm{i}n\theta}$ with respective amplitudes B_n and F_n:

$$W_M(\mathbf{x}) = \sum_{n=-\infty}^{\infty} \{B_n I_n(\beta r) + F_n K_n(\beta r)\} e^{in\theta}. \tag{8}$$

The amplitudes A_n, B_n, E_n, F_n are the multipole coefficients to be determined, and they are related by the boundary conditions (3). The coefficients A_n and B_n are expressed in terms of the coefficients E_n and F_n (see equations (S7), (S8)).

The next step in the argument is to write down the Rayleigh identity, which expresses the part of the expansion for W which is regular near the origin (i.e. the terms involving the Bessel functions J_n and I_n) as sums over the part which is irregular near the origin (i.e. the terms involving the Bessel functions $H_n^{(1)}$ and K_n), together with a term representing the expansion of the incident wave in multipoles. While we refer to Movchan et al. (2007), here the Rayleigh equations are written with the use of the appropriate grating lattice sums in place of the doubly periodic array lattice sums. Equations (S11), (S12) in the Supplementary Material give the connection between the two sets of multipole coefficients A_n, B_n and E_n, F_n. These equations are uncoupled for solutions of the Helmholtz and the modified Helmholtz type. The necessary coupling between the two types of waves is provided by the boundary conditions. The combination of the Rayleigh identity and the boundary conditions gives a system of linear equations (S13), (S14), which is then truncated (e.g. so that Bessel functions of orders $-L$ to L are retained) and solved to evaluate a set of multipole coefficients. These in turn are used to evaluate plane wave amplitude coefficients for the fields above and below the grating, using what we call reconstruction equations (S25)–(S28). The amplitudes of the reflected and transmitted waves are determined for the set of incident fields corresponding to a range of grating orders, both propagating and evanescent. These are assembled into scattering matrices for reflection and transmission that completely characterise the grating's scattering action.

3 Transmission resonance for pairs of aligned gratings: normal incidence

Here we consider the filtering action in the transmission of flexural waves by systems of two gratings of inclusions, with both of the gratings aligned. We concentrate on the effects of inclusion radius a and vertical separation η between gratings on the filtering action,

characterised by the Q-factor[1] of the transmittance peak.

The phenomenon of transmission resonance for gratings of rigid pins was addressed by Movchan et al. (2009) and Haslinger et al. (2012). The increase of the radius of inclusions within the gratings affects the scattered fields; this consequently leads to the change in the frequency of the resonance mode as well as its Q-factor. The analytical representation of the physical fields between the gratings incorporates multipole representations (7) and (8) outlined in Sect. 2.

The formulae for transmission and reflection matrices were already obtained for a single grating, together with the formulae for the transmission and reflection matrices for a pair of identical gratings (see Movchan et al. 2009)

$$\boldsymbol{\mathcal{T}}_2 = \boldsymbol{\mathcal{T}}_1[\mathbf{I} - (\boldsymbol{\mathcal{R}}_1)^2]^{-1}\boldsymbol{\mathcal{T}}_1,$$
$$\boldsymbol{\mathcal{R}}_2 = \boldsymbol{\mathcal{R}}_1 + \boldsymbol{\mathcal{T}}_1\boldsymbol{\mathcal{R}}_1[\mathbf{I} - (\boldsymbol{\mathcal{R}}_1)^2]^{-1}\boldsymbol{\mathcal{T}}_1. \quad (9)$$

The matrices $\boldsymbol{\mathcal{T}}_i, \boldsymbol{\mathcal{R}}_i$ comprise the matrices $\mathbf{T}_i$ and $\mathbf{R}_i$ respectively, where $i = 1, 2$, which are formed using the coefficients for the plane wave representations (S25)-(S28), together with a diagonal propagation matrix $\boldsymbol{\mathcal{P}}$:

$$\boldsymbol{\mathcal{R}}_i = \boldsymbol{\mathcal{P}}\mathbf{R}_i\boldsymbol{\mathcal{P}}, \ \boldsymbol{\mathcal{T}}_i = \boldsymbol{\mathcal{P}}\mathbf{T}_i\boldsymbol{\mathcal{P}}. \quad (10)$$

For the unshifted pair, this propagation matrix $\boldsymbol{\mathcal{P}}$ changes with the change of the vertical separation η between the gratings,

$$\boldsymbol{\mathcal{P}} = \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}, \quad \text{where} \quad P = [\delta_{tp} e^{i\tilde{\chi}_p \eta/2}], \quad (11)$$

with $\tilde{\chi}_p = \chi_p$ if p corresponds to a Helmholtz type plane wave and $\tilde{\chi}_p = \hat{\chi}_p$ if p corresponds to a plane wave of modified Helmholtz type. In contrast, the scattering matrices $\mathbf{T}_i$ and $\mathbf{R}_i$ change with β.

For the purpose of illustration, we consider several configurations. A symmetric pair of gratings of inclusions of finite radius is shown in Fig. 3. The geometrical dimensions are normalised so that the distance d between the centres of inclusions is unity. Table 1 gives a summary of the resonance frequencies for different configurations for the case of normal incidence.

The resonant frequency corresponding to the peak in transmission is denoted by β_*. For very high Q resonances, additional decimal places are required to evaluate the spectral parameter β_* because the sampling, needed to identify the characteristic values of the spectral parameter, must be much finer. Throughout the paper, appropriate accuracy is used in the Tables for resonant frequencies. The first line in Table 1 corresponds to the pair of gratings of rigid pins.

[1] For a resonant peak of transmittance T occurring at $\beta = \beta_*$ with $T = T_{\max}$ there, if $T = T_{\max}/2$ for $\beta = \beta_+$ and β_-, then $Q = \beta_*/|\beta_+ - \beta_-|$.

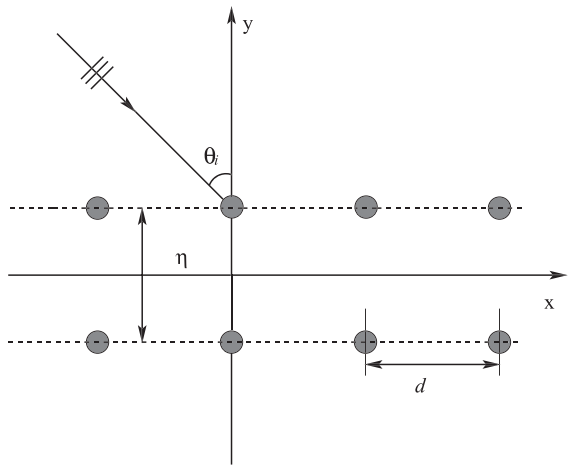

Fig. 3 Pair of gratings consisting of finite radius inclusions with period d and relative grating separation η

Table 1 Resonant frequencies β_* and the corresponding Q-factors for various pairs of unshifted gratings for normal incidence

a	η	Resonant frequency	Q
0	d	$\beta_* = 3.62810$	20.0
0	$2d$	$\beta_*^{(1)} = 1.95200$	6.19
		$\beta_*^{(2)} = 3.41020$	25.2
		$\beta_*^{(3)} = 4.63308$	539
$0.01d$	d	$\beta_* = 3.82890$	208
$0.01d$	$2d$	$\beta_*^{(1)} = 2.13669$	33.4
		$\beta_*^{(2)} = 3.53658$	138
		$\beta_*^{(3)} = 4.72190$	76.6
$0.02d$	d	$\beta_* = 3.90296$	477
$0.02d$	$2d$	$\beta_*^{(1)} = 2.17529$	50.8
		$\beta_*^{(2)} = 3.57980$	219
		$\beta_*^{(3)} = 4.76865$	68.2
$0.1d$	d	$\beta_* = 4.5572187$	6.51×10^4
$0.1d$	$2d$	$\beta_*^{(1)} = 2.40631$	617
		$\beta_*^{(2)} = 3.925255$	1.19×10^3
		$\beta_*^{(3)} = 5.23749$	202

In this table, $d = 1$

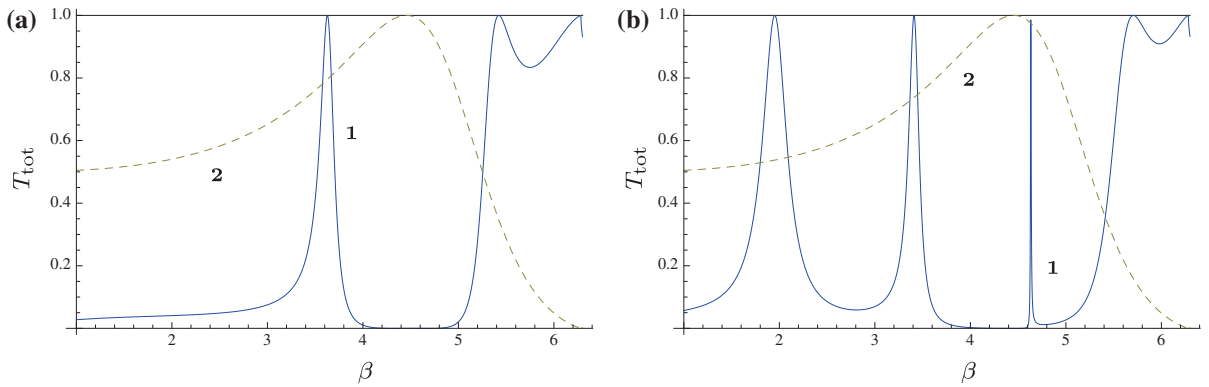

Fig. 4 Total transmittance $T_{\rm tot}$ (curve **1**) for a pair of rigid pin gratings and total reflectance $R_{\rm tot}$ (curve **2**) for a single grating of rigid pins as functions of β. Data used: **a** $\eta/d = 1$, **b** $\eta/d = 2$. Here $L = 0$

Throughout this paper we consider the range of β values for which only the zeroth order plane waves in reflection and transmission propagate i.e. $p = 0$. For rigid zero-radius pins, it is sufficient for the corresponding order of truncation L to be zero. Accordingly, the total energy transmitted, $T_{\rm tot}$, is given by $|T_0|^2$, $|T_0|$ being the amplitude of the zeroth order transmission coefficient (see equation (S26)).[2] However as the radius of the inclusion is increased, more terms are required in the grating sums thereby increasing L.

The normalised resonant frequency is $\beta_* = 3.62810$ for the transmission peak with Q-factor 20, as illustrated by curve **1** in Fig. 4a. Curve **2** shows the reflected energy for a single grating of rigid pins. It is clear that there is a significant difference between the transmission properties of the single grating and of the pair of gratings. The main emphasis is on the resonance mode which arises due to the interaction between the gratings, leading to a peak in transmission.

The increase in the vertical separation may bring additional resonances. This is illustrated in Fig. 4b for the vertical separation $2d$. We observe the peak at $\beta_* = 3.41020$ (rather than at $\beta_* = 3.62810$ as for the vertical separation $\eta = d$) and two additional peaks, one at a much lower frequency of $\beta_* = 1.95200$, and the other at a higher frequency ($\beta_* = 4.633083$) and with a significantly higher Q-factor (see Table 1). The reason for the very sharp peak at $\beta_* = 4.633083$ is its proximity in frequency to very high reflectance for a single grating of rigid pins for $\theta_i = 0°$. Figure 4b illustrates two requirements for delivering high Q-factor transmission peaks: the first is that the reflectance of a single grating has to be close to unity, and the second is that the separation of the gratings has to be chosen to align the interference peak with the point of high reflectance. We note that the physics here for flexural waves is in keeping with that for optical waves (see Born and Wolf 1959, Section 7.6).

For a grating of inclusions of radius $a = 0.1d$, Fig. 5a shows that the transmittance resonance for a pair of gratings lies in the region of β values where the reflectance of a single grating is very close to unity. This results in a high Q-factor (6.50×10^4) for the pair of such gratings. By comparison, for a pair of gratings with inclusions of radius $a = 0.01d$ (Fig. 5b), the Q-factor is 208, with the peak in transmittance for a pair being significantly separated from the reflectance peak for a single grating.

4 Transmission resonance for pairs of aligned gratings: oblique incidence

With the introduction of nonzero angle of incidence, additional physical effects become apparent. Firstly, from Eqs. (11) and (S3) it follows that, if η is fixed and the angle of incidence θ_i increases, then χ_0 decreases and thus the resonant value β_* must increase to preserve the resonance transmission condition. Secondly, the range of β values in which only the zeroth order of diffraction propagate changes according to the equation

$$\beta_{-1} = \frac{2\pi}{d(1 + \sin\theta_i)}, \tag{12}$$

[2] Similarly, the total reflectance $R_{\rm tot}$ is equal to $|R_0|^2$, with $|R_0|$ being the amplitude of the zeroth order reflection coefficient (see equation (S25)).

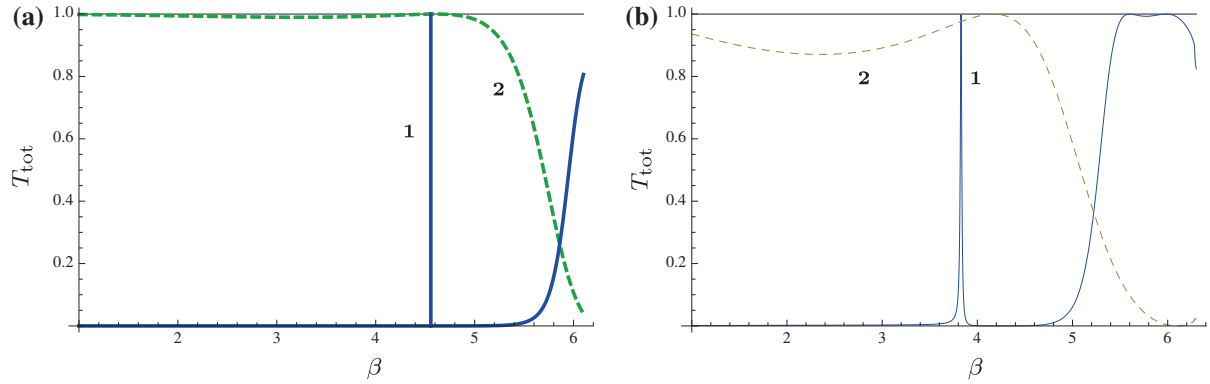

Fig. 5 Total transmittance $T_{\rm tot}$ (curve **1**) for a pair of gratings of inclusions of radius a and total reflectance $R_{\rm tot}$ (curve **2**) for a single grating of inclusions as functions of β for separation $\eta/d = 1$, and $L = 2$. Data used: **a** $a = 0.1d$, **b** $a = 0.01d$

Table 2 Resonant frequencies β_* and the corresponding Q-factors for various pairs of unshifted gratings for examples of oblique incidence

θ_i	a	η	Resonant frequency	Q
30°	0	d	$\beta_* = 3.58221$	5.42×10^3
30°	0	$2d$	$\beta_*^{(1)} = 3.6174737$	3.55×10^4
			$\beta_*^{(2)} = 2.15950$	10.4
30°	$0.01d$	d	$\beta_* = 3.86164$	238
30°	$0.05d$	d	$\beta_* = 4.18486$	253
20°	$0.01d$	d	$\beta_* = 3.8507342$	2.23×10^4
20°	$0.05d$	d	$\beta_* = 4.1606161$	2.57×10^3
15°	$0.1d$	$2d$	$\beta_*^{(1)} = 4.011852$	1.75×10^3
			$\beta_*^{(2)} = 2.45092$	639
10°	$0.1d$	$2d$	$\beta_*^{(1)} = 3.963273$	1.40×10^3
			$\beta_*^{(2)} = 2.42564$	626

In this table $d = 1$

with β required to be smaller than β_{-1}. Thirdly, the angle of incidence θ_i may be used to bring the resonant frequency β_* for a pair of gratings into proximity with the value for peak reflectance of a single grating.

Table 2 presents the results of computations of the resonance frequencies in transmission and of the quality factors for pairs of aligned gratings (of zero shift).[3]

In Fig. 6 we give the transmittance curves for a pair of rigid pin gratings with the separation being $\eta/d = 2$ and with the angle of incidence $\theta_i = 30°$, with corresponding curves for $\eta/d = 1$ being given in Fig. 2 of Haslinger et al. (2012). With the doubling of the grating separation, the Q-factor goes up from 5.42×10^3 to 3.55×10^4. Note the very strong contrast in the quality factors of the two transmission resonances in Fig. 6a, caused by the difference in the reflectance values of the single grating.

For the particular inclusion radius $a = 0.1d$ we show in Fig. 7 reflectance curves for a single grating for angles of incidence ranging from 0° to 30°. The possible frequency range, for which the interference peak may be aligned with the almost total reflectance, is $4.4 < \beta < 4.7$ for $0° \le \theta_i \le 15°$. Note that as the angle of incidence increases the reflectance peak moves closer and closer to the position of the Wood anomaly whose frequency is given by Eq. (12). Since this Wood anomaly marks the boundary of the region where two orders propagate in both reflection and transmission, it then becomes more and more difficult to choose parameters giving a sharp transmission resonance for a pair of gratings.

5 Controlling transmission resonances for triplets of gratings

In previous sections we have considered stacks in which all gratings were aligned. In this section we generalise the discussion to non-aligned stacks. The recurrence procedure of Sect. 3 is adapted to this latter case if we replace Eq. (10) by

$$\mathcal{R}_i = \bar{\mathcal{Q}}\mathcal{P}\mathbf{R}_i\mathcal{P}\bar{\mathcal{Q}}, \quad \mathcal{T}_i = \bar{\mathcal{Q}}\mathcal{P}\mathbf{T}_i\mathcal{P}\bar{\mathcal{Q}}, \tag{13}$$

[3] For rigid pin geometries, we use a more efficient method to evaluate resonant frequencies than Movchan et al. (2009) and Haslinger et al. (2012).

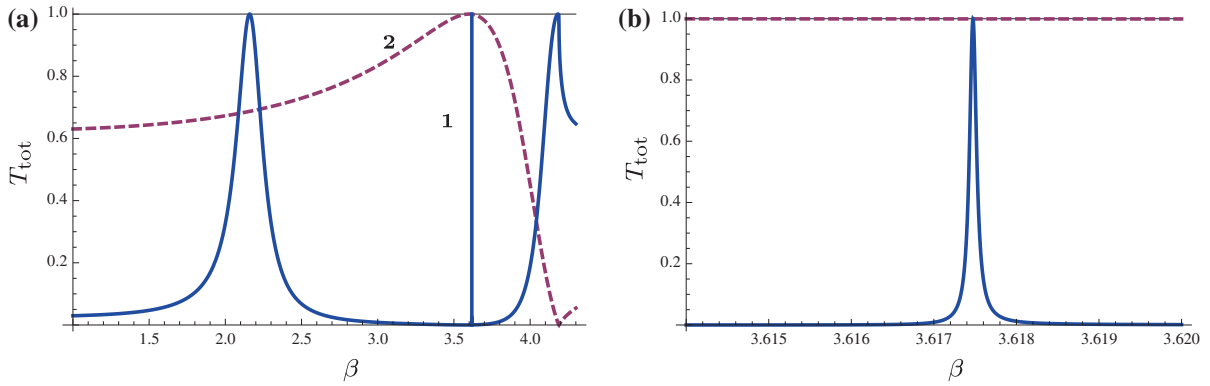

Fig. 6 Total transmittance $T_{\rm tot}$ (curve **1**) for a pair of rigid pin gratings and total reflectance $R_{\rm tot}$ (curve **2**) for a single grating of rigid pins as functions of β. Data used: $\eta/d = 2$, $\theta_i = 30°$, and $L = 0$. The diagram (**b**) is a blow up of the sharp transmission resonance from (**a**)

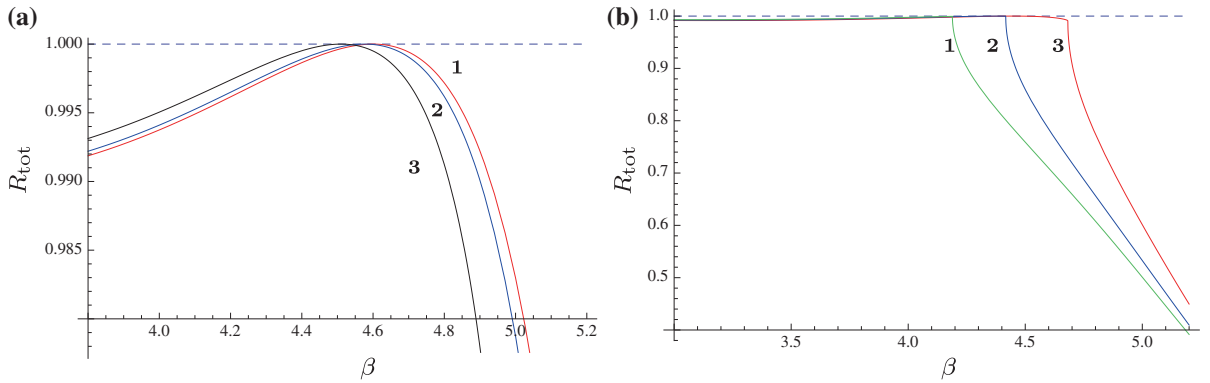

Fig. 7 Total reflectance $R_{\rm tot}$ as a function of β for a single grating with inclusions of radius $a = 0.1d$ (with order of truncation $L = 2$) for the angle of incidence **a** $\theta_i = 0°$ (curve **1**), 7.5° (curve **2**), 15° (curve **3**) and **b** $\theta_i = 30°$ (curve **1**), 25° (curve **2**), $\theta_i = 20°$ (curve **3**)

where $\mathcal{P}$ is as defined in (11) and

$$\mathcal{Q} = \begin{pmatrix} Q_\xi & 0 \\ 0 & Q_\xi \end{pmatrix}$$

$$\text{with} \quad Q_\xi = [\delta_{tp} e^{-i\alpha_p \xi/2}], \quad \alpha_p = \alpha_0 + \frac{2\pi p}{d}. \tag{14}$$

Three types of triplets are considered here:

- triplets of identical gratings consisting of inclusions with the same nonzero radius;
- triplets consisting of an outer pair of gratings of rigid pins with a central grating of inclusions with a nonzero radius;
- triplets with the central grating consisting of rigid pins surrounded by a pair of gratings with inclusions of nonzero radius.

The corresponding geometries are illustrated in Fig. 8.

5.1 Triplets of identical gratings

The results of computations of the transmitted energy, in normal incidence, for the triplet of gratings of small inclusions of radius $a = 0.02d$, are shown in Fig. 9. We observe a double peak for normal incidence. These double peaks are characterised by symmetric and anti-symmetric vibration modes, so that the flexural wave

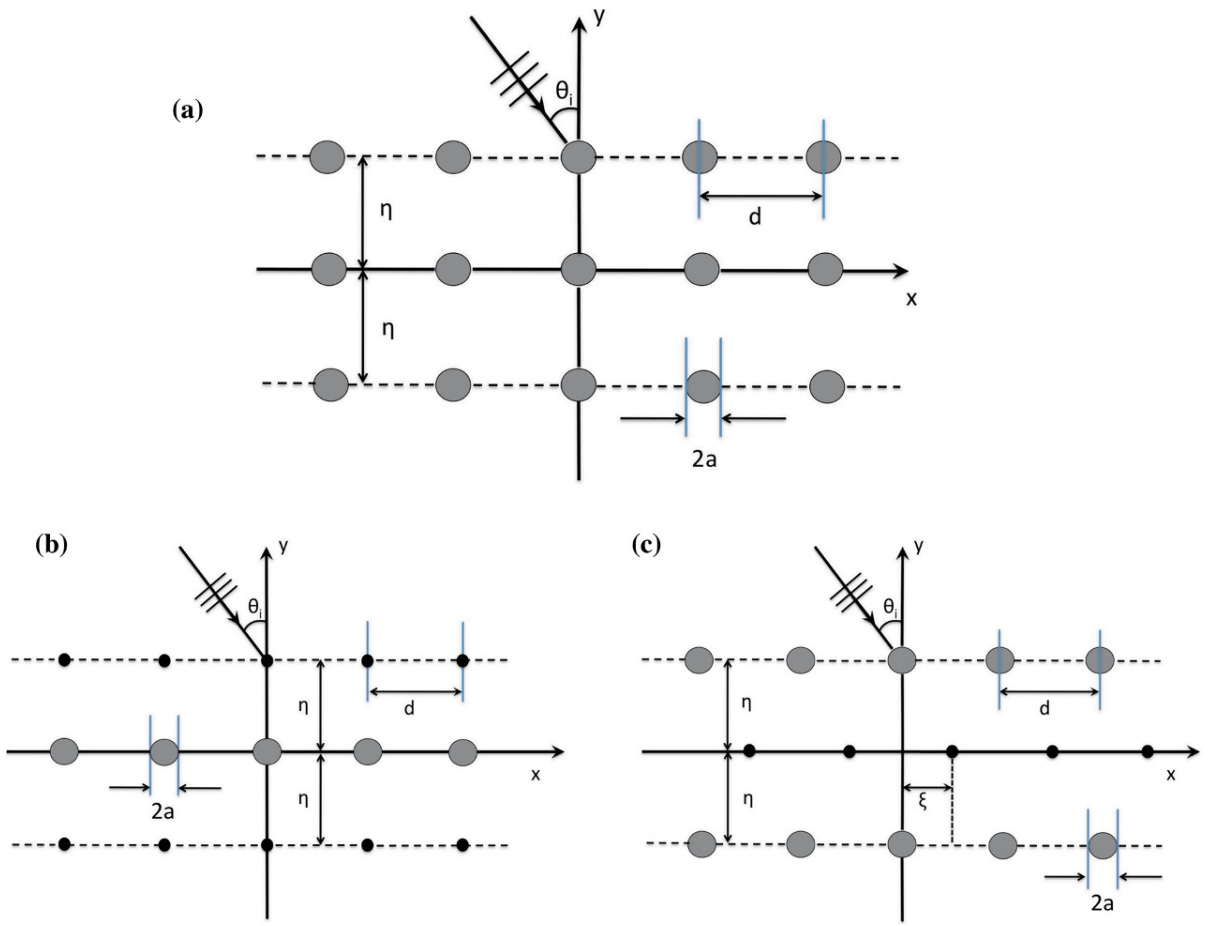

Fig. 8 **a** Triplet consisting of three identical gratings of inclusions with nonzero radius a; **b** triplet consisting of a central grating of inclusions of nonzero radius a, surrounded by a pair of rigid pin gratings; **c** triplet consisting of a pair of gratings with inclusions of nonzero radius a surrounding a grating of rigid pins

within the triplet is an even or odd function of y, respectively. A shift of the central grating results in a perturbative effect on the frequency for which enhanced transmission is observed. This shift in the double peak's frequency is illustrated in Fig. 9b where the central grating is shifted by $\xi = 0.2d$ relative to the outer pair. Although the change in the resonance frequencies in the diagrams (a) and (b) appears to be small, we note that the anti-symmetric mode depends on the longitudinal shift for the case of inclusions of nonzero radius.

As a consequence of the symmetry of the grating stack, we deduce that the resonant flexural vibration modes are either symmetric or anti-symmetric. The flexural displacement for frequencies $\beta_*^{(2)} = 3.93753$ and $\beta_*^{(2)} = 3.948562$ for the right hand peaks in Fig. 9 correspond to anti-symmetric flexural modes, whereas the other two frequencies $\beta_*^{(1)} = 3.857309$ and $\beta_*^{(1)} = 3.863065$ correspond to symmetric flexural modes.

The results given here for inclusions of nonzero radius differ from those obtained by Haslinger et al. (2012) for triplets of rigid pins. One of the main differences is related to anti-symmetric flexural modes, which vanish on the central line of the grating stack. The frequency of the transmission resonance for anti-symmetric vibrations within the triplet of gratings of rigid pins is invariant with respect to a horizontal shift of the middle grating. This feature disappears when the radius of inclusions is nonzero.

These comments are exemplified in the field plots of Fig. 10. For a triplet of unshifted gratings of radius $a = 0.01d$ we identify two resonance transmission

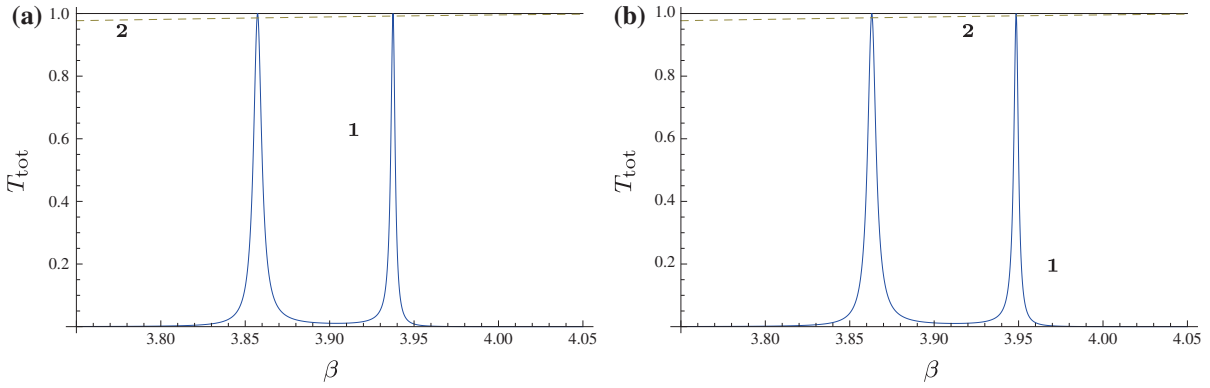

Fig. 9 Total transmittance $T_{\rm tot}$ (curve **1**) as a function of β for normal incidence for a triplet of gratings of inclusions with $a = 0.02d, \eta = d = 1$ and $L = 2$. Curve **2** represents total reflectance $R_{\rm tot}$ for a single grating of such scatterers: **a** aligned gratings, $\xi = 0$; **b** shifted central grating with $\xi = 0.2d$

peaks for the angle of incidence $\theta_i = 30°$, as illustrated in Fig. 10. It is noted that the symmetric mode corresponds to a lower value of $\beta_*^{(1)} = 3.819908$ than the anti-symmetric mode, which occurs at $\beta_*^{(2)} = 3.93678$. While plane waves outside the grating stack have the orientation of the incident wave, inside the stack the nodes and anti-nodes of the flexural displacement lie broadly parallel to the y-axis.

5.2 Non-uniform triplets, with rigid pins on the exterior boundary

We consider the triplet configuration shown in Fig. 8b, where all three gratings are aligned with each other (i.e. $\xi = 0$). The triplet has rigid pin gratings on the outer boundaries, and a grating of finite dimension inclusions in the middle. In order for this configuration to deliver sharp transmission resonances in normal incidence it is necessary for the radius of inclusions in the middle grating to be around $0.20d$ or larger. This is evident from Fig. 11 where transmittance curves are compared for radii $a = 0, a = 0.15d$ and $a = 0.20d$. The transmission peaks for the rigid pin triplet in part (a) are far from being aligned with the maximal reflectance for the single grating in normal incidence. To align the peaks, and thereby increase the quality factors of the resonances, the radius of the inclusions is increased to $a = 0.15d$ in part (b), and $a = 0.20d$ in part (c). The data for resonant frequencies and Q-factors for these and other radii, together with data for different angles of incidence, are given in Table 3. Of the two sharp peaks in Fig. 11c that at $\beta_*^{(1)} = 4.42186394$ has a Q-factor of 3.42×10^4 and corresponds to an anti-symmetric flexural mode. The other peak at $\beta_*^{(2)} = 4.445547468$ has the very high Q-factor of 7.68×10^5 and corresponds to a symmetric flexural mode.

High quality factor transmission resonances may be achieved either by specifying the angle of incidence and then tuning the radius or by varying the angle of incidence for a fixed radius. Examples of this can be seen in Table 3. In Fig. 12 we show field plots of the flexural displacement for the third case of Table 3. The anti-symmetric mode $\beta_*^{(1)} = 3.9995073$ in this case has a higher Q-factor (7.89×10^4) than the symmetric mode ($\beta_*^{(2)} = 4.0416640$, $Q = 4.48 \times 10^3$).

5.3 Non-uniform triplets, with rigid pins in the central grating

Triplets with a central grating of rigid pins have the important property that the frequency of the transmission resonance associated with the anti-symmetric mode is invariant with respect to the horizontal shift of the central grating. As the resonant frequency of the symmetric mode is affected by such horizontal shifts, they offer a means of controlling the frequency gap between these modes. We illustrate this property by studying the relationship between the transmission resonances of grating triplets and those for the outer pair of gratings.

In Fig. 13 we show total transmittance curves for an aligned triplet with the outer gratings having $a = 0.1d$ (part (a)) and for the structure with the central grating

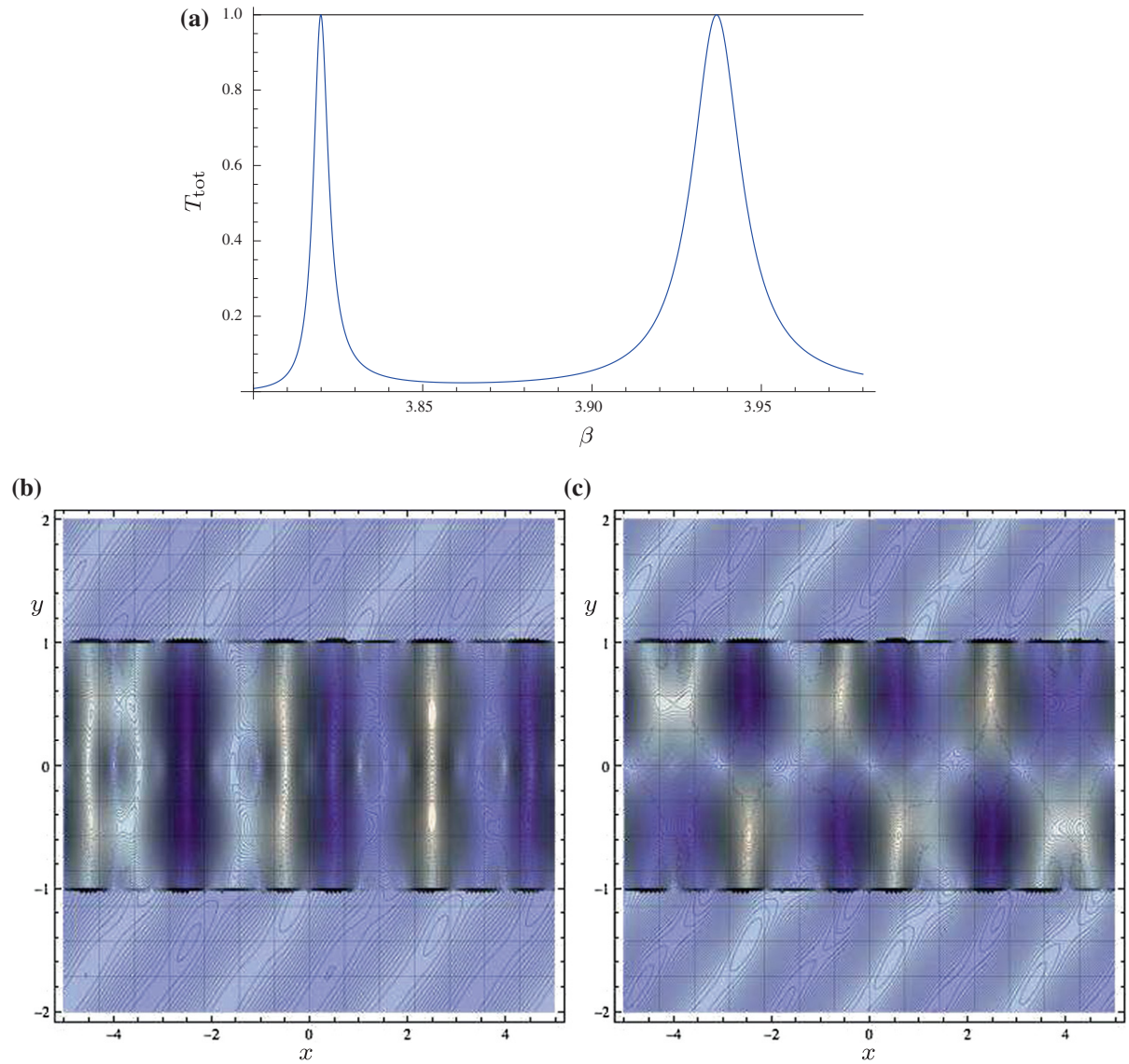

Fig. 10 **a** Total transmittance $T_{\rm tot}$ as a function of β for a triplet of unshifted gratings of rigid inclusions of radius $a = 0.01d$, $\eta = d = 1$, for $\theta_i = 30°$. Field plots for transmission resonances: **b** a symmetric mode with $\beta_*^{(1)} = 3.819908$; **c** an anti-symmetric mode with $\beta_*^{(2)} = 3.93678$

of rigid pins removed (part (b)). The transmission resonance shown in part (b) occurs at $\beta_* = 4.011852$ and it coincides with the leftmost peak in part (a). The flexural displacement plotted in Fig. 14a for this mode confirms its anti-symmetric nature and shows also its nodal line, coinciding with the central grating. The second transmission peak in Fig.13a, of course, corresponds to a symmetric mode, as confirmed in Fig.14b.

Figure 15 illustrates two different circumstances which can arise when the central grating of pins is shifted with respect to the outer gratings of inclusions with nonzero radius. In case (a) the angle of incidence is $\theta_i = 15°$ and the radius of inclusions is $a = 0.1d$. The anti-symmetric transmission resonance (peak **1**) occurs at $\beta_* = 4.011852$ for all values of $\xi \in [-0.5d, 0.5d]$. The symmetric mode resonance value is an even function of ξ which moves to higher values of β as ξ increases from 0 to $0.5d$. (Peaks **2** and **3** correspond respectively to $\xi = 0.25d$ and $\xi = 0.5d$.) Our calculations show that, for this system, the positions of

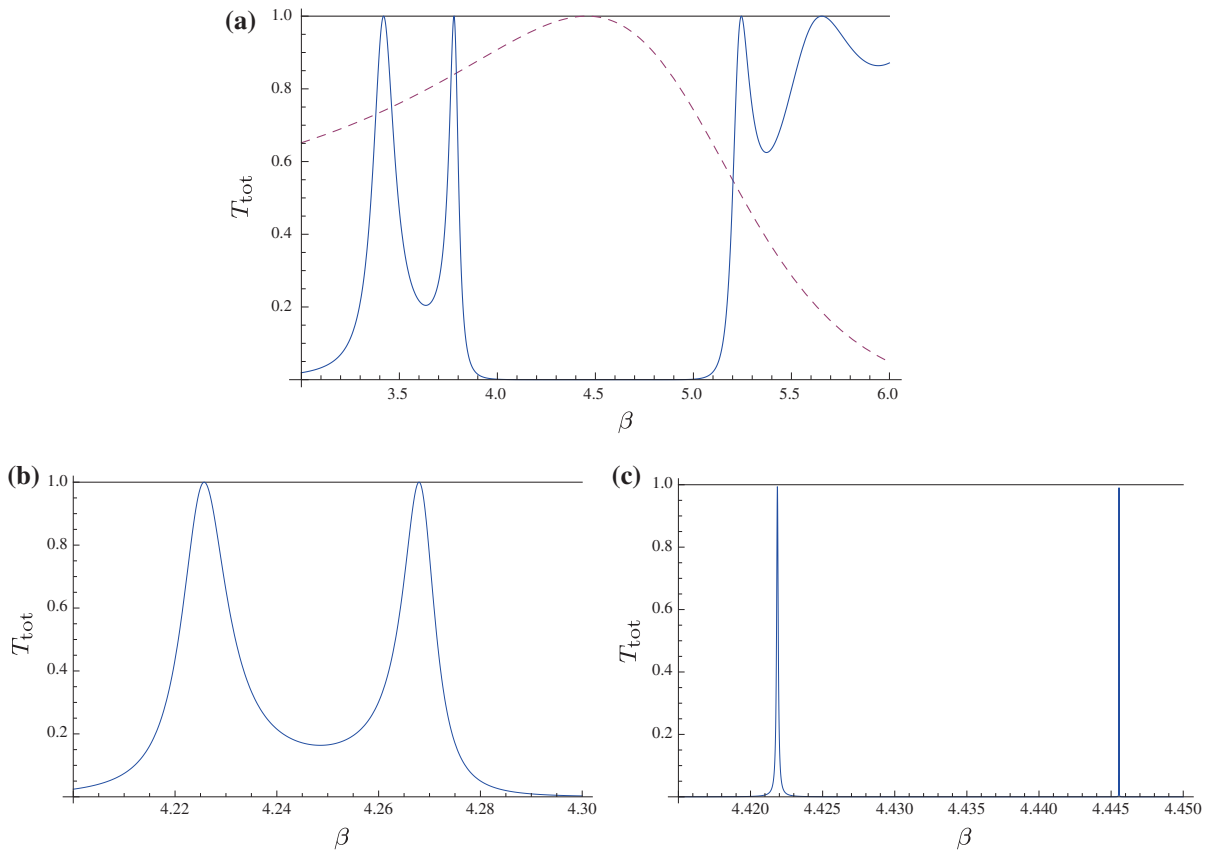

Fig. 11 Total transmittance $T_{\rm tot}$ as a function of β for normal incidence for a symmetric triplet with **a** $a = 0$ (total reflectance $R_{\rm tot}$ for a single grating of zero-radius scatterers is shown by the dashed curve) **b** $a = 0.15d$ and **c** $a = 0.20d$ for the central grating. Here $\xi = 0$, $\eta = d = 1$, and $L = 2$

Table 3 Resonant frequencies β_* and the corresponding quality factors for triplets of gratings with a middle grating consisting of finite radius inclusions, and an outer pair comprising rigid pins

θ_i	a	η	ξ	Resonant Frequency	Q
0°	$0.2d$	d	0	$\beta_*^{(1)} = 4.42186394$	3.42×10^4
				$\beta_*^{(2)} = 4.445547468$	7.68×10^5
0°	$0.15d$	d	0	$\beta_*^{(1)} = 4.22572$	377
				$\beta_*^{(2)} = 4.26794$	521
18°	$0.1d$	d	0	$\beta_*^{(1)} = 3.9995073$	7.89×10^4
				$\beta_*^{(2)} = 4.0416640$	4.48×10^3
18°	$0.1d$	d	$0.5d$	$\beta_*^{(1)} = 4.06325990$	4.79×10^4
				$\beta_*^{(2)} = 4.10531818$	1.99×10^4
25°	$0.01d$	d	0	$\beta_*^{(1)} = 3.72273076$	6.03×10^5
				$\beta_*^{(2)} = 3.7251407$	1.24×10^4

The relative lateral shift of the central grating is denoted by ξ and the relative vertical separation between successive gratings is denoted by η

the anti-symmetric and symmetric resonance peaks do not coincide for any values of ξ in this range. However, for the angle of incidence of $\theta_i = 20°$ and the inclusion radius of $a = 0.085d$, Fig. 15b shows that the symmetric transmission resonance peaks for $\xi = 0.25d$ (peak **2**) and $\xi = 0.5d$ (peak **3**) bracket the position of the anti-symmetric resonance (peak **1**). Therefore, the symmetric and anti-symmetric resonances can be brought into a coincidence for a value of ξ in the interval between $0.25d$ and $0.5d$.

6 Shifted systems of gratings and the EDIT phenomenon

The phenomenon of *Elasto-Dynamically Inhibited Transparency* (EDIT), which can be demonstrated via the analysis of resonance vibration modes within triplets of periodic gratings, was identified by Haslinger

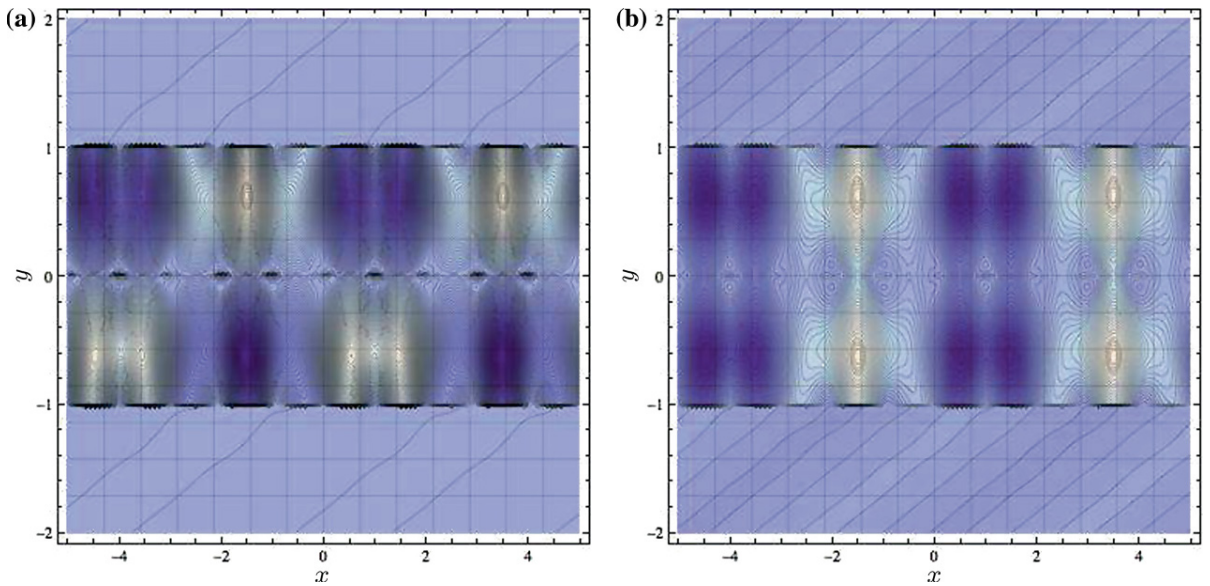

Fig. 12 Flexural displacement as a function of x and y for **a** an anti-symmetric mode ($\beta_* = 3.9995073$) and **b** a symmetric mode ($\beta_* = 4.041664$). Data: central grating inclusions of radii $a = 0.1d$, $\theta_i = 18°$ and $\eta = d = 1$

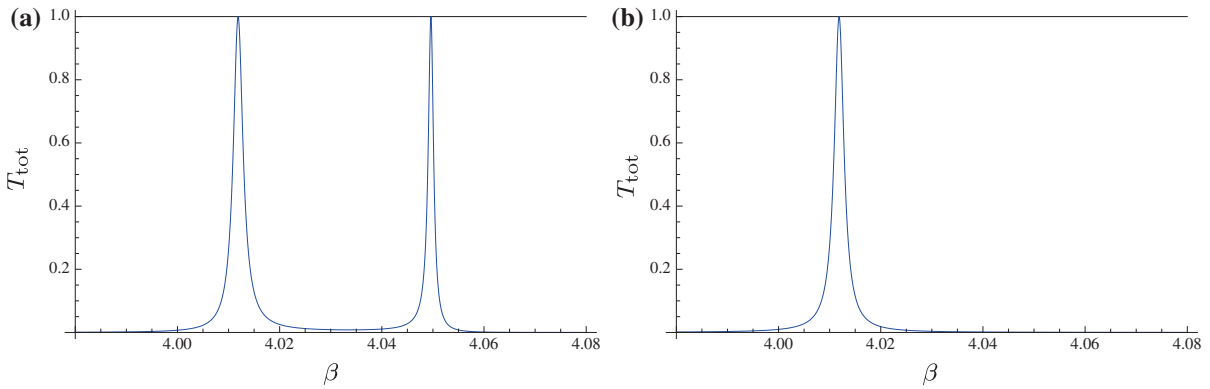

Fig. 13 Total transmittance T_{tot} as a function of β: **a** for a triplet consisting of an outer pair of gratings with $a = 0.1d$ and the central grating of rigid pins ($\eta/d = 1$); **b** for the outer pair of gratings with $a = 0.1d$ and $\eta/d = 2$. Data: $\theta_i = 15°$, $L = 2$.

et al. (2012) for the case of rigid pins. For a triplet exhibiting up-down symmetry, a plane wave characterised by the angle of incidence θ_i produces two transmission resonances, one linked to an anti-symmetric mode and the other associated with a symmetric mode. The latter mode's frequency depends on the relative horizontal shift ξ of the middle grating whilst the resonant frequency of the anti-symmetric mode is unaffected by the value of shift for the rigid pin triplets.

By adjusting the shift parameter ξ, the resonant frequencies of the two modes can be aligned. For the case of rigid pin triplets, three types of resonance coincide: the symmetric and anti-symmetric resonances occurring in the entire triplet, and the anti-symmetric resonance for the outer pair of gratings.

Figure 16a gives the total transmittance as a function of β for the triplet studied by Haslinger et al. (2012), and in addition in parts (b) and (c) it shows flexural displacements for both the outer pair of gratings and the entire triplet, at the β value corresponding to the point of minimum transmittance in part (a). The total transmittance is negligibly small at this minimum. The

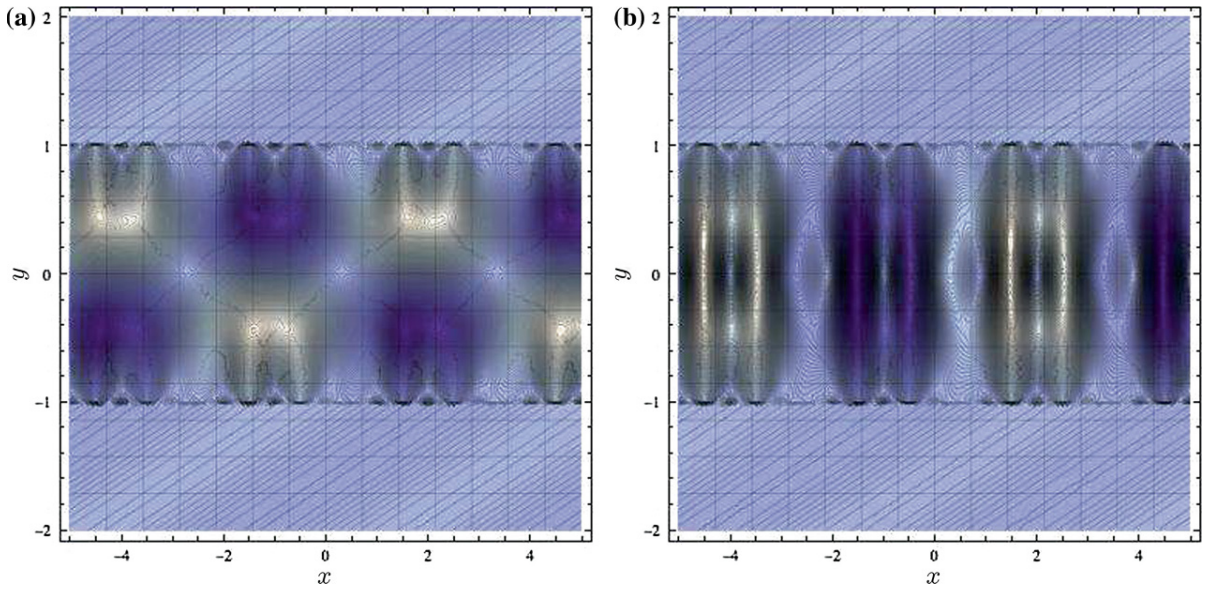

Fig. 14 Flexural displacement as a function of x and y for a triplet consisting of an outer pair of gratings with inclusions of radius $a = 0.1d$ and a central unshifted rigid pin grating, for an angle of incidence of 15°. **a** Anti-symmetric mode ($\beta_* = 4.011852$); **b** symmetric mode ($\beta_* = 4.0496094$). Data: $\eta = d = 1$ and $L = 2$

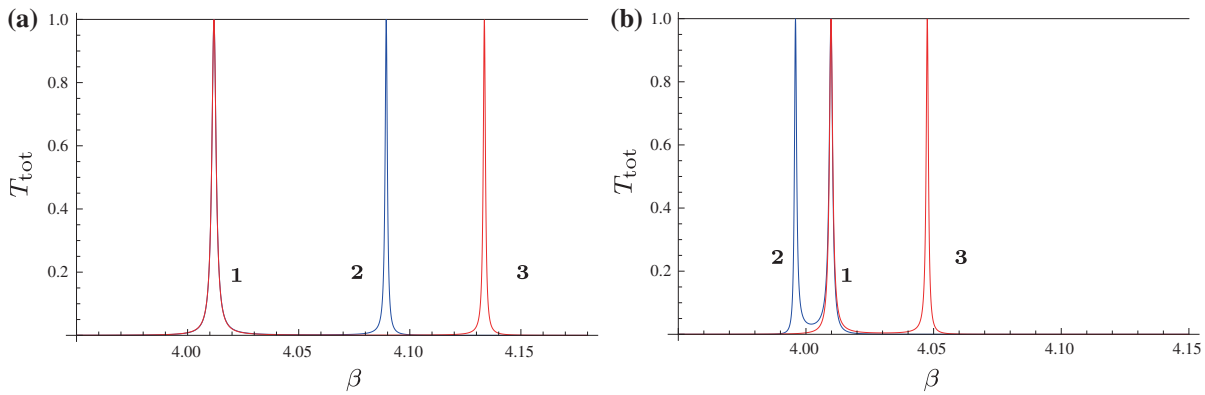

Fig. 15 Total transmittance $T_{\rm tot}$ as a function of β for two shifted triplets ($\xi = 0.25d$ and $\xi = 0.5d$) containing a central grating of rigid pins ($\eta = d = 1$). **a** Angle of incidence $\theta_i = 15°$, outer scatterer radius $a = 0.1d$; **b** $\theta_i = 20°$, $a = 0.085d$

symmetric and anti-symmetric modes add in the upper half of the structure and subtract in the lower half. The result is that the flux of energy into the region below the triplet is zero and the field above the triplet takes the form of a standing wave.

We will now show that it is possible to obtain a strong EDIT effect for grating triplets containing inclusions of finite radius. To construct examples of this, it is preferable to use triplets with the outer gratings having inclusions of small but nonzero radius, with the central grating consisting of rigid pins. Note that, if the inclusions in the outer gratings are not sufficiently small, the resonances of the anti-symmetric and symmetric modes can be made to coincide, but the transmittance peaks on either side of the zero transmittance point are quite weak.

We have been able to achieve striking examples of the EDIT phenomenon for triplets using inclusions with

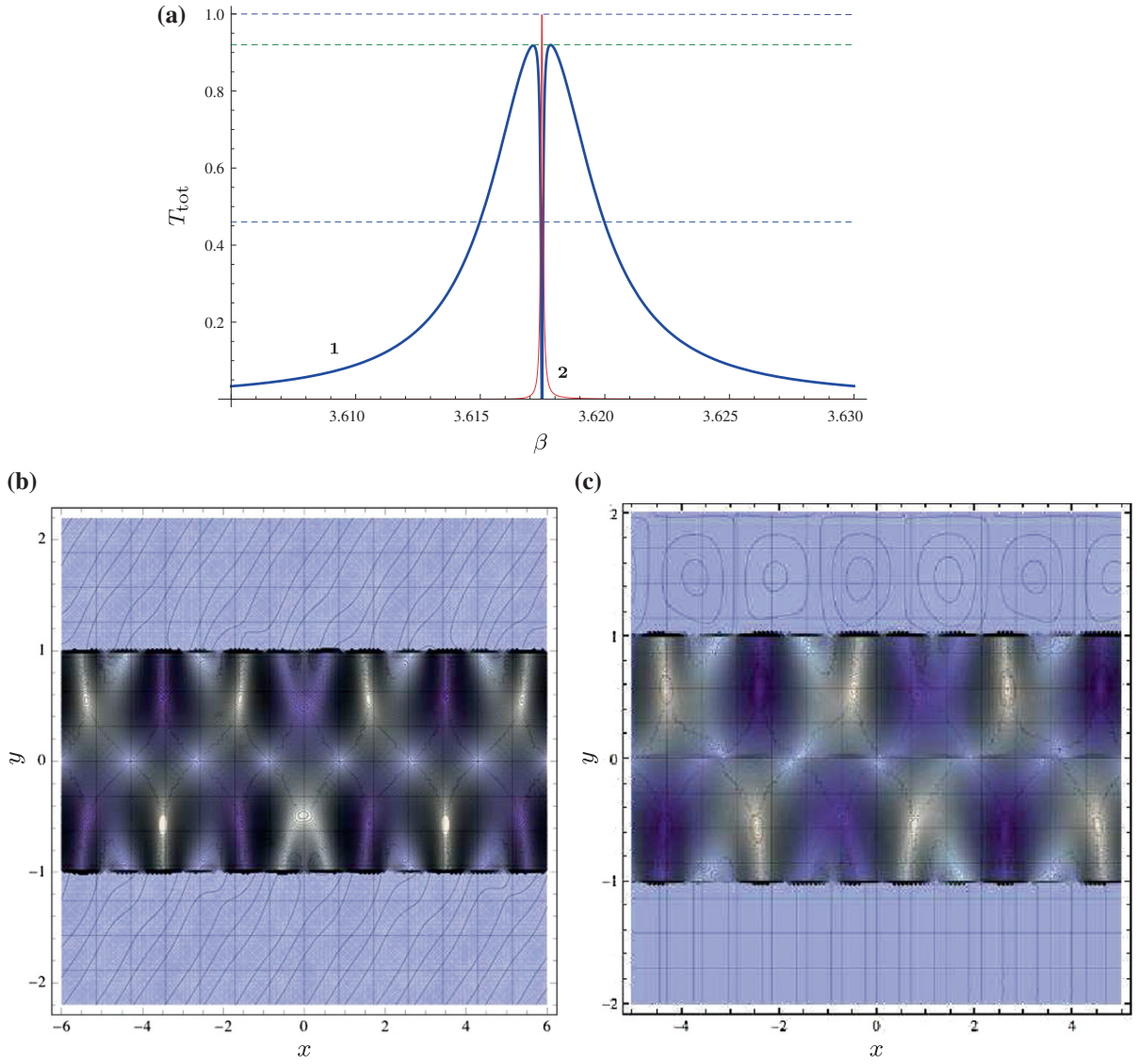

Fig. 16 **a** EDIT effect for a triplet of rigid pin gratings with the central grating shifted by $\xi = 0.25200d$ for the angle of incidence $\theta_i = 30°$. Total transmittance $T_{\rm tot}$ (curve **1**) as a function of β for the triplet. Curve **2** represents the total transmittance for the outer pair of gratings. **b** Flexural displacement for the outer pair of gratings of rigid pins ($\eta/d = 2$) as a function of x and y for $\beta_* = 3.61747$. **c** Flexural displacement for the triplet of rigid pin gratings ($\eta/d = 1$) as a function of x and y for $\beta_* = 3.61747$

radius $a = 0.01d$ for the outer pair of gratings. The results for an initial design with an angle of incidence of 27° are given in Fig. 17a while the results for an optimised system are shown in Fig. 17b for the angle of incidence of 26°. The transmittance notch in the latter case has an extremely high quality factor (around 1.80×10^5), five times the Q-factor for the rigid pin triplet of Fig. 16. We give the field plots in Fig. 18. The plot in part (a) shows the anti-symmetric mode for the outer pair of gratings, which is a plane propagating wave above and below the pair. The corresponding plot (part (b)) for the triplet shows a standing wave above the triplet and an evanescent field below it.

7 Concluding remarks

We have shown that a careful design of pairs or triplets of platonic gratings can deliver resonances in

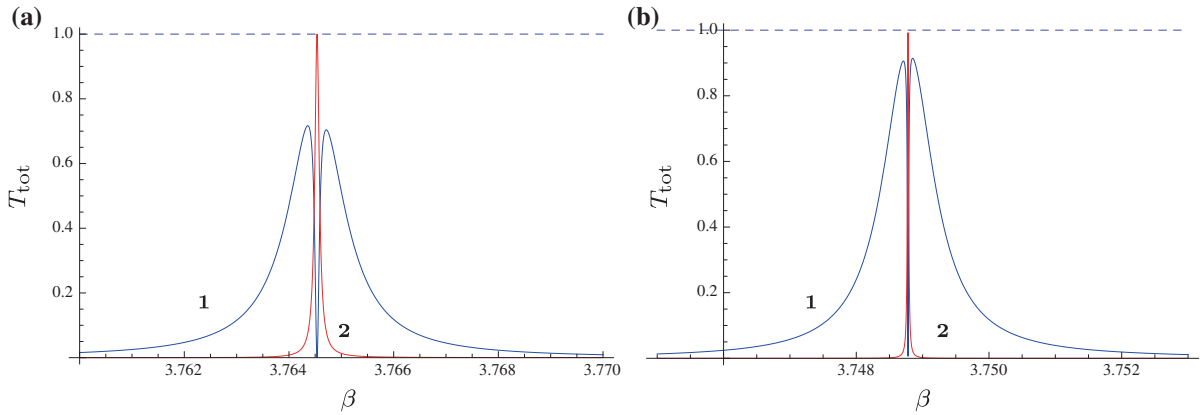

Fig. 17 Total transmittance $T_{\rm tot}$ (curve **1**) as a function of β for a triplet with a shifted central grating of rigid pins and the outer pair of gratings of inclusions of radius $a = 0.01d$. Curve **2** represents the total transmittance for the outer pair of gratings. Data used: **a** $\xi = 0.3112d, \theta_i = 27°$. **b** $\xi = 0.23265d, \theta_i = 26°$

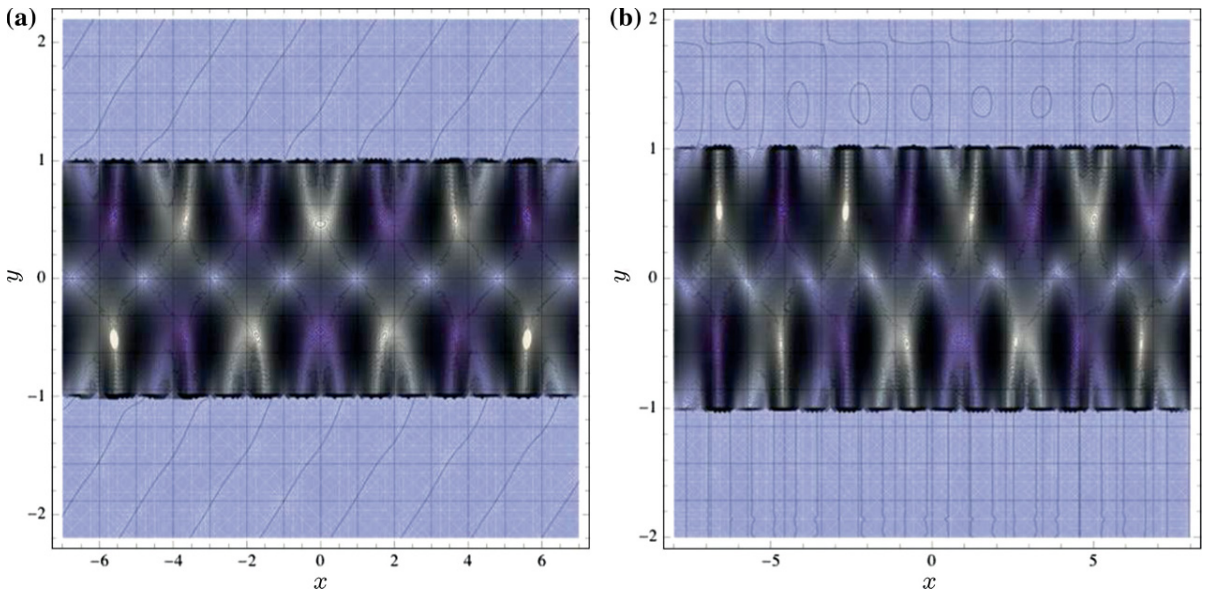

Fig. 18 **a** Flexural displacement as a function of x and y for a pair of gratings consisting of inclusions of radius $a = 0.01d$ with grating separation $\eta/d = 2$ for $\beta_* = 3.748779850$. **b** Flexural displacement as a function of x and y for a triplet with a shifted central grating of rigid pins ($\xi = 0.23265d$) and the outer pair of gratings consisting of inclusions of radius $a = 0.01d$ for $\beta_* = 3.748779784$, $\eta/d = 1$. Data: $\theta_i = 26°$

transmission with Q-factors ranging from 10^1 to 10^6. These Q-factors can be varied by altering the grating separation η, the angle of incidence θ_i and the inclusion radius a. Furthermore, by controlling the separation between the resonances for symmetric and anti-symmetric modes, we have shown that the EDIT phenomenon can deliver a high Q transmittance minimum in the middle of the transmittance peak, together with a strong variation in the magnitude and sign of the dispersive properties of the triplet. These demonstrated properties augur well for various applications of platonic grating systems in the moulding of flexural vibrations.

Acknowledgments S.G.H. gratefully acknowledges the financial support of the Duncan Norman Charitable Trust through

the Duncan Norman Research Scholarship. The participation of R.C.M. in this project was facilitated by a grant from the Research Centre in Mathematics and Modelling of the University of Liverpool; he also acknowledges support from the Australian Research Council through its Discovery Grants Scheme.

References

Abramowitz M, Stegun IA (1965) Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover Reprint

Bertoldi K, Bigoni D, Drugan JW (2007a) Structural interfaces in linear elasticity. Part I: nonlocality and gradient approximations. J Mech Phys Solids 55:1–34

Bertoldi K, Bigoni D, Drugan JW (2007b) Structural interfaces in linear elasticity. Part II: effective properties and neutrality. J Mech Phys Solids 55:35–63

Bertoldi K, Bigoni D, Drugan JW (2007c) A discrete-fibers model for bridged cracks and reinforced elliptical voids. J Mech Phys Solids 55:1016–1035

Bigoni D, Gei M, Movchan AB (2008) Dynamics of a prestressed stiff layer on an elastic half space: filtering and band gap characteristics of periodic structural models derived from long-wave asymptotics. J Mech Phys Solids 56:2494–2520

Bigoni D, Movchan AB (2002) Statics and dynamics of structural interfaces in elasticity. Int J Solids Struct 39:4843–4865

Born M, Wolf E (1959) Principles of optics. Pergamon Press, London

Botten LC, McPhedran RC, Lamarre JM (1985) Inductive grids in the resonant region: theory and experiment. Int J Infrared Millim Waves 6:511–575

Brun M, Guenneau S, Movchan AB, Bigoni D (2010) Dynamics of structural interfaces: filtering and focussing effects of elastic waves. J Mech Phys Solids 58(9):1212–1224

Evans DV, Porter R (2007) Penetration of flexural waves through a periodically constrained thin elastic plate in vacuo and floating on water. J Eng Maths 58:317–337

Farhat M, Guenneau S, Enoch S, Movchan AB (2009) Cloaking bending waves propagating in thin elastic plates. Phys Rev B 79:033102

Farhat M, Guenneau S, Enoch S (2010) High directivity and confinement of flexural waves through ultra-refraction in thin perforated plates. Euro Phys Lett 91:54003

Fleischhauer M, Imamoglu A, Marangos JP (2005) Electromagnetically induced transparency: optics in coherent media. Rev Mod Phys 77:633–673

Gei M, Movchan AB, Bigoni D (2009) Band-gap shift and defect-induced annihilation in prestressed elastic structures. J Appl Phys 105:063507

Haslinger SG, Movchan NV, Movchan AB, McPhedran RC (2012) Transmission, trapping and filtering of waves in periodically constrained elastic plates. Proc R Soc A 468:76–93

Movchan AB, Movchan NV, McPhedran RC (2007) Bloch-Floquet bending waves in perforated thin plates. Proc R Soc A 463:2505–2518

Movchan NV, McPhedran RC, Movchan AB, Poulton CG (2009) Wave scattering by platonic grating stacks. Proc R Soc A 465:3383–3400

Movchan NV, McPhedran RC, Movchan AB (2011) Flexural waves in structured elastic plates: mindlin versus bi-harmonic models. Proc R Soc A 467:869–880

McPhedran RC, Movchan AB, Movchan NV (2009) Platonic crystals: bloch bands, neutrality and defects. Mech Mater 41:356–363

Pelton EL, Munk BA (1979) Scattering from periodic arrays of crossed dipoles. IEEE Trans AP-27, 323–330

Poulton CG, McPhedran RC, Movchan NV, Movchan AB (2010) Convergence properties and flat bands in platonic crystal band structures using the multipole formulation. Waves Random Complex Media 20:702–716

Stenger N, Wilhelm M, Wegener M (2012) Experiments on elastic cloaking in thin plates. Phys Rev Lett 108:014301

Twersky V (1961) Elementary function representations of Schlömilch series. Arch Ration Mech Anal 8:323–332

Ulrich R, Tacke M (1973) Submillimeter wave-guiding on periodic metal structure. Appl Phys Lett 22:251–253

DOI 10.1007/978-3-319-04397-5_5
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9814-7

ORIGINAL PAPER

Fracture process in cortical bone: X-FEM analysis of microstructured models

Simin Li · Adel Abdel-Wahab · Emrah Demirci · Vadim V. Silberschmidt

Received: 3 November 2012 / Accepted: 28 January 2013 / Published online: 9 February 2013

Abstract Bones tissues are heterogeneous materials that consist of various microstructural features at different length scales. The fracture process in cortical bone is affected significantly by the microstructural constituents and their heterogeneous distribution. Understanding mechanics of bone fracture is necessary for reduction and prevention of risks related to bone fracture. The aim of this study is to develop a finite-element approach to evaluate the fracture process in cortical bone at micro-scale. In this study, three microstructural models with various random distributions based on statistical realizations were constructed using the global model's framework together with a submodelling technique to investigate the effect of microstructural features on macroscopic fracture toughness and microscopic crack-propagation behaviour. Analysis of processes of crack initiation and propagation utilized the extended finite-element method using energy-based cohesive-segment scheme. The obtained results were compared with our experimental data and observations and demonstrated good agreement. Additionally, the microstructured cortical bone models adequately captured various damage and toughening mechanisms observed in experiments. The studies of crack length and fracture propagation elucidated the effect of microstructural constituents and their mechanical properties on the microscopic fracture propagation process.

Keywords X-FEM · Microstructured model · Crack propagation · Fracture toughening mechanisms

S. Li · A. Abdel-Wahab · E. Demirci · V. V. Silberschmidt (✉)
Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK
e-mail: v.silberschmidt@lboro.ac.uk

A. Abdel-Wahab
e-mail: A.A.Abdel-wahab@lboro.ac.uk

E. Demirci
e-mail: E.Demirci@lboro.ac.uk

S. Li
e-mail: S.Li@lboro.ac.uk

1 Introduction

Fracture of cortical bone can significantly affect structural integrity of a load-bearing skeletal system, and, consequently, cause injuries, mobility loss and reduced life quality. As a naturally formed composite material, a cortical bone tissue is formed by heterogeneously distributed microstructural constituents that could be categorised into several hierarchical organizations from nano-scale to macro-scale levels (Currey 2011; Ritchie et al. 2005). At the nano-scale, bone is composed of mineralized collagen fiber matrix and extrafibrillar mineral particles known as carbonated hydroxyapatite (Currey 1999; Fratzl et al. 2004). At the micro-scale, cortical bone is laid down in layers of lamellar structure (3–7 μ m in thickness) that is similar to that of plywood composite—parallel with each other within layer, but

having a staggered arrangement between the adjacent layers (Ascenzi and Benvenuti 1986). Across a bone section, concentric layers of lamellae together with hollowed vascular channels form the most observable structure under microscope—a Haversian system (containing osteon and a Haversian canal) embedded into the remnants of a bone's remodelling process called *interstitial matrix*. Osteons are, on average, a 200 μm in diameter and up to 1 cm long cylindrical structure parallel to the bone's longitudinal axis (Ethier and Simmons 2007). In addition, a network of canals and channels formed across the bone's section accommodates blood vessels and a nervous system; those large canals, on average 50–90 μm in diameter, parallel to the bone's main axis are called *Haversian canals*. Bone has living cells called *osteocytes* that live within an interconnected network of microscopic channels called *canaliculi*. The latter are responsible for exchange of nutrients and waste between osteocytes (Ethier and Simmons 2007). The interface between osteons and interstitial matrix is called *cement line*; it is 2–5 μm in thickness. It plays a key role in the bone's mechanical behaviour, especially its fracture. However, the opinions in the literature with regard to the mechanical properties of cement line are rather controversial. Different experimental observations reported that the cement line can act either as a toughening mechanism deflecting a crack from osteons or as a weakening path that facilitates the crack initiation (Currey 2011; Ritchie et al. 2005). At the millimetre length scale, the dense and thick outer layer of cortical bone and the porous sponge-like trabecular bone make up the tissue-level bone structure (Peterlik et al. 2006). All these hierarchical levels work together in accord and complementing each other to achieve enhanced macroscopic mechanical properties of the bone tissue at the full-bone scale (Peterlik et al. 2006).

The micro-architecture of cortical bone has a significant effect on its mechanical and fracture properties. Moreover, preferential alignment of both collagen fibrils and mineral crystals at nano-scale and the osteons and Haversian cannals at micro-scale results in highly anisotropic mechanical and fracture behaviour of the tissue (Peterlik et al. 2006). The anisotropic ratio of fracture toughness for different crack propagation directions can be significantly large—from 2 to 3—depending on interaction of a propagating crack with the microstructural features, activation of various toughening mechanisms affecting fracture resistance: formation of microcracks in the vicinity of the main crack due to stress concentrations ahead of its tip (Vashishth et al. 2003; Zioupos and Currey 1994; Zioupos et al. 1996) and crack deflection and blunting at cement lines that create discontinuity at the boundary layers (Liu et al. 1999). Recently, it was reported that ligament bridging of crack in the wake zone is a dominant toughening mechanism in cortical bone as it reduces a driving force at the crack tip (Nalla et al. 2004, 2003, 2005). Several authors reported that toughening mechanisms are highly dependent on a crack propagation direction; therefore, fracture toughness of long bones is significantly higher in transverse and radial directions compared to the longitudinal one (Behiri and Bonfield 1986; Martin and Boardman 1993; Nalla et al. 2005).

Being a physiological living tissue, bone has the ability of continuously remodelling, repairing and adapting itself to the surrounding environment. Due to this inherent dynamics, both the microstructure of cortical bone and its mechanical behaviour vary dramatically from one part to another. Considering differences introduced by using various test methods and specimen's sizes, the spectrum of fracture energy per unit area of cortical bone varies from 920 to 2,780 N/m (Ritchie et al. 2005) for the same type of bone tested at same orientation. This variability is significant even for different cortices of a single bone (Bonney et al. 2011). Unlike traditional artificial composite materials, which have predefined average volume fraction and, consequently, a limited range of their mechanical properties, the volume fraction of each constituent of cortical bone is not unique and changes during the bone remodelling process. As a result, the randomly distributed elements of microstructure at the local region have a significant impact on variability of the mechanical behaviour of cortical bone at macro-scale level. However, a rather small number of studies was performed to unveil a correlation between the variation of microstructure and variability of the mechanical behaviour of cortical bone.

Finite-element simulations provide a powerful tool to analyse the fracture behaviour of materials at different length scales. Ural and Vashishth (Ural and Vashishth 2006) developed a cohesive-zone-element model to capture an experimentally observed rising crack-growth behaviour and age-related loss of bone toughness. Later, the same authors investigated the effects of age-related changes and orientation of crack

growth on a toughening behaviour of human cortical bone using the same model. The used approach—cohesive-zone (CZ) method—accounts for the nonlinear fracture mechanism and describes the nonlinear fracture process in terms of a traction-separation law; it has been broadly used in the literature (Ural et al. 2011; Yang et al. 2006) to investigate the fracture mechanics of cortical bone. However, it has an inherent drawback: the crack extension has to follow a predefined path around elements of the mesh. Obviously, in the case of fracture of real bones such a crack path is hard to predict. Similar to the CZ technique, an analytical approach based on perturbation technique were recently extended by Piccolroaz et al. (Piccolroaz et al. 2012) to solve an interfacial crack problem in heterogeneous material where a crack path was assumed coincide with a bonding interface between dissimilar materials. The extended finite-element method (X-FEM) was used in a small number of papers on fracture in bones to overcome this issue. For instance, Budyn and Hoc (Budyn and Hoc 2007) introduced a multiscale method to simulate multiple crack growth in a cortical bone tissue using X-FEM under simplified tensile loading conditions. In another recent attempt, Liu et al. (Liu et al. 2010) developed a homogenised X-FEM model to predict fracture of a proximal femur due to impact. Despite many attempts by various researchers, the model development for fracture of cortical bone is still limited to simplified formulations: simplified material properties; a microstructured model with applied linear-elastic material properties and boundary conditions for a very small region (less than 1 mm in length) (Budyn and Hoc 2007); a full-size bone model but with continuum homogenised material properties (Liu et al. 2010). There is still a need in a comprehensive X-FEM model that can reflect adequately the main features characteristic to the bone fracture process.

Therefore, in this paper, a microstructured model of cortical bones is proposed to study the effect of its microstructure on variability of fracture toughness and crack-propagation process for the case of three-point bending, typical for experimental fracture analysis.

2 Materials and methods

2.1 Modelling approach

To investigate the variability of fracture toughness and various toughening mechanisms induced by random microstructure, three two-dimensional microstructured models of cortical bone were developed in commercial finite-element software—Abaqus (Dassault Systèmes 2012). The models were established based on configuration of our three-point-bending experiments detailed in (Li et al. 2012). These models were constructed using a submodelling technique that focuses the computational power at the crack-propagation region while maintaining the full-scale approach of the model. The submodelling technique allows development of multiple models based on the same modelling object and extends the level of interest into a pre-defined region (usually with a finer mesh or more local geometric details) to achieve adequate and accurate results. The computational cost of submodelling technique is usually lower when compared with the whole-size model having the same level of accuracy. The developed approach employed two different levels of modelling of the bone tissue: full size global model for the macroscopic response of the entire specimen under three-point bending and three submodels reflecting microscopic responses of different localized microstructures during the crack propagation process. The boundary conditions in the submodel were derived for the correspondent region from the results of the global model using the displacement-control criterion based on the nodal field variables.

The model geometries correspond to those of the specimens used in our experiment: 25 mm × 2.72 mm × 5.43 mm (total length × width × thickness). Cylindrical loading pins at the three-point setup were modelled as analytical rigid bodies with a radius of 5 mm. The span (length between the centres of two holding pins) and the pre-crack length were chosen according to the experimental setup: 21.72 and 2.72 mm, respectively. Then, the submodel was extruded from the central un-cracked region of the global model with dimension of 2.72 mm × 2.72 mm (Fig.1). The pre-crack is mostly outside the submodel with only one element of its bottom middle surface cut by it.

2.2 Model configuration

The random microstructural distributions inside the submodels were constructed as four-phase composite structures consisting of interstitial bones (i) and Haversian systems that include osteons (ii), Haversian canals (iii) and cement line (iv). All geometrical parameters of

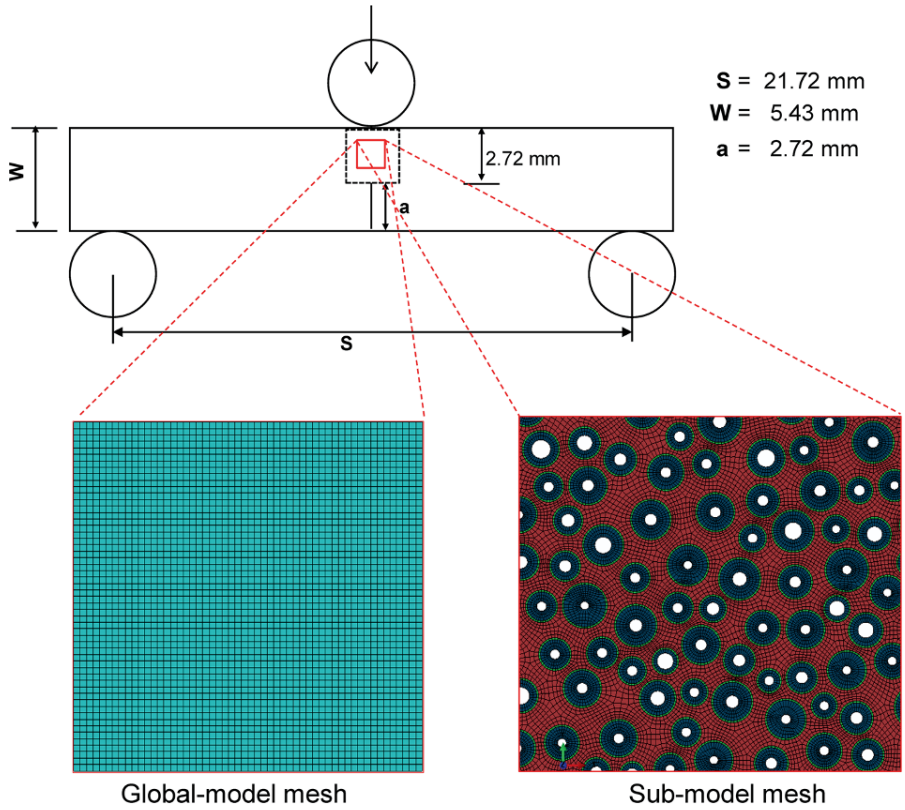

Fig. 1 Schematic illustration of model configuration for the three-point-bending setup using global model and microstructured sub-model

each model were defined based on statistical analysis of real microstructures (for details see (Abdel-Wahab et al. 2010b)) obtained with optical microscopy of radial-transverse sections from a lateral cortex where a large portion of osteonal structure can be observed. That analysis confirmed the volumetric fraction of osteons in the range from 28 to 55 %. The porosity ratio was measured between 5 and 13 %. The average width of cement line was close to 5 μm. The distributions of diameters of osteons and Haversian canals were regularized statistically using best-fit functions described in detail in (Abdel-Wahab et al. 2010b). The average diameters for osteons and Haversian canals were 99.89 and 23.1 μm, respectively.

The algorithm to generate random microstructures in the submodels was first programed in a custom-developed Matlab code according to the statistical data for real bone specimens, and then all the geometrical parameters were encoded into a python script to construct the microstructural model in Abaqus. Three representative microstructured cortical bone models were developed and employed in this study based on the statistical measurements for each constituent: osteons volume fraction varies from 30 to 51 %, while porosity changes from 5 % to around 8 % (Fig. 2). Full data on the volume fractions of microstructure constituents used in the models are listed in Table 1.

2.3 Material properties

In this study, the mechanical behaviour of cortical bone was introduced using an elasto-plastic transverse isotropic material formulation with regard to the radial-transverse section plane (see Fig. 1). At macroscopic level, the effective homogeneous material was used in the global model neglecting microscopic heterogeneity. The effective elasto-plastic material properties obtained from our macroscopic experiments (Abdel-Wahab et al. 2010a) were applied in the global model. On the other hand, at microscopic level, microstructural constituents play an important role in the localized fracture process and formation of toughening mechanisms. Consequently, the four-phase microstructured models of cortical bone were employed in the sub-model, and individual material properties based on

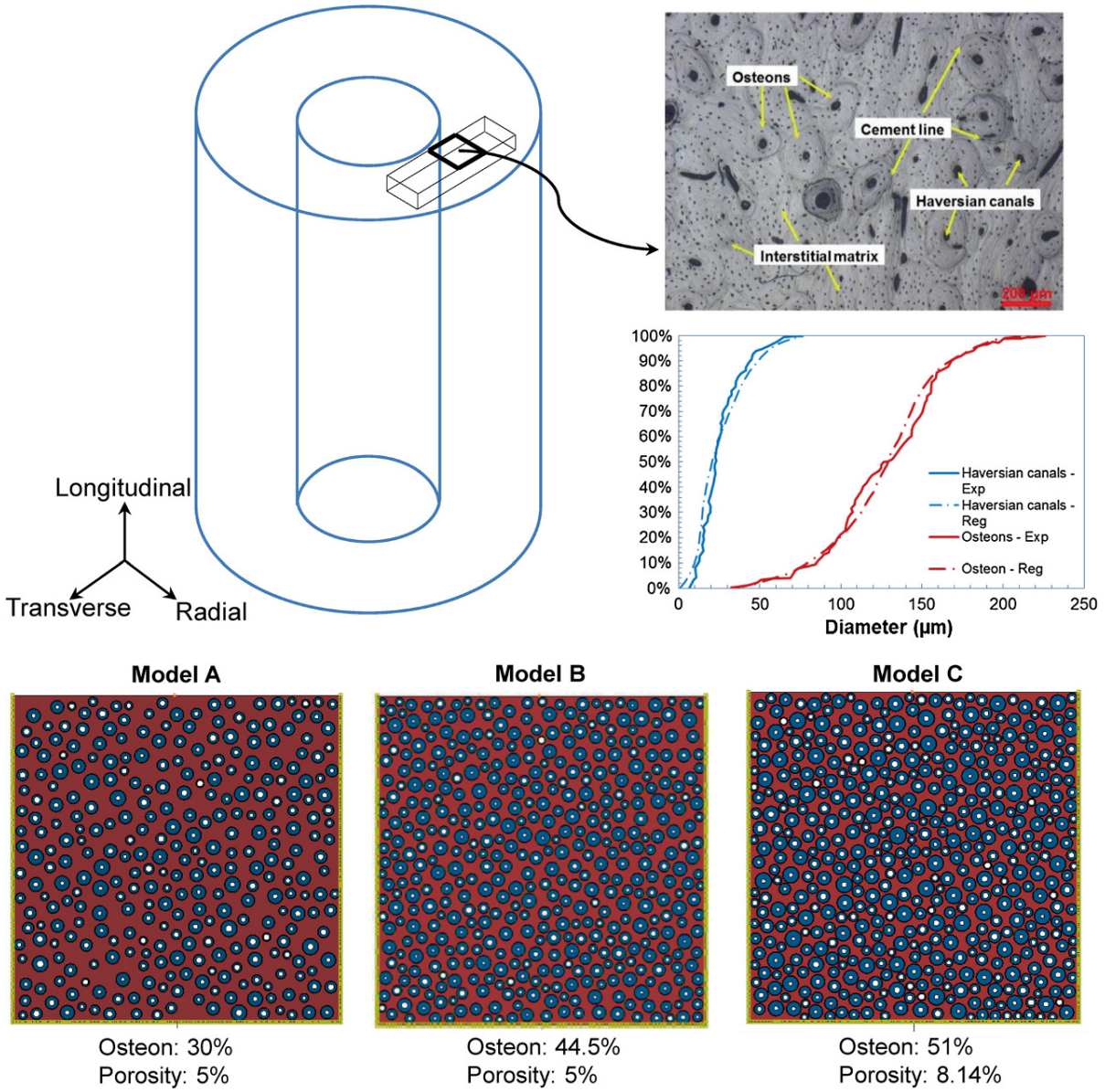

Fig. 2 Schematic illustration of specimen, image of real microstructure, Cumulative distribution functions and three statistical realizations of random microstructures with different fractions of constituents for three submodels

nano-indentation results (Abdel-Wahab et al. 2010b) were assigned to constituents. The elastic modulus of cement line was initially set to be 25 % lower than that of osteon based on the findings in (Budyn and Hoc 2007; Montalbano and Feng 2011), and two other levels—equal to that of osteon and 25 % higher—were also used to investigate the effect of cement line's properties on the fracture process in cortical bone. A strain-based yield criterion was implemented both in the global model and submodels, and a yield strain of 0.6 % was chosen based on (Abdel-Wahab et al. 2010a). The post-yield material behaviours in both global and sub-models were based on flow stress-strain curves obtained experimentally (Abdel-Wahab et al. 2010a,b). A summary of material properties used in this study is given in Table 2.

Table 1 Volume fractions of microstructure constituents for Models A, B and C

Constituent	Model A (%)	Model B (%)	Model C (%)
Osteon	30	44.5	51.2
Porosity	5.01	5.02	8.14
Interstitial matrix	58.77	41.25	30.04
Cement line	6.22	9.23	10.62

Damage and crack propagation in this study were modelled using the X-FEM technique in Abaqus (Dassault Systèmes 2012) that allows a crack to initiate and propagate through an arbitrary, solution-dependent path subject on the local material response. Hence, the X-FEM enrichment was applied to the whole model for all the cases. The local crack initiation and evolution were evaluated continuously based on chosen criteria. Crack initiation in a hard biological tissue (cortical bone) was commonly described with a strain-driven criterion (Nalla et al. 2003). Therefore, a strain-based crack-initiation criterion was set up both in global and microstructured models. It assumes that crack initiates when the maximum principal strain reaches its critical value and the newly defined crack direction is orthogonal to that of the maximum principal strain. A crack initiation strain of 0.65 % was chosen based on our observations of material response for the radial-transverse direction detailed in (Abdel-Wahab et al. 2010a). Once initiated, crack conforms to the energy-based damage evolution criterion, and the cracked element starts degradation and eventually fails. The governing formulation for the onset of crack utilizes a cohesive traction-separation constitutive behaviour to define the damage evolution of the cracked surface. It describes the rate, at which the cohesive stiffness of the cracked surface degrades once the crack-initiation criterion is fulfilled at particular element. The energy dissipated (fracture energy) as a result of damage progress is equal to the area under the traction-separation curve at the point of complete damage. The fracture energy in our models (Table 2) was defined according to the previous results (Abdel-Wahab et al. 2010b; Li et al. 2012; Ritchie et al. 2005).

2.4 Mesh-convergence analysis

The global model was discretised into 14,000 of four-node bilinear plane-strain quadrilateral elements and ran on an eight-processor (quad-core Intel I7 970 CPU) PC while the submodels were meshed using 150,000 to 200,000 elements of the same type and ran on a 60-processors (five six-core Intel Westmere Xeon X5650 CPUs) high-performance cluster. The Abaqus implicit solver was used in both types of simulations. A mesh-convergence study was carried out for the global model using six different element sizes, and the obtained results were analysed in terms of peak reaction forces as demonstrated in Fig. 3. Apparently, the reaction force converges when the minimum element size reduces to 100 μm or below. Therefore, the minimum element size for the global model was chosen to be 50 μm and the minimum element size for the submodel was around 5 μm.

3 Results

3.1 Effect of microstructure on variability of fracture toughness

Three microstructured models of cortical bone were analysed in this study. Their results are compared with that of the effective homogeneous model as well as experimental data in terms of force—displacement

Table 2 Material properties used in global model and microstructured submodels (Budyn and Hoc 2007; Katz 1984; Abdel-Wahab et al. 2010b; Ritchie et al. 2005)

	Effective homogenised material	Osteons	Interstitial matrix	Cement line
Elastic modulus (GPa)	11.18	12.85	14.12	9.64
Poisson's ratio	0.167	0.17	0.153	0.49
Yield strain	0.6 %	0.6 %	0.6 %	0.6 %
Fracture initiation strain	0.65 %	0.65 %	0.65 %	0.65 %
Fracture energy release rate (N/m)	2043	860	238	146

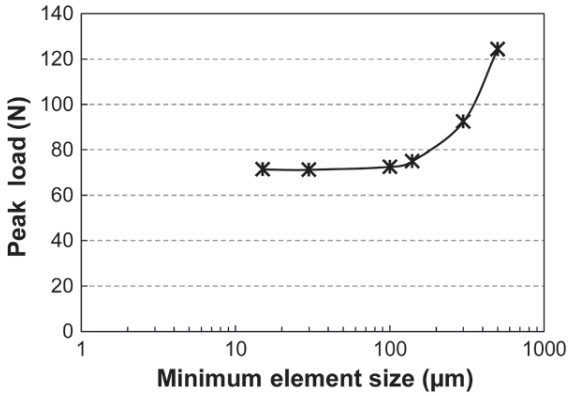

Fig. 3 Mesh convergence study; mesh sizes varied from 500 to 30 μm

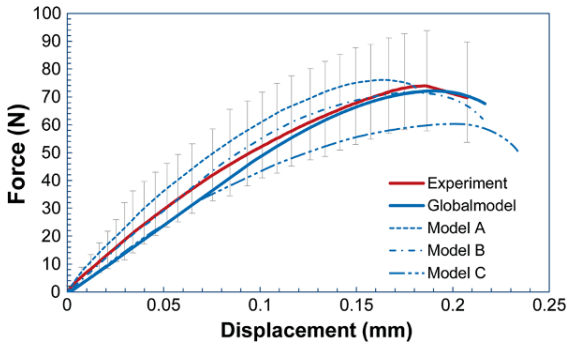

Fig. 4 Comparison of experimental (Li et al. 2012) force—displacement diagram with results for different FE models (*error bar* indicates variations of experimental results)

diagram in Fig. 4. Dissimilar fracture-resistance behaviours were evidenced for three different microstructured models. The change in the microstructure at microscopic level has a significant impact on the macroscopic fracture toughness of cortical bone. Among the three models, Model A has the highest critical value of J integral—2,503 N/m, while Models B and C result in 2,369 and 2,212 N/m, respectively. This decreasing trend in the fracture resistance is apparently linked to the increasing volume fractions of osteons and porosity (Haversian canals in this case). From the morphological point of view, the bone remodelling process generates new Haversian systems (each including an osteon, a Haversian canal and a cement line) to replace the old, damaged regions as an adaptive process. The newly formed bone cell is usually less mineralized than its surrounding area due to the fact that mineral concentration period lasts longer than the remodelling process (Currey 2011). As a result, a large fraction of

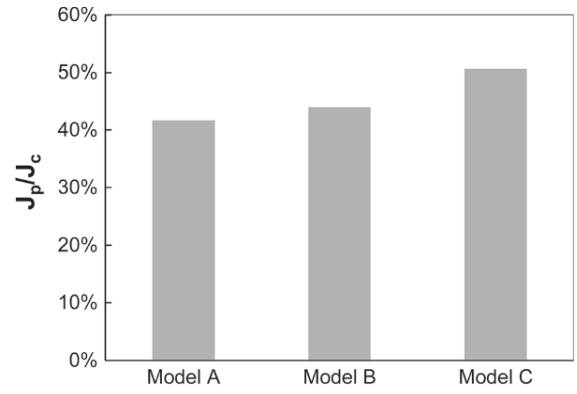

Fig. 5 Comparison of proportions of energy associated with plastic deformation for different models

less mineralized osteons associated with the bone-mass and stiffness reduction has a negative impact on the overall fracture resistance of cortical bone. Still, benefiting from their low stiffness but high fracture toughness, osteons demonstrate a higher failure strain when compared with interstitial matrix and, in general, offer a positive effect on fracture toughness. On the other hand, the increasing proportion of Haversian system leads to the increase in structural compliance as a result of cavitation, hence, to increased overall fracture strain (Fig. 4). These mutual effects of microstructural constituents result in the variation of macroscopic fracture toughness. Significant nonlinearity observed at the initial loading stage during the experiment was successfully captured using the microstructured model. Comparing the proportion of the plastic component (J_p) of the critical value of J-integral (J_c) in each model, an increased tendency for the energy associated with plastic deformation is observed for the increase of osteon and porosity volume fractions (Fig. 5). Based on the above findings, the bone remodelling process related with the increasing fraction of osteons and porosity changes the bone's fracture resistance from a stress-based mode to a more strain-based mode—fracture stress resistance reduces but fracture strain resistance increases.

3.2 Heterogeneous fracture process due to microstructure

At the global level, the effective homogeneous material model was able to capture a macroscopic response in terms of force—displacement curve (Fig. 4). However,

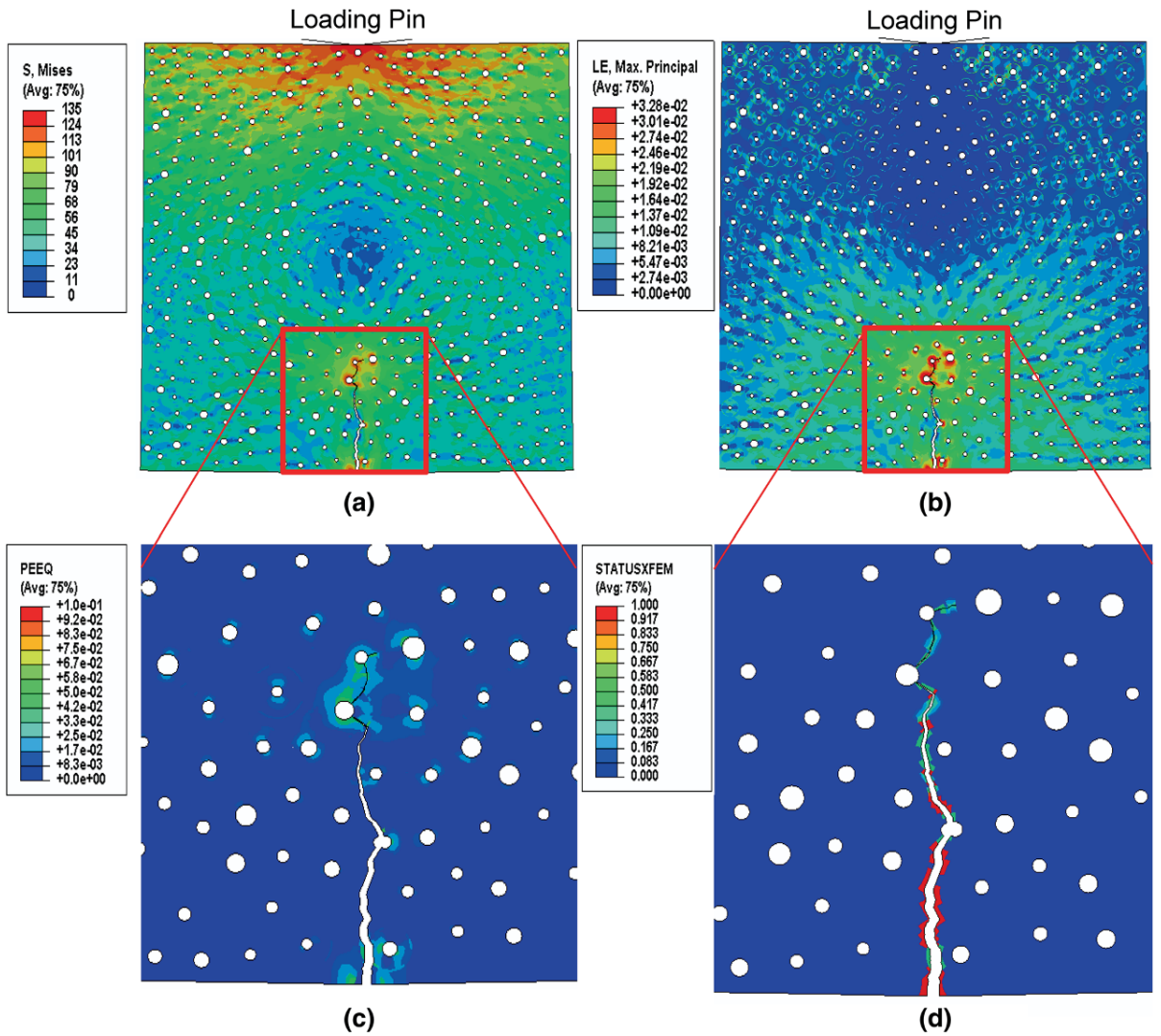

Fig. 6 Contour plots for von-Mises stress (**a**), maximum principal strain (**b**), equivalent plastic strain (PEEQ) (**c**) and damage scale factor for X-FEM (STATUS) (**d**) for Model B for crack propagation approaching for maximum reaction force (**a** and **b** represent full Model B)

the detailed fracture evolution process, especially the localized damage zone was neglected at this level. On the contrary, the heterogeneous models with random microstructures, operating within the framework of the global model using direct displacement-controlled boundary conditions, emphasise the effect of the local non-uniform stress-strain field on the crack propagation process at microscopic level. Figure 6 presents contour plots for von-Mises stress, maximum principal strain, equivalent plastic strain (PEEQ) and a damage scale factor for X-FEM (STATUS) for Model B when the crack is approaching the state of the maximum reaction force. As evidenced from the figure, a diffused stress pattern is characteristic for the von-Mises contour, while a cross-hatched strain pattern for the maximum principal strain contour is located ahead of the crack tip (in the compressive region of the specimen) with a diffused strain pattern near it (in the tensile region). These dissimilar stress and strain patterns around the crack tip coincide with results of the previous experimental studies (Boyce et al. 2005; Ebacher and Wang 2008; Nyman et al. 2009), in which the

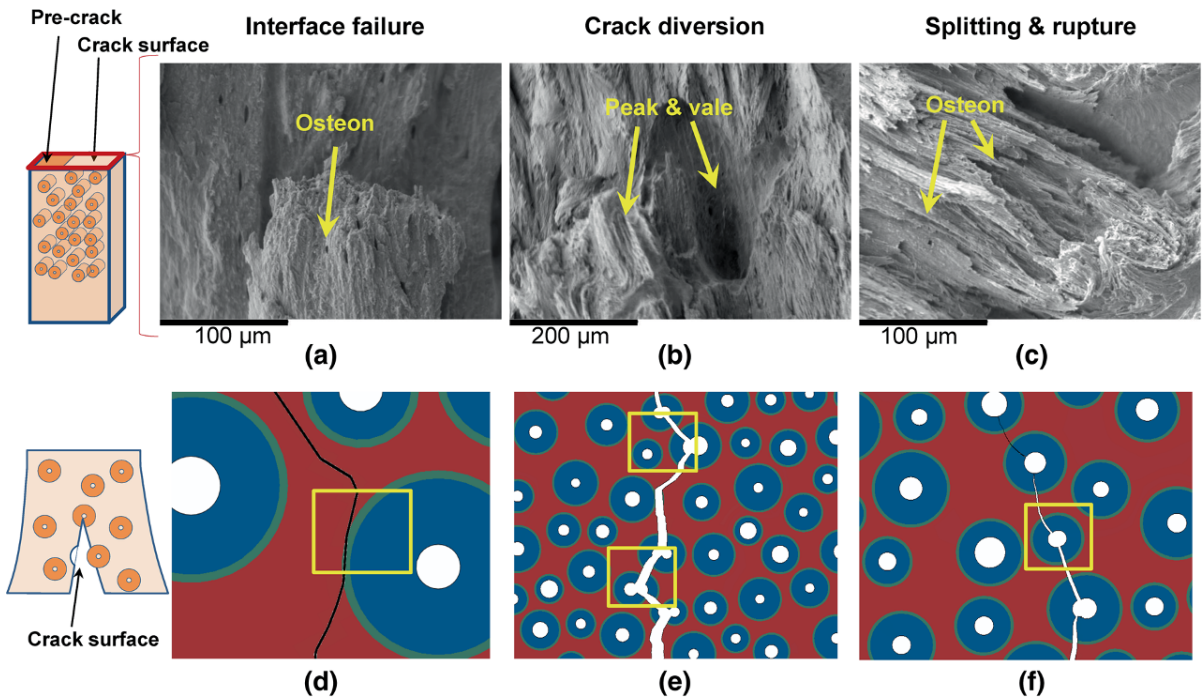

Fig. 7 Comparison of toughening mechanisms in radial-transverse crack plane between experimental results (**a**–**c**) and numerical simulations (**d**–**f**): **a**, **d** interface failure between osteon and interstitial matrix; **b**, **e** crack deviation from its central line towards vulnerable but twisted and deflected crack path as a result of material heterogeneity; **c**, **f** splitting of osteons and breakage of ligament due to crack opening observed in SEM image (**c**)

authors indicated that such distinctive stress and strain fields in tension and compressive regions could lead to realization of different damage fracture mechanisms. Equivalent plastic strain illustrated in Fig. 6c indicates that the area undergoes plastic deformation during the crack propagation process. The identified plastic zone around the crack tip is within 1–2 osteonal radius i.e. approximately 100 μm in length. The value seems to be higher than 17 μm reported in the experimental work (Robertson et al. 1978). One possible reason for this larger plastic-zone size predicted in our model is the lack of multiple cracks formation in the current model, while, in reality, micro-cracks and natural imperfections inclusions in front of the crack tip may develop into mini cracks frontal that can release local stress concentration, thus, reducing the plastic-zone size.

3.3 Microstructure-related difference in toughening mechanisms

The damage scale factor denoted as STATUS in Fig. 6d indicates that 1/3 of the crack surface is still under

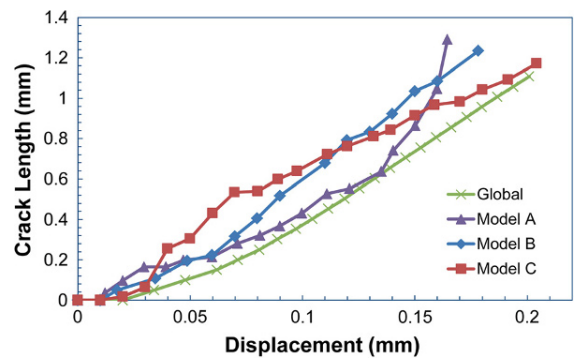

Fig. 8 Crack length—displacement diagram for global and microstructured models (the total crack length is measured until reaching the maximum reaction force)

traction force and acts as toughening mechanisms that contributes to the non-linear fracture process. The toughening mechanisms active in a radial-transverse crack specimen can be divided predominantly into three types (Ritchie et al. 2005): (i) interfacial debonding as a result of the material's discontinuity at the interface between osteons and interstitial matrix—the formation

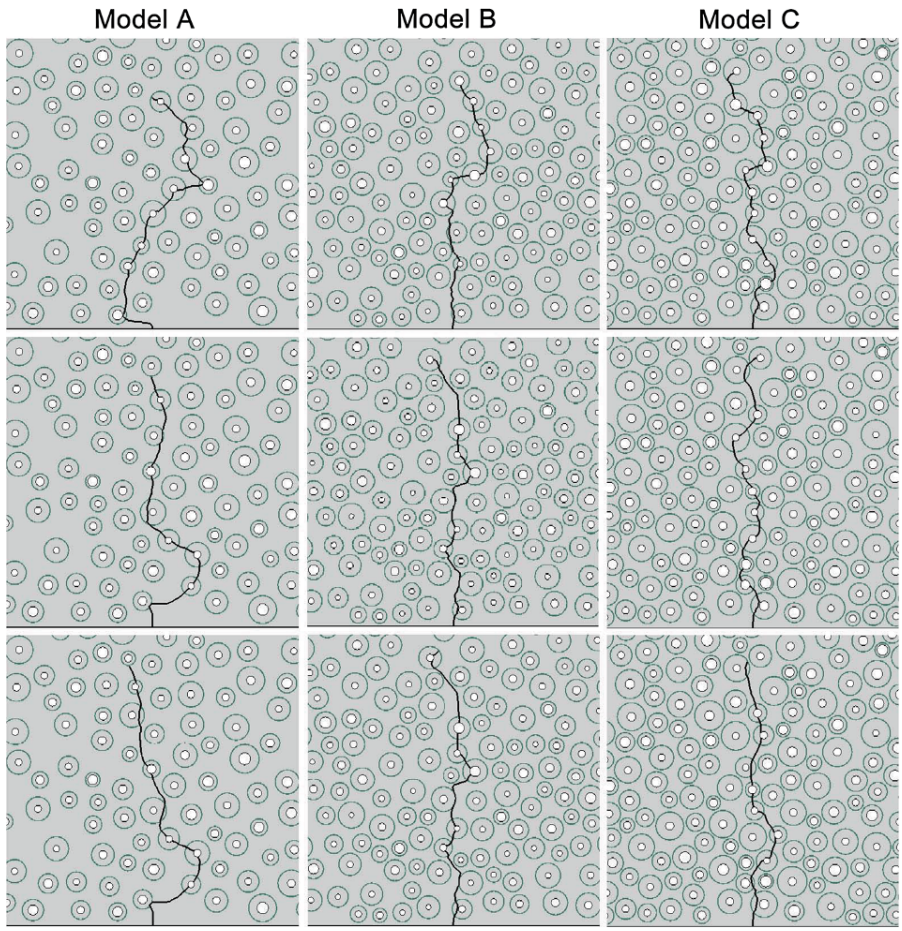

Fig. 9 Crack propagation trajectories for various elastic moduli of *cement line* for three microstructured models: row (**a**): 25 % *lower than* that of osteon; row (**b**): equal to that of osteon; row (**c**): 25 % *higher than* that of osteon

of the weak path of cement line; (ii) crack diversion due to microstructural heterogeneity and material imperfections at which crack is redirected towards the most vulnerable part producing a twisted and deflected fracture path; (iii) uncracked-ligament bridging caused by osteon splitting and rupture acting as a post-crack toughening mechanism behind the crack tip. In this study, the microstructured models were able to capture these main features of the toughening mechanisms as shown in Fig. 7. Figure 7d demonstrates an interface failure predicted by the model as the crack bends away from the osteon due to the discontinuity in the cement-line region. Figure 7e reveals the crack-diversion mechanism as crack deviates from its central line towards a vulnerable but twisted and deflected crack path. The uncracked-ligament bridging behind the crack tip is presented as a cohesive traction force between the damaged elements along the crack path (Fig. 7f).

3.4 Crack lengths analysis

To investigate the effect of microstructural constituents on the crack propagation process, the crack length is plotted in Fig. 8 as a function of displacement of loading pin for the global model and three different sub-models. Their respective crack propagation paths are demonstrated in Fig. 9, row a. The total crack length was measured until reaching the maximum reaction force. It is clear from Fig. 8 that Model A has the longest overall crack length, while Model C has the shortest one. Comparing the respective crack trajectories, the higher crack length related to Model A is largely defined by significant crack deflections observed in Fig. 9a. As a result of increase in the fractions of osteons and porosity from Model A to Model C, the effect of crack-deflection mechanism gradually reduces (Fig. 9a–c). On the other hand, the crack-propagation rate (with respect to the loading-pin displacement) in Model C is higher at

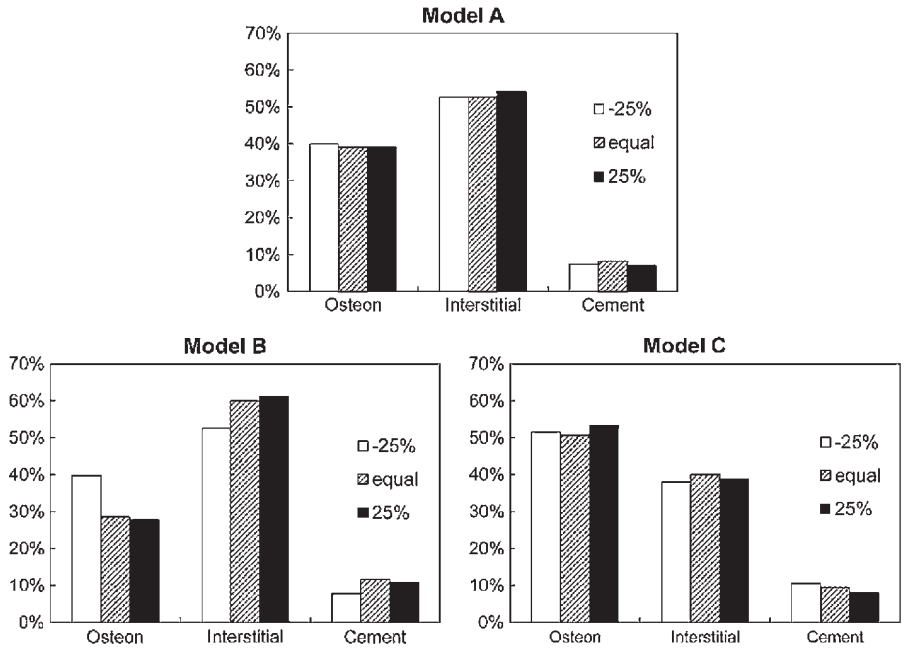

Fig. 10 Fractions of crack path in microstructure constituents for various magnitudes of cement line's modulus

initial stage, but gradually reduces as the crack propagates through more Haversian systems, whereas Model B shows a moderate linear evolution process and Model A demonstrates an increased crack-propagation rate. It seems that an increased fraction of Haversian systems has a negative effect on the crack-propagation rate and constrains the crack-diversion magnitude. This finding is consistent with experimental observation in (Zimmermann et al. 2011), where the authors concluded that age-related changes in morphology of microstructure as a result of remodelling process may lead to suppression of the crack-deflection mechanism and reduction of the total crack length.

3.5 Effect of cement line

The effect of cement line was studied by changing the magnitude of its elastic modulus within the range 25 % below and above that of osteon. The respective results for the crack propagation trajectory are compared in Fig. 9 for three different microstructural models. The result indicates that an increase in cement line's modulus to the levels equal to, or 25 % higher than that of the osteon results in similar crack trajectories, that differ from the initial ones (i.e. for 25 % lower modulus) for both Model A and Model B. This higher stiffness of cement line leads to some rise of fracture propagation in the regions with low fracture toughness—interstitial areas (Fig. 10). Moreover, higher stiffness also results in a higher rate of interface debonding in Model A and Model B (Fig. 9b, c) where cement lines facilitate crack propagation around osteons. However, no substantial difference was found between the two groups (equal to and 25 % higher). On the other hand, the lower cement modulus increase the chance of osteonal fracture and penetration into Haversian canal in Models A and B, where high fracture toughness and high compliance regions could potentially increase the overall fracture resistance and may lead to more crack deflections and arrests. As the osteon and porosity density increase in Model C, the effect of the local heterogeneity becomes more dominant. Cracks are likely to grow along the most vulnerable path, and the effect of cement line relents. Therefore, the influence on the crack-propagation trajectory is less pronounced than in two other models. In summary, cement line plays an important role in the crack-propagation process in cortical bone. Variation of its mechanical properties can considerably affect the shape of local crack trajectory. Both scenarios demonstrated in our models have been widely discussed in previous research (Currey 2011; Ritchie et al. 2005). Considering the fact that bone is a dynamic living tissue, the mechanical properties of cement line are likely to vary with time and locations. It is thus sensible that a 25 % differences in the cement

line's modulus within the local area can cause both toughening and weakening mechanisms as observed in experiment (Chan and Nicolella 2012).

4 Conclusions

The fracture process in cortical bone was evaluated in this study based on the developed X-FEM models. Three models with different random microstructures were developed and imbedded into a homogeneous global model to investigate the effect of microstructural changes and the related varying local mechanical behaviour on the fracture propagation process in cortical bone. The results obtained in this study indicate that local changes in volume fractions of microstructural constituents have a significant effect on variability of the macroscopic fracture toughness. The developed microstructured models of cortical bone are able to represent accurately the non-uniform plastic deformation associated with the nonlinear fracture process as well as realization of distinct damage and fracture toughening mechanisms observed in experiments. Moreover, the use of different statistical realization of random microstructure demonstrated importance of the local heterogeneity on the fracture propagation process. Various crack propagation trajectories and crack lengths were observed with different microstructured bone models. The changes in the underlying microstructure of cortical bone and its mechanical properties result in different toughening mechanisms, that, in turn, affect the crack propagation process in a dissimilar manner. High volume fractions of osteons and porosity result in a smoother fracture surface as a result of a lack of crack-diversion mechanisms; the higher stiffness of cement line supresses the osteonal crack and facilitates the interstitial damage and interface debonding. The knowledge obtained through the development of these microstructured X-FEM models provides an additional insight into the micro-scale fracture process in cortical bone and might be used in the future to provide support and guidance for treatments against bone fracture.

Acknowledgments The authors acknowledge the financial support from EPSRC UK (grant no. EP/G048886/1).

References

Abdel-Wahab AA, Alam K, Silberschmidt VV (2010a) Analysis of anisotropic viscoelastoplastic properties of cortical bone tissues. J Mech Behav Biomed Mater 4:807–820

Abdel-Wahab AA, Maligno AR, Silberschmidt VV (2010b) Micro-scale numerical model of bovine cortical bone: analysis of plasticity localization. In 10th ASME Biennial Conference on Engineering Systems Design and Analysis, Proceedings of the ASME 10th Biennial Conference on Engineering Systems Design and Analysis, 2010, vol 1. Istanbul, Turkey, pp 821–829

Ascenzi A, Benvenuti A (1986) Orientation of collagen fibers at the boundary between two successive osteonic lamellae and its mechanical interpretation. J Biomech 19:455–463

Behiri JC, Bonfield W (1989) Orientation dependence of the fracture mechanics of cortical bone. J Biomech 22:863–867, 869–872

Bonney H, Colston BJ, Goodman AM (2011) Regional variation in the mechanical properties of cortical bone from the porcine femur. Med Eng Phys 33:513–520

Boyce TM, Fyhrie DP, Glotkowski MC, Radin EL, Schaffler MB (2005) Damage type and strain mode associations in human compact bone bending fatigue. J Orthop Res 16:322–329

Budyn E, Hoc T (2007) Multiple scale modeling of cortical bone fracture in tension using X-FEM. R E M N 16:213–236

Chan KS, Nicolella DP (2012) Micromechanical modeling of R-curve behaviors in human Cortical Bone. J Mech Behav Biomed Mater http://dx.doi.org/10.1016/j.jmbbm.2012.09.009

Currey JD (2011) The structure and mechanics of bone. J Mater Sci 47:41–54

Currey JD (1999) The design of mineralised hard tissues for their mechanical functions. J Exp Biol 202:3285–3294

Dassault Systèmes. (2012) Abaqus v6.12 Documentation-ABAQUS analysis user's manual. ABAQUS Inc; 6.12

Ebacher V, Wang R (2008) A unique microcracking process associated with the inelastic deformation of haversian bone. Adv Funct Mater 19:57–66

Ethier CR, Simmons CA (2007) Introductory biomechanics: from cells to organisms. Cambridge University Press, New York

Fratzl P, Gupta HS, Paschalis EP, Roschger P (2004) Structure and mechanical quality of the collagen-mineral nanocomposite in bone. J Mater Chem 14:2115–2123

Katz JL, Yoon HS, Lipson S, Maharidge R, Meunier A, Christel P (1984) The effects of remodeling on the elastic properties of bone. Calcif Tissue Int 36:31–36

Li S, Abdel-Wahab A, Silberschmidt VV (2012) Analysis of fracture processes in cortical bone tissue. Eng Fract Mech: http://dx.doi.org/10.1016/j.engfracmech.2012.11.020

Liu D, Weiner S, Daniel Wagner H (1999) Anisotropic mechanical properties of lamellar bone using miniature cantilever bending specimens. J Biomech 32:647–654

Liu XC, Qin X, Du Z (2010) Bone fracture analysis using the extended finite element method (XFEM) with abaqus. The 34th annual meeting of the American society of biomechanics. 2010. Brown University

Martin RB, Boardman DL (1993) The effects of collagen fiber orientation, porosity, density, and mineralization on bovine cortical bone bending properties. J Biomech 26:1047–1054

Montalbano T, Feng G (2011) Nanoindentation characterization of the cement lines in ovine and bovine femurs. J Mater Res 26:1036–1041

Nalla RK, Kruzic JJ, Ritchie RO (2004) On the origin of the toughness of mineralized tissue: microcracking or crack bridging? Bone 34:790–798

Nalla RK, Kinney JH, Ritchie RO (2003) Mechanistic fracture criteria for the failure of human cortical bone. Nat Mater 2:164–168

Nalla RK, Kruzic JJ, Kinney JH, Ritchie RO (2005) Mechanistic aspects of fracture and R-curve behavior in human cortical bone. Biomater 26:217–231

Nyman JS, Leng H, Dong XN, Wang X (2009) Differences in the mechanical behavior of cortical bone between compression and tension when subjected to progressive loading. J Mech Behav Biomed Mater 2:613–619

Peterlik H, Roschger P, Klaushofer K, Fratzl P (2006) Orientation dependent fracture toughness of lamellar bone. Int J Fract 139:395–405

Piccolroaz A, Mishuris G, Movchan A, Movchan N (2012) Perturbation analysis of mode III interfacial cracks advancing in a dilute heterogeneous material. Int J Solids Struct 49: 244–255

Ritchie RO, Kinney JH, Kruzic JJ, Nalla RK (2005) A fracture mechanics and mechanistic approach to the failure of cortical bone. Fatigue Fract Eng Mater Struct 28:345–371

Robertson DM, Robertson D, Barrett CR (1978) Fracture toughness, critical crack length and plastic zone size in bone. J Biomech 11:359–364

Ural A, Vashishth D (2006) Cohesive finite element modeling of age-related toughness loss in human cortical bone. J Biomech 39:2974–2982

Ural A, Zioupos P, Buchanan D, Vashishth D (2011) The effect of strain rate on fracture toughness of human cortical bone: a finite element study. J Mech Behav Biomed Mater 4:1021–1032

Vashishth D, Tanner KE, Bonfield W (2003) Experimental validation of a microcracking-based toughening mechanism for cortical bone. J Biomech 36:121–124

Yang QD, Cox BN, Nalla RK, Ritchie RO (2006) Fracture length scales in human cortical bone: the necessity of nonlinear fracture models. Biomater 27:2095–2113

Zimmermann EA, Schaible E, Bale H, Barth HD, Tang SY, Reichert P et al (2011) Age-related changes in the plasticity and toughness of human cortical bone at multiple length scales. Proc Natl Acad Sci 108:14416–14421

Zioupos P, Currey JD (1994) The extent of microcracking and the morphology of microcracks in damaged bone. J Mater Sci 29:978–986

Zioupos P, Wang X, Currey JD (1996) The accumulation of fatigue microdamage in human cortical bone of two different ages in vitro. Clin Biomech 11:365–375

DOI 10.1007/978-3-319-04397-5_6
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9818-3

ORIGINAL PAPER

Minimum theorems in 3D incremental linear elastic fracture mechanics

A. Salvadori · F. Fantoni

Received: 15 August 2012 / Accepted: 6 February 2013 / Published online: 5 March 2013

Abstract The crack propagation problem for linear elastic fracture mechanics has been studied by several authors exploiting its analogy with standard dissipative systems theory (see e.g. Nguyen in Appl Mech Rev 47, 1994, Stability and nonlinear solid mechanics. Wiley, New York, 2000; Mielke in Handbook of differential equations, evolutionary equations. Elsevier, Amsterdam, 2005; Bourdin et al. in The variational approach to fracture. Springer, Berlin, 2008). In a recent publication (Salvadori and Carini in Int J Solids Struct 48:1362–1369, 2011) minimum theorems were derived in terms of crack tip "quasi static velocity" for two-dimensional fracture mechanics. They were reminiscent of Ceradini's theorem (Ceradini in Rendiconti Istituto Lombardo di Scienze e Lettere A99, 1965, Meccanica 1:77–82, 1966) in plasticity. Following the cornerstone work of Rice (1989) on weight function theories, Leblond et al. (Leblond in Int J Solids Struct 36:79–103, 1999; Leblond et al. in Int J Solids Struct 36:105–142, 1999) proposed asymptotic expansions for stress intensity factors in three dimensions—see also Lazarus (J Mech Phys Solids 59:121–144, 2011). As formerly in 2D, expansions can be given a Colonnetti's decomposition (Colonnetti in Rend Accad Lincei 5, 1918, Quart Appl Math 7:353–362, 1950) interpretation. In view of the expression of the expansions proposed in Leblond (Int J Solids Struct 36:79–103, 1999), Leblond et al. (Int J Solids Struct 36:105–142, 1999) however, symmetry of Ceradini's theorem operators was not evident and the extension of outcomes proposed in Salvadori and Carini (Int J Solids Struct 48:1362–1369, 2011) not straightforward. Following a different path of reasoning, minimum theorems have been finally derived.

Keywords Fracture mechanics · Crack growth · Brittle fracture · Variational formulations

A. Salvadori (✉)· F. Fantoni
CeSiA—Research Center on Applied Seismology and Structural Dynamics, Dipartimento di Ingegneria Civile, Architettura, Territorio, Ambiente e di Matematica (DICATAM), Università di Brescia, via Branze 43, 25123 Brescia, Italy
e-mail:alberto.salvadori@ing.unibs.it

1 Introduction

Fracturing processes reveal three distinct phases: loading without crack growth, stable and unstable crack propagation. Energy dissipation due to fracture growth takes place in the process region, in the plastic region outside the process region, and eventually in the wake of the process region. When a fracture process is idealized to infinitesimally small scale yielding, energy dissipation is concentrated at the crack front. Smoothness of the latter together with isotropic linear elasticity on a domain $\Omega \in \mathbb{R}^3$ is assumed in the present note, making use of Hooke's law without limitation of stress and strain magnitudes.

The material response to the following quasi-static external actions is sought: tractions $\overline{\mathbf{p}}(\mathbf{x})$ on $\Gamma_p \subset \partial\Omega$,

displacements $\overline{\mathbf{u}}(\mathbf{x})$ on $\Gamma_u \subset \partial\Omega$, bulk forces $\mathbf{f}(\mathbf{x})$ in Ω. External actions are all assumed to be *proportional*, i.e. that they vary only through multiplication by a time-dependent scalar $k(t)$, termed load factor, taken to be zero at initial time[1] $t = 0$ when the crack attained its initial length.[2] The stress-strain fields in the crack tip vicinity are uniquely determined by the stress intensity factors (SIFs) denoted as usual with K_1, K_2, K_3 and collected in vector $\mathbf{K}$.

Similarly to the determination of the elastic limit, the concept of incipient crack growth is difficult to identify: in both cases, the difficulty is solved by a convention. An *onset of crack propagation* and a *safe equilibrium domain* are governed theoretically by a local condition on $\mathbf{K}$, describing when the process region reaches a critical state which, in most cases of engineering interest, is independent of body, geometry and loading: this property is termed autonomy (see Barenblatt 1959).

The *global incremental quasi-static fracture propagation problem* consists in seeking an expression of the crack growth rate $\dot{l}(s,t)$ at a generic point s along the crack front in the presence of a variation of the load factor $\dot{k}$ for all three phases of a fracturing process. The problem can be framed in the mechanics of *standard dissipative systems* and posed in the following way: given the state of stress and the history of crack propagation (if any) at time t, express the crack propagation rate (if any) of the crack front due to a variation of the external actions as a function of the stress and of the history.

For two-dimensional linear elastic fracture mechanics, the problem was studied in Salvadori and Carini (2011), Salvadori (2008). A maximum dissipation principle at the crack tip during propagation was postulated. Associated flow rule and loading/unloading conditions in Kuhn–Tucker complementarity form descended. Consistency condition led to the formulation of an algorithm for crack advancing, which was driven by the increment of external actions and allowed the evaluation of crack length increment at the crack tip.

Such an approach will be here extended to three-dimensional fracture mechanics. The asymptotic expansion of the SIFs in power of the crack extension length in the normal plane to the crack front for the three-dimensional case is presented in Sect. 2. The expansion of SIFs can be put in analogy with Colonnetti's decomposition of stresses in plasticity as due to an elastic contribution and to a distortion (Colonnetti 1918, 1950).

Assuming that the maximum energy release rate (MERR) criteria models the onset of crack propagation and the safe equilibrium domain, three conclusions are stated: i) mixed mode crack propagation must satisfy mandatory requirement (19); ii) the transition between stable and unstable propagation regimes is ruled by condition (22); iii) crack front "velocity" minimizes linearly constrained quadratic functionals (38, 39).

Requirement (19) for mixed mode crack propagation is established in Sect. 3. It is detailed in Appendix A, under the assumption that between times $\tau > t$ and t *the principle of local symmetry applies* and the propagation is continuous in time, namely $K_2 = 0$ at all instants.

In Sect. 4, moving from the consistency condition of the plasticity analogy, condition (22) for stable crack growth and, inherently, for the transition to the unstable phase is derived. This transition between stable and unstable propagation regimes is a crucial information for the safety of a structural component, assuming that unstable propagation leads to structural collapse.

Section 5.1 is devoted to the non trivial proof of the symmetry property of the Gateaux derivative of the linear operator defined in Eq. (23). As a consequence of such a property, two variational statements are proved in Sect. 5.2, in the range of stable crack growth. They are reminiscent of Ceradini's theorems for plasticity and characterize velocity $\dot{l}(s,t)$ that solves the global quasi static fracture propagation problem at time t as the unique minimizer of linearly constrained quadratic functionals. Uniqueness is a consequence of SIFs expansion (1) and can be avoided only by using expansions for branched elongations.

The theoretical framework is tested on a simple problem, the axialsymmetric example of a penny-shaped crack subjected to uniform tensile stress, in Sect. 6. Results confirm the potential of the proposed incremental formulation.

2 Fracture propagation as a standard dissipative system

Notation: Consider within a three dimensional elastic body a crack of arbitrary shape as in Fig. 1, except that

[1] In what follows, "time" represents any variable which monotonically increases in the physical time and merely orders events; the mechanical phenomena to study are time-independent.

[2] This hypothesis is not fundamental and can easily be substituted with a smooth variation of SIFs around the critical state.

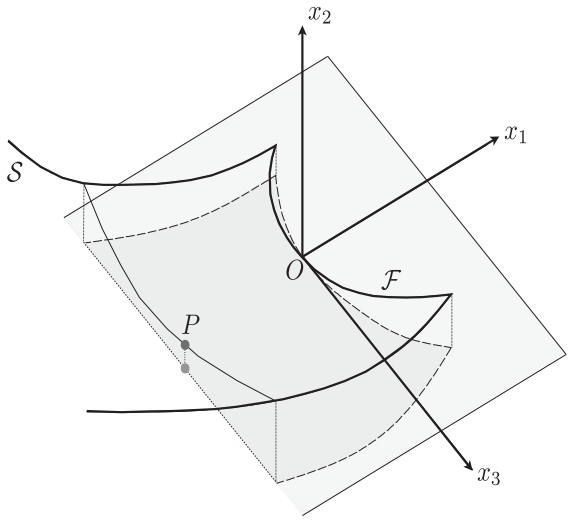

Fig. 1 Arbitrarily shaped crack in a three dimensional body. Reference $\{O, x_1, x_2, x_3\}$ is the Frenet frame at point O along the crack front $\mathcal{F}$. *Dark gray shadow* shows the projection of the crack surface $\mathcal{S}$ onto the tangent plane $\{O, x_1, x_3\}$ (enlighten in *light gray*)

both its surface $\mathcal{S}$ and front $\mathcal{F}$ are assumed to be of class C^∞, at least in the vicinity of $\mathcal{F}$ (Leblond 1999; Leblond et al. 1999).

Let O denote an arbitrary point on $\mathcal{F}$. Cartesian coordinates x_1, x_2, x_3 are attached to that point with Ox_1 in the tangent plane to $\mathcal{S}$ and orthogonal to $\mathcal{F}$, Ox_2 perpendicular to $\mathcal{S}$ and Ox_3 coincident with the tangent to $\mathcal{F}$. Denote with s the curvilinear length of O on $\mathcal{F}$, and with s' that of a generic point on that curve. Denote with P a generic point on $\mathcal{S}$. The local geometry of $\mathcal{S}$ will be described with a degree of accuracy such that distance $x_2(P)$ from P to its projection onto the tangent plane at O be specified up to order $O(x_1^2 + x_3^2)$. This is achieved as in Leblond (1999), Leblond et al. (1999) by prescribing the components C_{11}, C_{13}, C_{33} of the curvature tensor $\mathbf{C}$ of $\mathcal{S}$ at point O. The local geometry of $\mathcal{F}$ is then also described with a similar accuracy by prescribing the curvature Γ of its projection onto the tangent plane at O.

A small arbitrary deviated extension will be then added to the crack, as in Fig. 2. One makes the fundamental assumption that the original crack front belongs to both the old and new crack surfaces, i.e. the crack extension develops continuously and irreversibly from that original front. At each point s' of $\mathcal{F}$, the new tangent plane is obtained by rotating the old one about the local tangent to $\mathcal{F}$ by an angle $\theta(s')$ (the *kink angle*);

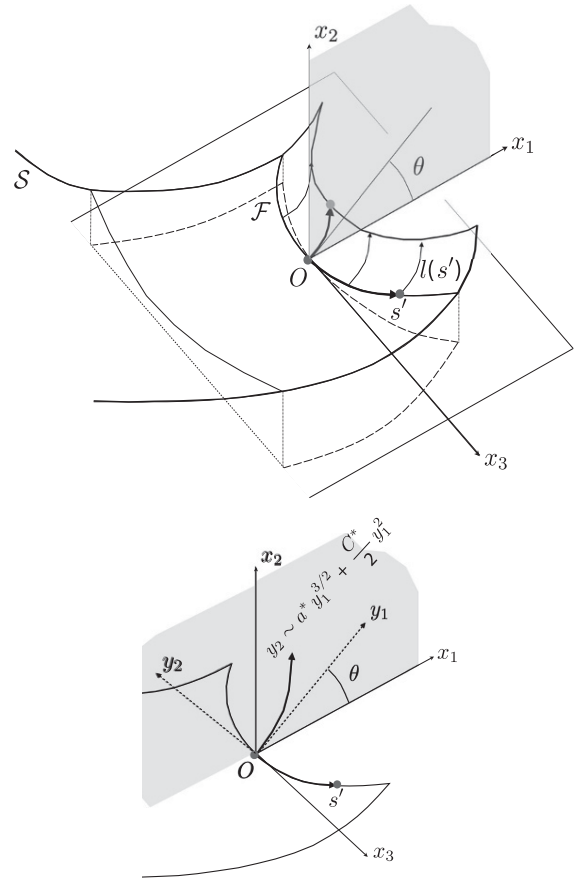

Fig. 2 Arbitrary crack with a small kinked and curved extension $l(s')$. Kink angle is denoted with θ and defined in the normal plane at the abscissa s along the crack front. The normal plane at the origin is highlighted in *gray*. Each kink is surrounded by an unbranched, curvilinear crack path in the normal plane. In agreement with Amestoy and Leblond (1992), Leblond (1999), Leblond et al. (1999), the elongation in the normal plane is taken to be: $y_2 = a^* y_1^{3/2} + \frac{1}{2} C^* y_1^2 + O(y_1^{5/2})$ with axis $\{y_1, y_2\}$ denoting the Frenet frame right after the kink

the value $\theta(s)$ of function $\theta(s')$ at point O will often be simply denoted with θ and the same convention will be used for other functions defined along the crack front. Function $\theta(s')$ will be assumed to be of class C^∞ with respect to s'. The present paper will not concern *how to determine the crack kinking angle* during crack growth. Such an issue was largely debated in the literature, including a few papers from one of the authors (Salvadori 2008, 2010). It suffices here to point out that given the geometry and the load history at time t, $\theta(s')$ is uniquely defined for any s', i.e. the problem of evaluating $\theta(s')$ is uncoupled with the evaluation of the crack extension.

Recall the formulation of the *global incremental quasi-static fracture propagation problem*, namely: given the state of stress and the history of crack propagation (if any) at time t, express the crack propagation rate (if any) of the crack front as a function of the stress and of the history.[3] To this aim, consider a "time" interval $]t, t^*]$ in which the crack front elongates at point s and an instant $\tau \in]t, t^*]$. Let $l(s', \tau) > 0$ denote the curvilinear length of the crack extension, as measured in the normal plane at s'. The function $l(s', \tau)$ will be assumed to be of class C^∞ with respect to s'. The value $l \equiv l(s, \tau)$ of $l(s', \tau)$ at point O will be assumed to be non-zero; this means that O will be supposed not to be an endpoint of the active, effectively propagating part of the front. The velocity of the elongation of the crack at time t, also called *crack tip quasi static velocity*, will be denoted with $\dot{l}(s', t)$. In view of the irreversibility hypothesis $\dot{l}(s', t) \geq 0$.

SIFs expansion: It will be assumed that $l(s', \tau)$ is of the form $\varepsilon(\tau)\eta(s')$ where $\varepsilon(\tau)$ is a small positive parameter and $\eta(s')$ a given, fixed non-negative function. The point of view that $\eta(s')$ is a given function is purely formal, since in reality that function is unknown a priori. Denoting with $\mathbf{K}(s', t) = \{K_1(s', t), K_2(s', t), K_3(s', t)\}$ the SIFs vector at any point s' along the crack front at time t, the expansion of $\mathbf{K}$ at time τ at the extended point in powers of $\varepsilon(\tau)$ was provided in Leblond (1999), Leblond et al. (1999). At point s it takes the form:

$$\mathbf{K}(s, \tau) = \mathbf{K}^*(s, \tau) + \mathbf{K}^{(1/2)}(s, \tau)\sqrt{\varepsilon(\tau)\eta(s)} + \mathbf{K}^{(1)}[k(\tau), \eta(s)]\,\varepsilon(\tau) + O(\varepsilon^{3/2}) \quad (1)$$

Terms $\mathbf{K}^*$, $\mathbf{K}^{(1/2)}$ are local factors at point s that depend linearly on the load factor k. They are given componentwise (using the Einstein summation convention) by:

$$K_i^* = F_{ij}(\theta)K_j \quad (2)$$

$$K_i^{(1/2)} = G_{ij}(\theta)T_j + a^* H_{ij}(\theta)K_j \quad (3)$$

Symbol $\mathbf{K}^{(1)}[\cdot, \cdot]$ denotes a non local operator that acts along the whole front $\mathcal{F}$, whose expression was provided in Leblond et al. (1999) moving from the cornerstone work of Rice (1989) on weight function theories: (see also Lazarus 2011)

$$K_i^{(1)}[k(\tau), \eta(s)] = \frac{1}{2}\left[\frac{\partial^2 K_i}{\partial(\sqrt{l})^2}\right]_{l(s')\equiv l(s)}^{C^*=0} \eta(s) + C^* M_{ij}(\theta)\, K_j\, \eta(s) + N_{ij}(\theta)\, K_j\, \frac{\partial \eta}{\partial s} + \fint_{\mathcal{F}} Z_{ij}(\Omega, s, s', \theta(s), \theta(s'))\, K_j(s')\, \big(\eta(s') - \eta(s)\big)\, ds' \quad (4)$$

In these equations the $K_j s$ are the SIFs prior to the crack kinking, $\mathbf{T}$ is the *T-stress vector* of components T_j and a^* and C^* define the curvature of the extension. The $F_{ij}s$, $G_{ij}s$, $H_{ij}s$, $M_{ij}s$ and $N_{ij}s$ are functions of the kink angle $\theta(s)$, which are termed *universal* because they obey to the autonomy concept; on the contrary the terms $\frac{1}{2}\left[\frac{\partial^2 k_i}{\partial(\sqrt{l})^2}\right]_{l(s')\equiv l(s)}^{C^*=0}$ and $Z_{ij}s$ depend on the geometry of body Ω. Symbol $\fint_{\mathcal{F}}$ stands for the Cauchy Principal Value along $\mathcal{F}$. Expansion (1) details the behaviour of SIFs along crack front $\mathcal{F}$ due to an irreversible change in the geometry of the same front $\mathcal{F}$.

Onset of crack propagation and criteria for crack kinking angle evaluation: The mathematical representation of the *onset of crack propagation* at point s and time t can be given a general form:

$$\varphi(\mathbf{K}, \theta) = \vartheta(\mathbf{K}, \theta) - \vartheta(K_1^C, \theta^C) = 0 \quad (5)$$

in the normal plane of the Frenet reference defined in Fig. 1. In Eq. (5) K_1^C is the fracture toughness and $\theta^C = 0$ is the propagation angle in a mode 1 experimental test. For each φ, there is a "related magnitude" ϑ which increases monotonically with the level k of applied loads and which is supposed to obtain a critical value at the onset of crack growth (Salvadori 2008). Specific examples for ϑ are: i) Maximum Energy Release Rate (shortened in MERR) G in incipient crack growth; ii) maximum hoop tensile stress in the $r^{-1/2}$ near-tip singular field. Cracks cannot advance at "time" t if

$$\varphi(\mathbf{K}, \theta) < 0 \quad (6)$$

The latter inequality defines the *safe equilibrium domain*.

As seen in Eq. (5) and discussed in Salvadori (2010) the onset of crack propagation is always related to a prediction of the kinking angle θ in the eventuality of a crack elongation. The safety of a configuration

[3] The global incremental quasi-static fracture propagation problem at "time" t therefore depends on the geometry at "time" t as well as on the external actions that drive the problem itself. In the present note, external loads are merely considered. Propagation by reason of temperature or environmentally driven are not taken into account.

at *time* t, no matter how far away it is from the critical state, depends on the angle the crack is going to kink at the time it grows. Historically, any onset of crack propagation has been provided with a *criteria for crack kinking evaluation*. With the mere exception of the local symmetry (Goldstein and Salganik 1974), the two notions of onset of crack propagation and criteria for crack kinking evaluation[4] correspond one another in the general formulation

$$find \ \{k, \theta\} \quad s.t. \ \ \varphi = 0\,, \quad \frac{\partial \varphi}{\partial \theta} = 0 \tag{7}$$

derived in Sect. 2.2 in Salvadori (2010).

There are two main "streams" of literature on crack kinking criteria, that correspond to two different views. An approach assumes that a crack propagates "as soon as it can", which means at the lowest values of external actions that allow the onset of propagation to be reached. In this view: (1) formula (7) insures that safety is measured against the worst value of θ for the MERR; (2) it was shown in Salvadori (2010) that crack propagation criteria different from the MERR must be ruled out and attempts of capturing crack growth by means of a propagation criteria different from the MERR should not not be undertaken in the Griffith standpoint of fracture. This strong statement includes the Local Symmetry (henceforth shortened as LS) principle at least in mixed mode. In particular, this implies that the view of kinking $K_2^*(t^-) = 0$ by means of LS (Goldstein and Salganik 1974) criteria is incorrect, as it requires a value for k higher than the one predicted by (7). Assuming that a crack propagates at minimum external actions implies that the energy stored in the system is minimal as well. This seems to be in the line of the revisitation of brittle fracture proposed in Francfort and Marigo (1998), provided that the onset of propagation is reached. Moving from energy minimality principle, Chambolle et al. (2009) shown that the kinking never occurs with a propagation which is continuous in time (see Property 4 in Chambolle et al. 2009). Analysis made in Salvadori (2010) confirm this statement in the light of formula (7).

On the other hand, it seems intuitive to consider a smooth in space and continuous in time crack evolution at least after the kink. It has been proved (Chambolle et al. 2009) that condition $K_2^* = 0$ is mandatory for any continuous propagation in time.[5] Therefore, denoting with t the time at which the crack kinks, a continuous propagation in time after the kink requires $K_2^*(t^+) = 0$. Assuming that such a condition holds also at t^- (and thus a continuous in time propagation at a kink) leads to the view of LS criteria for crack kinking, which still makes use of MERR as an onset of propagation. This view is not compatible with formula (7), in the sense that the angle of propagation is different and also the load at which crack elongates is larger than the one required by formula (7). Assuming LS at a kink implies accepting that the energy inserted in the system to propagate the crack is not the minimal one, i.e. that some energy barriers must be present at a kink.

Linking conditions at t^- and $K_2^*(t^+) = 0$ is an open, challenging problem in the fracture mechanics community. The choice made by mother nature has been not understood clearly so far. In 2D, LS and MERR criterion are so close one another that differences are far below the accuracy of experimental analysis. As envisaged in Appendix B, the scenario in 3D propagation appears to be different so that experimental campaigns might be planned. Solving this controversy is in any case far out of the scope of the present note, which naturally develops in the framework of continuous propagation at kinks.

The latter condition is not mandatory, however. Throughout the paper it is assumed that the load factor has no requirements apart being non negative, $k(t) \geq 0$. In other terms, $k(t)$ may increase with t, decrease at some point, eventually jump so that the crack proceeds quasi-statically and continuously in time,[6] keeping the system at the onset of propagation (5). Starting from $K_2^*(t^+) = 0$ a continuous evolution in time along a smooth crack path may proceed. The crack evolution at any $\tau > t$ is stable with respect to $k(t^+)$ but not with respect to $k(t^-)$ if MERR is assumed at a kink. Of course, in realistic analysis, the jump $k(t^-) - k(t^+)$ in the load factor is unfeasible. A dynamic evolution

[4] Appendix B entails a short decryption of most relevant onsets and criterions.

[5] See property 3 in Chambolle et al. (2009), namely: assuming the validity of the Energy Conservation and Stability Criterion, a crack cannot propagate continuously in space and time in a homogeneous, isotropic material unless it propagates in mode I.

[6] Obviously, an arbitrary behavior of $k(t)$ may be not "realistic", in the usual sense of k being a control parameter tuned within a laboratory. Yet, the approach is not uncommon in mechanics. Essentially, it is the same assumption that leads to model snapping and buckling problems, allowing to follow incrementally the equilibrium path after a peak (Riks 1979).

is therefore localized at any crack kinking, similarly to what happens in a snap through experiment. The crack length at which a quasi-statical evolution may eventually start again corresponds to a time $\tau > t$ at which $k(\tau) = k(t^-)$, if a unique crack evolution takes place.

Maximum energy release rate: When cracks—idealized to infinitesimally small scale yielding—advance, energy dissipation is concentrated at the crack fronts. Whereas in plasticity the choice of a yield function is free and the relevant amount of dissipation descends, Irwin (1958) formula in the Griffith standpoint of fracture restricts the choice of the onset of crack propagation φ to the MERR:[7]

$$\varphi = \frac{1-\nu^2}{E}(K_1^{*2} + K_2^{*2}) + \frac{1+\nu}{E}K_3^{*2} - G_C \qquad (8)$$

where G_C is the fracture energy, i.e. the dissipated energy per unit crack elongation. The expression of the energy release rate G is referred to an embedded crack in which prevails a plane strain condition in a core of the crack front. The radius of this core is a function of the distance to the free surface.

Recent investigations (Chambolle et al. 2010) revisited the notion of MERR in 2D and proved formula (8) also by means of the l'Hopital theorem (Salvadori and Giacomini 2012). Similar analysis in 3D have not been carried out and the validity of Irwin's formula (8) for 3D, widely accepted in the fracture mechanics community, is here assumed.

Phases of the fracturing process: If $\varphi < 0$ at "time" t, a "sufficiently small" load increment δk between instants t and $\tau > t$ exists that does not elongate the crack:

$$\begin{aligned}
&\text{at any } t \text{ s.t. } \varphi(\mathbf{K}(s,t),\theta) < 0 \text{ for all } \theta \text{ it exists } \delta k \\
&\quad = k(\tau) - k(t) > 0 \text{ s.t.} \\
&\delta\mathbf{K}(s,\tau) = \frac{\mathbf{K}(s,t)}{k(t)}\delta k \quad \text{and} \quad \varphi(\mathbf{K}(s,t)+\delta\mathbf{K}(s,\tau),\theta) \\
&< 0 \text{ for all } \theta
\end{aligned} \qquad (9)$$

Such an incremental process describes the first phase of the fracturing process, namely loading without crack growth. When the onset of crack propagation is reached at a point s at time t, the second phase, when present, is triggered off: stable crack growth. A further increase of load δk causes crack elongation at s. Denoting with $\delta\mathbf{K} = \mathbf{K}(s,\tau) - \mathbf{K}(s,t^+)$, $\delta\mathbf{K}^* = \mathbf{K}^*(s,\tau) - \mathbf{K}^*(s,t^+)$

$$\begin{aligned}
&\text{at time } t \text{ s.t. } \varphi(\mathbf{K}(s,t),\theta(s,t)) = 0 \\
&\text{for at least one } \theta(s,t) \text{ it exists } \delta k = k(\tau) \\
&\quad - k(t^+) > 0 \text{ s.t.} \\
&\delta\mathbf{K}^* = \frac{\mathbf{K}^*(s,t^+)}{k(t^+)}\delta k \\
&\delta\mathbf{K} = \delta\mathbf{K}^*(s,\tau) + \mathbf{K}^{(1/2)}(s,t)\sqrt{l(s,t)} + O(l)
\end{aligned} \qquad (10)$$

All details of Eq. (10) will be discussed in Sect. 3 and Appendix A. Conceptually, it states that a quasi-static fracture extension $l(s,t)$ due to external actions requires a contemporary variation δk such that the global equilibrium is guaranteed. It is a reminiscence of Colonnetti's decomposition of stresses in plasticity (Colonnetti 1918, 1950), as the variation of SIFs is additively decomposed as due to an elastic contribution ($\delta\mathbf{K}^*$) and to a distortion (in fracture: crack elongation l; in plasticity: plastic strain rate) which reverses itself into SIFs (stresses in plasticity) by means of a stiffness factor (in fracture: $\mathbf{K}^{(1/2)}$, in plasticity: the action of the $\mathbb{Z}$ matrix over the plastic part of the volume). Equation (10) states also implicitly that the extension $l(s,t)$ cannot be arbitrary along the crack front. Equilibrium, in the sense that δk is unique for all points s, requires $l(s,t)$ to assume a precise shape with respect to s. Such a constraint is provided in plasticity by Ceradini's functional which in fact will be extended to fracture in Sect. 5.2.

The third phase of crack propagation, unstable crack growth, is reached when condition $\delta k > 0$ in Eq. (10) is no longer required at some point s. A quasi-static growth, merely academic, can be simulated only with a decrease of external action $\delta k \leq 0$ to recover the post peak behaviour as typical in "arc length" techniques for softening materials.

In reality, dynamics effects come into play. They fall out of the scope of the present note.

Fracture propagation as a standard dissipative system: In the Griffith theory (see Griffith 1921 but also its review in Bourdin et al. 2008) and in the light of Irwin's formula (Irwin 1958), propagation is governed *at time t* by the following conditions, reminiscence of Kuhn-Tucker conditions of plasticity:

$$\begin{aligned}
&\varphi(\mathbf{K}(s,t),\theta(s,t)) \leq 0, \quad \dot{l}(s,t) \geq 0, \\
&\varphi(\mathbf{K}(s,t),\theta(s,t))\,\dot{l}(s,t) = 0
\end{aligned} \qquad (11)$$

Conditions (11) can be derived on a thermodynamical basis as done in Salvadori and Carini (2011) for two-dimensional LEFM. Moving from a rigid-plasticity

[7] In Eq. (8), E is Young modulus and ν Poisson's coefficient.

analogy between SIFs and stresses, crack propagation induces a dissipation which satisfies Clausius–Duhem's inequality through the introduction of a convex dissipation potential, $\mathcal{D}$. Such an idea extends straightforwardly to three dimensional LEFM simply considering any point s along the crack front as a locus of possible dissipation. A "safe equilibrium domain" is defined as:

$$\mathbb{E} = \{\{K_1^*, K_2^*, K_3^*\} \in \mathbb{R}_0^+ \times \mathbb{R} \times \mathbb{R} \mid \varphi(K_1^*, K_2^*, K_3^*) < 0\} \tag{12}$$

It has a "local" nature at each point s of the crack front. A curve "onset of crack propagation" $\partial\mathbb{E}$ as the boundary of the "safe equilibrium domain" can be defined at any point s:

$$\partial\mathbb{E} = \{\{K_1^*, K_2^*, K_3^*\} \in \mathbb{R}_0^+ \times \mathbb{R} \times \mathbb{R} \mid \varphi(K_1^*, K_2^*, K_3^*) = 0\} \tag{13}$$

When $\mathbf{K}^*(s,t) \in \mathbb{E}$ no elongation occurs, eventually corresponding to an elastic unloading. Vectors $\mathbf{K}^*(s,t) \notin \overline{\mathbb{E}}$ are ruled out. The definitions above are reminiscent to the elastic domain and to the yield surface in the plasticity theory (Han and Reddy 1999). They implicitly label the SIFs vector $\mathbf{K}^*$ as an internal force for the LEFM problem.

Mechanical dissipation in LEFM is due to the irreversible nature of the crack extension, measured in an incremental setting by the quasi-static crack tip velocity vector $\dot{\boldsymbol{l}}(s,t)$. It is defined in the Frenet frame at point s of $\mathcal{F}$ as the vector, slanted by kinking angle $\theta(s,t)$, whose norm is equal to the quasi static crack tip velocity $\dot{l}(s,t)$.

The internal variable conjugated to $\mathbf{K}^*$ is here termed "dissipation rate vector" $\dot{\mathbf{a}}^*$ and unfortunately cannot coincide with $\dot{\boldsymbol{l}}(s,t)$. Its modulus is related to $\dot{l}(s,t)$ by the following identity, which is part of the definition of $\dot{\mathbf{a}}^*$:

$$\|\dot{\mathbf{a}}^*\| = \frac{G_C}{K_1^C}\,\dot{l} \tag{14}$$

A "maximum dissipation" principle for LEFM is postulated at any point s in terms of the dissipation function $\mathcal{D}(\mathbf{k}^*; \dot{\mathbf{a}}^*) = \mathbf{k}^* \cdot \dot{\mathbf{a}}^*$ among all possible SIFs $\mathbf{k}^* \in \overline{\mathbb{E}}$. It leads to the associative flow rule:

$$\dot{\mathbf{a}}^* = \frac{\partial \varphi}{\partial \mathbf{K}^*}\dot{\lambda} \tag{15}$$

and loading/unloading conditions in Kuhn–Tucker complementarity form at any point s along the crack front:

$$\dot{\lambda} \geq 0, \qquad \varphi \leq 0, \qquad \dot{\lambda}\,\varphi = 0 \tag{16}$$

Analogously to 2D, it can be proved in view of definition (14) that $\dot{\lambda} = \dot{l}$, thus Kuhn-Tucker complementarity Eq. (11) are recovered from (16). *Consistency condition* finally read:

$$\text{When } \varphi = 0, \quad \dot{l} \geq 0, \qquad \dot{\varphi} \leq 0, \quad \dot{l}\dot{\varphi} = 0 \tag{17}$$

3 On the mixed mode propagation

Expansion (1) at an instant $\tau \in]t, t^*]$ may be rewritten in view of (2–4) as:

$$\begin{aligned}\mathbf{K}(s,\tau) = \mathbf{K}^*(s,\tau) &+ \mathbf{K}^{(1/2)}(s,\tau)\sqrt{l(s,\tau)} \\ &+ \mathbf{K}_0^{(1)}\, l(s,\tau) + \mathbf{K}_1^{(1)}\,\frac{\partial l}{\partial s}(s,\tau) \\ &+ \mathbf{K}_{nl}^{(1)}[l(s',\tau) - l(s,\tau)] + o(l)\end{aligned} \tag{18}$$

Vector

$$\mathbf{K}_0^{(1)} = \frac{1}{2}\left[\frac{\partial^2 \mathbf{K}(s,l)}{\partial(\sqrt{l})^2}\right]_{l(s')=l(s)}^{C^*=0} + C^*(s)\mathbf{M}(\theta(s))\,\mathbf{K}(s)$$

accounts for the locally linear contribution of $l(s,t)$ to the variation of SIFs at s, whereas vector

$$\mathbf{K}_1^{(1)} = \mathbf{N}(\theta(s))\,\mathbf{K}(s)$$

conveys the influence of derivative of l with respect to the abscissa s on the crack front. Finally

$$\mathbf{K}_{nl}^{(1)}[f(s')] = \fint_{\mathcal{F}} \mathbf{Z}(\Omega, s, s', \theta(s), \theta(s'))\,\mathbf{K}(s') f(s') ds'$$

is the non local operator that provides, once applied to $l(s',\tau) - l(s,\tau)$ the contribution of the fluctuation of crack advancing at s' to SIFs at s.

During stable crack growth, propagation is meant to be a sequence of equilibrium states. At each load $k(\tau)$ a geometry configuration $l(s,\tau)$ corresponds which eventually evolves quasi-statically, keeping crack front $\mathcal{F}$ at the onset of crack propagation $\mathbf{K}^*(s,\tau) \in \partial\mathbb{E}$. In any "time" interval $]t, t^*]$ in which the crack front elongates at point s one has:

$$\begin{aligned}\varphi(\mathbf{K}^*(\tau)) - \varphi(\mathbf{K}^*(t)) = &\frac{1-\nu^2}{E}(K_1^{*2}(\tau) + K_2^{*2}(\tau)) \\ &+ \frac{1+\nu}{E}K_3^{*2}(\tau) \\ &- \frac{1-\nu^2}{E}(K_1^{*2}(t) + K_2^{*2}(t)) \\ &- \frac{1+\nu}{E}K_3^{*2}(t) = 0 \;\tau \in]t, t^*]\end{aligned}$$

in view of (8). *Assuming that the Principle of Local Symmetry is the criteria for crack kinking and that the elongation is a curve at least of class C^1*, in the limit $\tau \to t^+$, point s on the crack front never propagates unless:

$$\frac{\partial \varphi}{\partial \mathbf{K}^*}(s,t) \cdot \mathbf{K}^{(1/2)}(s,t) = 0 \tag{19}$$

which appears to be a mandatory requirement for mixed mode crack propagation. Proof is detailed in Appendix A. Outcome (19) together with the required assumptions will be taken in the rest of the paper.

With the aim of using SIFs expansion (1) with respect to to the values "before the kink" at time t, the assumption is made that the crack path in the normal plane $\{O, x_1, x_2\}$ at any $s \in \mathcal{F}$ is smooth[8] in the "time" interval $]t, t^*]$. Then the kinking angle $\theta(s,\tau) = 0$ and $\mathbf{F} = \mathbf{1}$ in expansion (1) for all $\tau > t$. This in turn implies $K_2^* = K_2 = 0$ for all $\tau > t$ and in fact also $a^* = 0$ at time t.

The latter proposition is easily verified: all functions G_{ij} vanish at $\theta = 0$ in Eq. (3), that reduces in this case to

$$K_i^{(1/2)} = a^* \;\; H_{ij}\big|_{\theta=0} \;\; K_j$$

for all instants $\tau > t$. For being $K_2 = K_2^* = 0$, expansion (1) implies $K_2^{(1/2)} = a^* \;\; H_{21}|_{\theta=0} \;\; K_1 = 0$. As $H_{21}|_{\theta=0} = 3/4$, the geometrical restriction $a^* = 0$ comes out for all instants $\tau > t$. As a^* at time t is defined in the Frenet frame "right after the kink" and as in the limit $\tau \to t$ the Frenet frame $\{O, y_1, y_2\}$ at time τ converges to the one at time t, so does a^*. One concludes therefore that $a^* = 0$ at time t as well.

Property (19) has been proven in Appendix A making use of the MERR onset of propagation. The proof of vanishing a^* at time t stems merely on the criteria of local symmetry. Onset and criterion convey separate outcomes. Typically, it is not possible to obtain $a^* = 0$ via the Griffith approach, in view of the fact that (19) holds for all a^* if $\theta = 0$. On the other hand, the proof in Appendix A shows that the crack tip velocity at the kink must be zero, which cannot be proved by the local symmetry criterion.

If a non smooth propagation after the kink is allowed, unfortunately $\mathbf{F} \neq \mathbf{1}$ and it is not possible to use expansion (1) with respect to the values "before the kink" at time t.

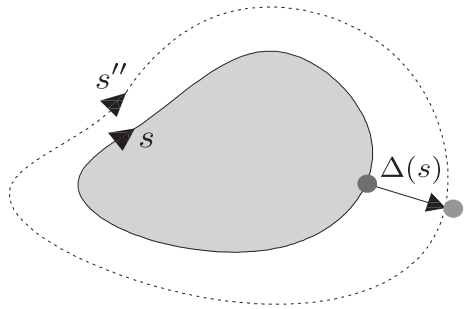

Fig. 3 Finite propagation length $\Delta(s)$. The *shadowed area* represents the crack surface before elongation

4 A stability condition

Provided that mandatory condition (19) holds, the "time" derivative of function $\varphi(\mathbf{K}^*(s,t), \theta(s,t))$ can be properly defined. In view of consistency condition (17) and of definition (8) for φ one writes:

$$\dot{\varphi}(s,t) = \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \left(\dot{\mathbf{K}}^* + \mathbf{L}[\dot{l}(s,t)]\right) = G_c \frac{\dot{k}(t)}{k(t)} + \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)] = 0 \tag{20}$$

at a point s along the crack front at which $\varphi = 0$ and $\dot{l}(s,t) > 0$. In Eq. (20) $\mathbf{L}[\cdot]$ is a linear operator defined by:

$$\mathbf{L}[\dot{l}(s,t)] = \mathbf{K}_0^{(1)} \dot{l}(s,t) + \mathbf{K}_1^{(1)} \frac{\partial \dot{l}}{\partial s}(s,t) + \mathbf{K}_{nl}^{(1)} [\dot{l}(s',t) - \dot{l}(s,t)] \tag{21}$$

in view of expansion (18). Note that Eq. (20) is defined at time t, i.e. there is no reason to consider a further instant τ after t.

In Colonnetti's framework, $\dot{\mathbf{K}}^*(t)$ is a mere elastic contribution to $\dot{\mathbf{K}}(t)$ due to $\dot{k}(t)$ and $\mathbf{L}[\dot{l}(s,t)]$ corresponds to the crack elongation rate $\dot{l}(s,t)$ considered as an inelastic distortion. Consistency condition yields:

$$\dot{\varphi}\, \dot{l} = \left(G_c \frac{\dot{k}(t)}{k(t)} + \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)] \right) \dot{l}(s,t) = 0$$

[8] Such a smoothness implies that each kink is surrounded by an unbranched, curvilinear crack path in the normal plane $\{O, x_1, x_2\}$ of Fig. 2. In agreement with Leblond and coworkers notation (Amestoy and Leblond 1992; Leblond 1999; Leblond et al. 1999), the elongation in the normal plane is taken to be:

$$y_2 = a^*(s)\, y_1^{3/2} + \frac{1}{2} C^*(s)\, y_1^2 + O(y_1^{5/2})$$

with axis $\{y_1, y_2\}$ denoting the Frenet frame right after the kink, as in Fig. 2b.

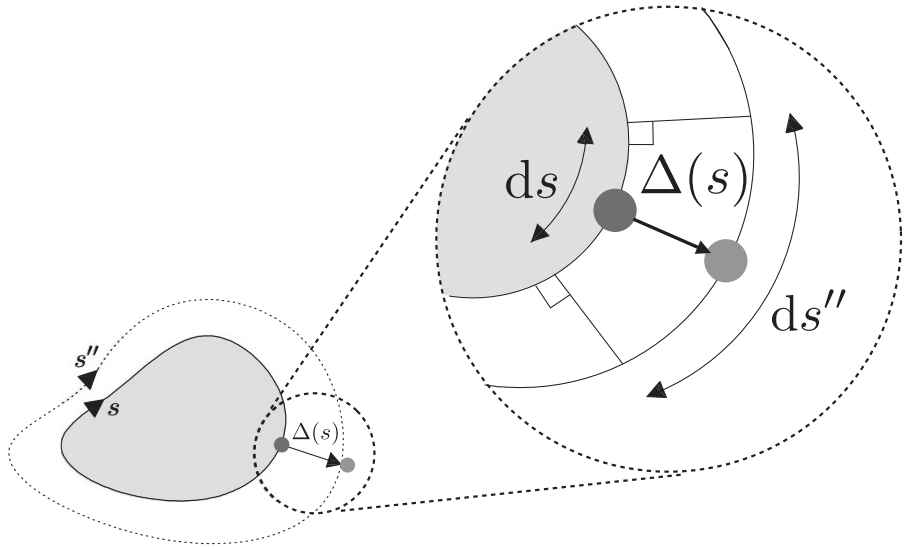

Fig. 4 A zoom of Fig. 3 allows to highlight the crack front length increment ds''

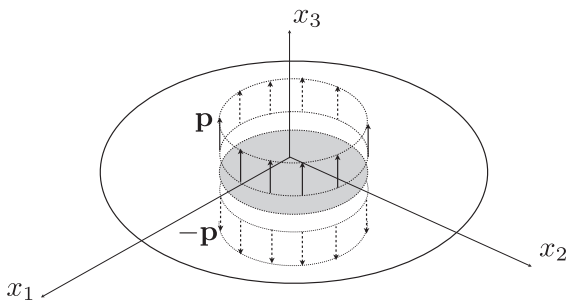

Fig. 5 Penny shaped crack of variable radius $R(t)$ in un unbounded linear elastic media, subject to a uniform tensile stress $\mathbf{p} = -k(t)\mathbf{n}$ on a constant circular area of radius R_{imp} here depicted in *gray*. $\mathbf{n}$ stands for the outer normal, so that $\mathbf{p}$ "opens the crack"

whence:

$$G_c \frac{\dot{k}(t)}{k(t)} \dot{l}(s,t) = -\frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)]\, \dot{l}(s,t)$$

The latter sets a condition for stable (i.e. $\dot{k}(t) > 0$) crack growth $\dot{l}(s,t) > 0$ at any point s along the crack front:

$$\dot{k}(t) > 0 \rightarrow \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)] < 0 \text{ at all } s \in \mathcal{F}|_{\varphi=0} \qquad (22)$$

Inherently, Eq. (22) is the (local) condition for the transition to the unstable phase at a point s. In other words, when condition (22) is not met at point s, an unstable propagation may take place in a neighborhood of s.

5 Variational formulation

5.1 A property of symmetry

Consider the affine operator $N[\cdot]$

$$N[\dot{l}] = \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \dot{\mathbf{K}}^* + \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)] \qquad (23)$$

The associated problem *find* $\dot{l} \geq 0$ *s.t.* $N[\dot{l}] = 0$ is equivalent to consistency conditions (17, 20). From well known theorems of variational calculus, in order

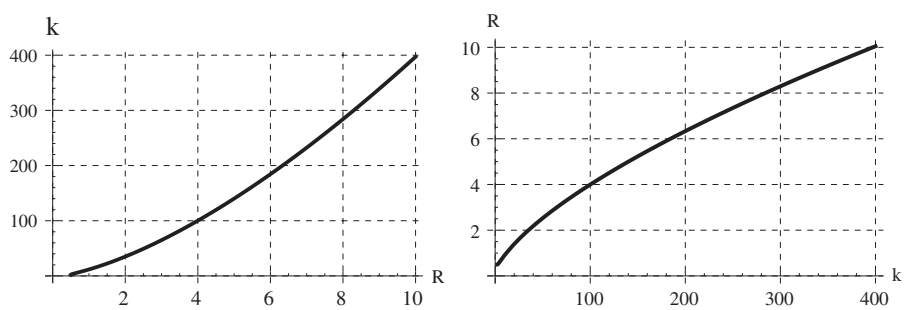

Fig. 6 Curves (45), (46) in the case $R_{imp} = 1/2$, $K_1^C = 1$. They confirm the stable nature of propagation

to give a variational formulation in "a restricted sense" (see Tonti 1984) to the nonlinear problem $N[\dot{l}] = 0$, it is necessary that the Gateaux derivative of N, defined by virtue of $\varepsilon \in \mathbb{R}$ as:

$$N'_{\dot{l}}[\dot{w}] = \left.\frac{dN[\dot{l} + \varepsilon\dot{w}]}{d\varepsilon}\right|_{\varepsilon=0} = \frac{\partial\varphi}{\partial\mathbf{K}^*} \cdot \mathbf{L}[\dot{w}] \tag{24}$$

is symmetric with respect to the usual bilinear form, namely:

$$\int_{\mathcal{F}} N'_{\dot{l}}[\dot{w}]\dot{z}\ \mathrm{d}s = \int_{\mathcal{F}} N'_{\dot{l}}[\dot{z}]\dot{w}\ \mathrm{d}s \tag{25}$$

By noting that φ has been defined as the maximum energy release rate criterion in Eq. (8), it turns out that $N'_{\dot{l}}[w]$ is precisely the energy release rate associated to elongation w at constant boundary conditions, the term $\frac{\partial\varphi}{\partial\mathbf{K}^*} \cdot \dot{\mathbf{K}}^*$ in definition (23) being the variation of energy due to the variation of external loads $k(t)$, i.e.

$$\dot{\mathbf{K}}^* = \mathbf{K}^* \frac{\dot{k}}{k}$$

It is not straightforward to envisage symmetry for $N'_{\dot{l}}$ from definition (21) of linear operator $\mathbf{L}$, although this property seems quite natural if one thinks the energy release rate as the derivative of the energy. On the contrary, term by term unsymmetry is apparent and one is lead to erroneously conclude that a variational formulation for problem $N[\dot{l}] = 0$ cannot be given. To prove symmetry for operator $N'_{\dot{l}}$ one must follow a different path of reasoning, based on the physical meaning of the operator itself.

Consider a finite propagation length $l(s)$ at a point s of the crack front, given by the sum of two contributions:

$$\Delta(s) = A\phi(s) + B\psi(s) \tag{26}$$

where A and B are two positive parameters and $\phi(s)$ and $\psi(s)$ are formally two fixed and assigned non negative functions expressing the elongation of the front in the normal plane.

The total (elastic) energy of body Ω, denoted henceforth by W, is a function of the geometry, namely of the advancing of the crack front which is uniquely defined by parameters A and B. Then $W = W(A, B)$ and apart from higher order terms one has:

$$\frac{\partial W(A, B)}{\partial A} = -\int_{\mathcal{F}} G(A, B, s)\ \phi(s)\ \mathrm{d}s'' \tag{27}$$

In Eq. (27) term $G(A, B, s)$ represents the *energy release rate* which is function of the location of the crack front (A and B), but also of the position s along the crack front. Furthermore, if $\mathrm{d}s$ refers to the front before the elongation $\Delta(s)$ and $\mathrm{d}s''$ to the final position of the front, it turns out:

$$\mathrm{d}s'' = \mathrm{d}s(1 + \gamma\ \Delta(s)) \tag{28}$$

where γ is a local corrective factor that depends upon the curvature of $\mathcal{F}$. Fig. 3 and 4

Second order terms in A and B arise during propagation, as the normal at point s changes direction at s''. Propagation in the normal plane modifies to:

$$\phi(s)\mathbf{n}(s) \cdot \mathbf{n}(s'') = \phi(s) + \text{ second order terms} \tag{29}$$

For all these reasons, Eq. (27) can be written as:

$$\frac{\partial W(A, B)}{\partial A} = -\int_{\mathcal{F}} G(A, B, s)\phi(s) [1 + \gamma(A\phi(s) + B\psi(s))]\ \mathrm{d}s \tag{30}$$

from which:

$$\frac{\partial^2 W(A, B)}{\partial B \partial A} = -\int_{\mathcal{F}} \frac{\partial G(A, B, s)}{\partial B}\phi(s) [1 + \gamma(A\phi(s) + B\psi(s))]\ \mathrm{d}s - \int_{\mathcal{F}} G(A, B, s)\phi(s)\gamma\psi(s)\ \mathrm{d}s$$

Setting $A = B = 0$ the latter becomes:

$$\left.\frac{\partial^2 W}{\partial B \partial A}\right|_{A=B=0} = -\int_{\mathcal{F}} \left.\frac{\partial G(A, B, s)}{\partial B}\right|_{A=B=0} \phi(s)\ \mathrm{d}s - \int_{\mathcal{F}} G(0, 0, s)\ \gamma\ \phi(s)\ \psi(s)\ \mathrm{d}s \tag{31}$$

Analogously one has:

$$\frac{\partial W(A, B)}{\partial B} = -\int_{\mathcal{F}} G(A, B, s)\psi(s)[1 + \gamma(A\phi(s) + B\psi(s))]\,\mathrm{d}s$$

and setting again $A = B = 0$ one has:

$$\left.\frac{\partial^2 W}{\partial A \partial B}\right|_{A=B=0} = -\int_{\mathcal{F}} \left.\frac{\partial G(A, B, s)}{\partial A}\right|_{A=B=0} \psi(s)\,\mathrm{d}s - \int_{\mathcal{F}} G(0, 0, s)\ \gamma\ \phi(s)\ \psi(s)\,\mathrm{d}s \tag{32}$$

Comparing Eqs. (31) and (32) one states:

$$\int_{\mathcal{F}} \frac{\partial G(s)}{\partial A}\psi(s)\,\mathrm{d}s = \int_{\mathcal{F}} \frac{\partial G(s)}{\partial B}\phi(s)\,\mathrm{d}s \tag{33}$$

Define infinitesimal increments of propagation length $l_1(s) = \delta A \cdot \phi(s), l_2(s) = \delta B \cdot \psi(s)$ and the corresponding infinitesimal variation of the energy release rate $\delta_1 G(s) = \frac{\partial G(s)}{\partial A}\delta A$, $\delta_2 G(s) = \frac{\partial G(s)}{\partial B}\delta B$. It holds:

$$\int_{\mathcal{F}} \delta_1 G(s) \cdot l_2(s)\, \mathrm{d}s = \delta A\, \delta B \int_{\mathcal{F}} \frac{\partial G(s)}{\partial A}\, \psi(s)\, \mathrm{d}s \quad (34)$$

and

$$\int_{\mathcal{F}} \delta_2 G(s) \cdot l_1(s)\, \mathrm{d}s = \delta A \delta B \int_{\mathcal{F}} \frac{\partial G(s)}{\partial B}\phi(s)\, \mathrm{d}s \quad (35)$$

Owing to Eq. (33) one concludes:

$$\int_{\mathcal{F}} \delta_1 G(s)\, l_2(s)\, \mathrm{d}s = \int_{\mathcal{F}} \delta_2 G(s)\, l_1(s)\, \mathrm{d}s \quad (36)$$

As it was pointed out that $N_i'[w]$ is precisely the energy release rate associated to elongation w, Eq. (36) is nothing but the proof of symmetry property (25).

5.2 Variational statements

In view of the symmetry property (25), the following two variational statements can be given. They are reminiscent of Ceradini's theorems (Ceradini 1965) for plasticity and hold under the assumption of stable crack growth

$$\frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\dot{l}(s,t)] < 0 \quad \text{for all } \dot{l}(s,t) \geq 0 \quad \text{at any point } s \in \mathcal{F}|_{\varphi=0} \quad (37)$$

Proposition 1 *Under hypothesis* (37)*, the crack tip "velocity" $\dot{l}(s,t)$ that solves the global quasi-static fracture propagation problem at "time" t minimizes the functional:*

$$\chi[\, v(s,t)\,] = -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[v(s,t)]\; v(s,t)\, \mathrm{d}s - \int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \dot{\mathbf{K}}^*\, v(s,t)\, \mathrm{d}s \quad (38)$$

under the constraint $v(s,t) \geq 0$.

To prove the theorem, denote with $v(s,t) = \dot{l}(s,t) + \Delta\dot{l}(s,t)$. Omitting the variable dependence on s and t for the sake of clearness, one writes:

$$\chi[\, v\,] - \chi[\, \dot{l}\,] = -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \left(\mathbf{L}[\dot{l}]\; \Delta\dot{l} + \mathbf{L}[\Delta\dot{l}]\; \dot{l} + \mathbf{L}[\Delta\dot{l}]\; \Delta\dot{l}\right) \mathrm{d}s - \int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \dot{\mathbf{K}}^*\; \Delta\dot{l}\, \mathrm{d}s$$

Owing to the symmetry property (36), in view of the linearity of $\mathbf{L}$ one writes:

$$\begin{aligned}
\chi[\, v\,] - \chi[\, \dot{l}\,] &= -\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \left(\mathbf{L}[\dot{l}] + \dot{\mathbf{K}}^*\right)\; \Delta\dot{l}\, \mathrm{d}s \\
&\quad -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\Delta\dot{l}]\; \Delta\dot{l}\, \mathrm{d}s \\
&= -\int_{\mathcal{F}|_{\varphi=0}} \dot{\varphi}\; (v - \dot{l})\, \mathrm{d}s \\
&\quad -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\Delta\dot{l}]\; \Delta\dot{l}\, \mathrm{d}s \\
&= -\int_{\mathcal{F}|_{\varphi=0}} \dot{\varphi}\, v\; \mathrm{d}s + \int_{\mathcal{F}|_{\varphi=0}} \dot{\varphi}\, \dot{l}\; \mathrm{d}s \\
&\quad -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\Delta\dot{l}]\; \Delta\dot{l}\; \mathrm{d}s \geq 0
\end{aligned}$$

because of the stable crack growth hypothesis, consistency conditions $\dot{\varphi} \leq 0$ and $\dot{\varphi}\,\dot{l} = 0$ when $\varphi = 0$, and constraint $v \geq 0$.

Proposition 2 *Under hypothesis* (37)*, the crack tip "velocity" $\dot{l}(s,t)$ that solves the global quasi-static fracture propagation problem at "time" t minimizes the functional:*

$$\omega[v(s,t)] = -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\, v(s,t)\,]\, v(s,t)\, \mathrm{d}s \quad (39)$$

under the constraint: $\frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \left\{\dot{\mathbf{K}}^* + \mathbf{L}[\, v(s,t)\,]\right\} \leq 0$

To prove the theorem, denote again with $v(s,t) = \dot{l}(s,t) + \Delta\dot{l}(s,t)$. Omitting the variable dependence on s and t for the sake of clearness and in view of the symmetry property (36), one writes:

$$\omega[\, v\,] - \omega[\, \dot{l}\,] = -\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\, \dot{l}\,]\; \Delta\dot{l}\; \mathrm{d}s - \frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial \varphi}{\partial \mathbf{K}^*} \cdot \mathbf{L}[\, \Delta\dot{l}\,]\; \Delta\dot{l}\; \mathrm{d}s$$

By adding and subtracting the amount $\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*} \cdot \dot{\mathbf{K}}^* \, \dot{l} \, \mathrm{d}s$ the latter holds:

$$\begin{aligned}\omega[\,v\,]-\omega[\,\dot{l}\,] &= -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{L}[\,\Delta\dot{l}\,]\,\Delta\dot{l}\;\mathrm{d}s\\ &+\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\left(\dot{\mathbf{K}}^*+\mathbf{L}[\,\dot{l}\,]\right)\,\dot{l}\;\mathrm{d}s\\ &-\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\dot{\mathbf{K}}^*\,\dot{l}\;\mathrm{d}s\\ &-\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{L}[\,\dot{l}\,]\,v\;\mathrm{d}s\\ &= -\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{L}[\,\Delta\dot{l}\,]\,\Delta\dot{l}\;\mathrm{d}s\\ &+\int_{\mathcal{F}|_{\varphi=0}} \dot{\varphi}\,\dot{l}\;\mathrm{d}s\\ &-\int_{\mathcal{F}|_{\varphi=0}} \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\left(\dot{\mathbf{K}}^*+\mathbf{L}[\,v\,]\right)\,\dot{l}\;\mathrm{d}s\end{aligned}$$

in view of the symmetry property (36). Owing to the stable crack growth hypothesis, consistency conditions $\dot{\varphi}\le 0$ and $\dot{\varphi}\,\dot{l}=0$ when $\varphi=0$, and constraint $\frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\left\{\dot{\mathbf{K}}^*+\mathbf{L}[\,v(s,t)\,]\right\}\le 0$ the thesis follows.

6 A benchmark

As usual denote with t a variable that orders events and consider a penny shaped crack (see Fig. 5) with radius $R(t) > R_{imp} > 0$ embedded in a continuum body, subject to an internal pressure $p(t,r)$ that "opens the crack" merely acting on a concentric circle of radius R_{imp}:

$$p(t,r) = k(t)\,H(R_{imp}-r) \tag{40}$$

k being the *load factor* and $H(x)$ the Heaviside step function

$$H(x)=\begin{cases}1 & if\ x\ge 0\\ 0 & if\ x<0\end{cases}$$

In such mode I conditions, the crack evolution in space is smooth, without kinking. The SIFs vector has a single non vanishing component that amounts to Kassir and Sih (1975):

$$K_1(R(t)) = \frac{2}{\pi}\frac{1}{\sqrt{R(t)}}\int_0^{R(t)}\frac{r\;p(t,r)}{(R(t)^2-r^2)^{1/2}}dr$$

whence, in view of (40):

$$K_1(R(t)) = 2\,\frac{k(t)}{\pi}\,\frac{R(t)-\sqrt{R^2(t)-R_{imp}^2}}{\sqrt{R(t)}} \tag{41}$$

Closed form solution (41) can be exploited in order to benchmark the variational framework developed in the previous sections.

When $k(t)$ reaches a propagation threshold so that $K_1 = K_1^C$, the onset of crack propagation is reached. Further increase of external actions allows fracture propagation and the radius R becomes $R+dR$. As the crack growth is axis-symmetric the amount dR is independent on the abscissa s along the crack front and Eq. (20) reduces to:

$$\dot{\varphi} = \frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\left[\mathbf{K}^*\frac{\dot{k}(t)}{k(t)}+\mathbf{K}_0^{(1)}\,\dot{R}(t)\right]$$

The closed form for the scalar $\mathbf{K}_0^{(1)}$ can be derived from the series expansion of Eq. (41) in terms of R:

$$\begin{aligned}&K_1(R+dR) = K_1(R)\\ &-\frac{k}{\pi R\sqrt{R}}\,\frac{R^2+R_{imp}^2-R\sqrt{R^2-R_{imp}^2}}{\sqrt{R^2-R_{imp}^2}}dR+O(dR^2)\end{aligned} \tag{42}$$

According to definition (21), operator $\mathbf{L}[\dot{l}(s,t)] = \mathbf{K}_0^{(1)}\,\dot{R}(t)$ and stability condition (22) is satisfied. The crack growth is therefore stable, as expected again.

Functional (38) holds:

$$\begin{aligned}\chi[\dot{R}(s,t)] = &-\frac{1}{2}\int_{\mathcal{F}|_{\varphi=0}}\frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{K}_0^{(1)}\;\dot{R}^2(s,t)\;\mathrm{d}s\\ &-\int_{\mathcal{F}|_{\varphi=0}}\frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{K}^*\,\frac{\dot{k}(t)}{k(t)}\,\dot{R}(s,t)\;\mathrm{d}s\end{aligned}$$

under the unilateral constraint $\dot{R}(s,t)\ge 0$. Consider a positive parameter ε and a positive elongation $\dot{Q}(s,t)\ge 0$, so that the configuration $\dot{R}(s,t)+\varepsilon\,\dot{Q}(s,t)\ge 0$ is in the set of admissible configurations for functional (38). Optimality implies

$$\chi[\dot{R}(s,t)+\varepsilon\,\dot{Q}(s,t)]\ge\chi[\dot{R}(s,t)]$$

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,\chi[\dot{R}(s,t)+\varepsilon\,\dot{Q}(s,t)]\Big|_{\varepsilon=0}\ge 0$$

The crack front $\mathcal{F}|_{\varphi=0}$ can be split into two parts. In the former, say $\mathcal{F}_1$, $\dot{R}$ is strictly positive. The complementary, say $\mathcal{F}_2$, is the part of $\mathcal{F}|_{\varphi=0}$ with vanishing

velocity $\dot{R} = 0$. Along $\mathcal{F}_1$ the usual Euler–Lagrange equation $\chi'[\dot{R}(s,t)] = 0$ hold, whereas along $\mathcal{F}_2$ the inequality $\chi'[\dot{R}(s,t)] \geq 0$ has to be satisfied. Accordingly, at all $s \in \mathcal{F}|_{\varphi=0}$ the Karush–Kuhn–Tucker conditions hold:

$$\dot{R}(s,t) \geq 0 \quad \chi'[\dot{R}(s,t)] \geq 0 \quad \chi'[\dot{R}(s,t)]\dot{R}(s,t) = 0$$

In view of closed form (42), it becomes:

$$\chi[\dot{R}(t)] = -\frac{1-\nu^2}{E}\,\dot{R}(t)\,\left[q[R(t)]\,\dot{R}(t) + 2\,r[R(t)])\,\frac{\dot{k}(t)}{k(t)}\right] \tag{43}$$

with:

$$q[R(t)] = \int_0^{2\pi R(t)} K_1 \cdot K_0^1 \,\mathrm{d}s$$

$$= -\frac{4k^2\left(R(t) - \sqrt{R^2(t) - R_{imp}^2}\right)\left(R^2(t) + R_{imp}^2 - R(t)\sqrt{R^2(t) - R_{imp}^2}\right)}{\pi\sqrt{R^4(t) - R^2(t)R_{imp}^2}}$$

$$r[R(t)] = \int_0^{2\pi R(t)} K_1^2 \,\mathrm{d}s = \frac{8k^2}{\pi}\,\left(R(t) - \sqrt{R^2(t) - R_{imp}^2}\right)^2$$

Along $\mathcal{F}_1$ the minimizer of functional (43) must satisfy the Euler–Lagrange equation:

$$-2\,\frac{1-\nu^2}{E}\,\left[q[R(t)]\,\dot{R}(t) + r[R(t)]\,\frac{\dot{k}(t)}{k(t)}\right] = 0$$

By "time" integration one gets:

$$\log\frac{k(t)}{k_0} = \frac{1}{2}\log\frac{R(t)}{R_0} + \log\frac{R(t) + \sqrt{R^2(t) - R_{imp}^2}}{R_0 + \sqrt{R_0^2 - R_{imp}^2}} \tag{44}$$

having set $R(0) = R_0, k(0) = k_0$. Equation (44) expresses the critical load factor corresponding to the evolution of radius $R(t)$ along $\mathcal{F}_1$. Due to the axial-symmetry of the problem, either $\mathcal{F}_1$ coincides with the whole circular crack front or is empty. In the latter case, $\dot{R}(t) = 0$ and the inequality

$$-2\,\frac{1-\nu^2}{E}\,\left[q[R(t)]\,\dot{R}(t) + r[R(t)]\,\frac{\dot{k}(t)}{k(t)}\right] \geq 0$$

is satisfied only by $\dot{k}(t) \leq 0$.

For example, setting $K_1(R_0) = K_1^C$ and $R_0 = R_{imp}$, from Eq. (42) one has:

$$k_0 = \frac{\pi\,K_1^C}{2\,\sqrt{R_0}}$$

from which the *load factor* as a function of the radius follows:

$$k(t) = K_1^C\,\pi\,\sqrt{R(t)}\,\frac{R(t) + \sqrt{R^2(t) - R_{imp}^2}}{2R_{imp}^2} \tag{45}$$

Equation (45) can be inverted in terms of $R(t)$:

$$R(t) = \frac{1}{48\alpha k^2 {K_1^C}^2}\left[\frac{\alpha^2}{\pi^{2/3}} + \pi^2\alpha\,{K_1^C}^4 + \pi^{2/3}\,{K_1^C}^4 \times\left(384k^4\,{R_{imp}}^2 + \pi^4\,{K_1^C}^4\right)\right] \tag{46}$$

with

$$\alpha = \sqrt{3}\,{K_1^C}^4\left(576\pi^4k^4\,{K_1^C}^4\,{R_{imp}}^2 + 1536k^6\,{R_{imp}}^3 \times\left(\sqrt{3}\sqrt{432k^4\,{R_{imp}}^2 + \pi^4\,{K_1^C}^4} + 36k^2R_{imp}\right) + \pi^8\,{K_1^C}^8\right)$$

Curves (45) and (46) are pot in Fig. 6 in the case $R_{\mathrm{imp}} = 1/2$, $K_1^C = 1$.

7 Conclusions

The present note aims at extending to three dimensional problems the variational formulation for the global incremental quasi-static linear elastic fracture propagation problem presented in Salvadori and Carini (2011). Laws that describe onset and propagation of cracks fall under the general Griffith theory (Griffith 1921), that puts linear elastic fracture mechanics (LEFM) in analogy with standard dissipative systems thermodynamics (Lemaitre and Chaboche 2000). The SIFs vector "right after the kink, if any" $\mathbf{K}^*$ acts as an internal force, conjugated to the internal variable $\mathbf{a}^*$. The latter is related to the crack length from definition (14) and its variation induces a dissipation which must satisfy Clausius–Duhem's inequality. Following this path of reasoning, a plasticity analogy for LEFM was presented in Salvadori (2008): as expected, it stems from a maximum dissipation principle, the counterpart of the maximum internal entropy generation (Tadmor et al. 2011) postulate for standard dissipative systems thermodynamics. "Constitutively" the crack elongation is analogous to a rigid-plastic behaviour. Griffith's criterion is recovered following a rigorous setting.

The key ingredients for this approach are: i) the stress intensity factor expansion (1) with respect to the crack elongation, provided in Leblond (1999), Leblond et al. (1999). In the analogy it plays the role of a Colonnetti's decomposition of stress; ii) the 3D extension of

Irwin's formula, that relates the Energy Release Rate to the SIFs; whereas a formal proof for mixed mode has been recently provided for 2D in Chambolle et al. (2010), a formal 3D derivation seems to be lacking. In this regard, formula (8) is taken as a generally accepted result. In the analogy, it is equivalent to the yield function in plasticity and allows the definition of the elastic domain and of its boundary; iii) the maximum dissipation principle, whence the normality and the complementarity laws come out. The rest follows from well known results of incremental plasticity (Ceradini (1965, 1966)).

Whereas in the 2D case the discrete nature of the number of crack tips made the formulation simple and the variational formulation a minimum of functions, in the 3D case the presence of a curved crack front requires a detailed investigation of the symmetry property of operator $N(\dot{l})$ defined in (23). As shown in Sect. 5.1 the form (21) of non local operator $\mathbf{L}[\,\dot{l}\,]$ does not allow a straightforward proof of symmetry for N. The latter was here provided on the basis of the physical meaning of operator N, because of its link with the concept of energy release rate.

Once symmetry is provided, the two minimum theorems follows. They allow the formulation of effective algorithms to determine the size of the increment of crack growth, based upon the corresponding increment for external loads. Such an outcome for incremental fracture has the potential to place numerically calculated crack trajectories on a much firmer basis. In particular, it could be implemented in techniques such as standard finite element or boundary element (Salvadori and Gray 2007) method in which crack growth direction is usually determined based upon standard criteria (Salvadori 2010), but the crack growth rate is by far less well defined.

Conditions for stable crack growth and, inherently, for the onset of unstable propagation have been firmly formulated is Sect. 4. If these theoretical results will be confirmed, they may reveal of great importance. The safety of a structural component is usually measured against the stable/unstable crack growth transition, as one assumes in fact that unstable propagation leads to structural collapse. Numerical validations are in progress.

The plasticity analogy in which the present paper is rooted allows the interpretation of "fracture as a standard dissipative system". The details of such a theory have been reviewed in Nguyen (1994, 2000), Mielke (2005), firmly formulating as variational inequalities the stability and bifurcation conditions. A current yet long lasting research target of us aims at relating stability and variational formulation presented in Sects. 4 and 5.2 to the general theory and in particular with the notion of "non negativeness of second order work density" attributed to Hill (1958). It did not reach a sufficient maturity to be included here as a discussion and will be considered in future publications.

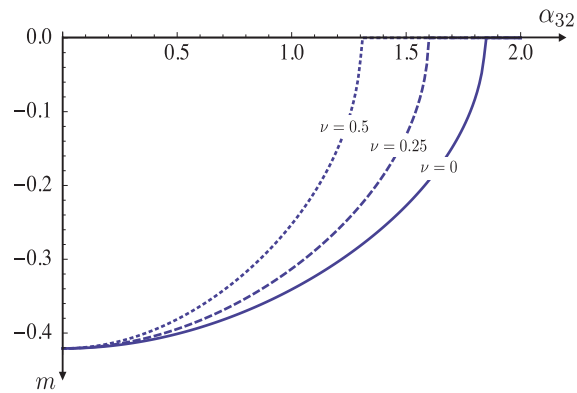

Fig. 7 Plot of the kink angle $\theta_{MERR} = m\,\pi$ as a function of the ratio $\alpha_{32} = K_3/K_2$ at $K_1 = 0$

Authors are aware that the proposed approach narrows the picture very much with respect to other formulations (Bourdin et al. 2008): going out of LEFM, where SIFs are not defined, does not seem to be obvious. Nevertheless, even within the proposed framework some extension can be clearly devised. The issues of thermally driven fracture processes and of environmentally assisted brittle fracture might be included naturally, once the correct driving force is described as done for the increment of external loads. More ambitious extensions pertain to fatigue of materials, which involves the multiscale analysis of the growth of defects, still embedded in a macroscopically linear elastic material behavior.

Acknowledgments Authors are gratefully indebted with J. B. Leblond for several, long, deep discussions and suggestions. A large part of the proof in Sect. 5.1 was made by him. Fruitful and, as always, exhaustive discussions with A. Giacomini are gratefully acknowledged.

Appendix A: Crack propagation requirements

Stable crack propagation is meant to be a sequence of equilibrium states. At each load $k(\tau)$ corresponds

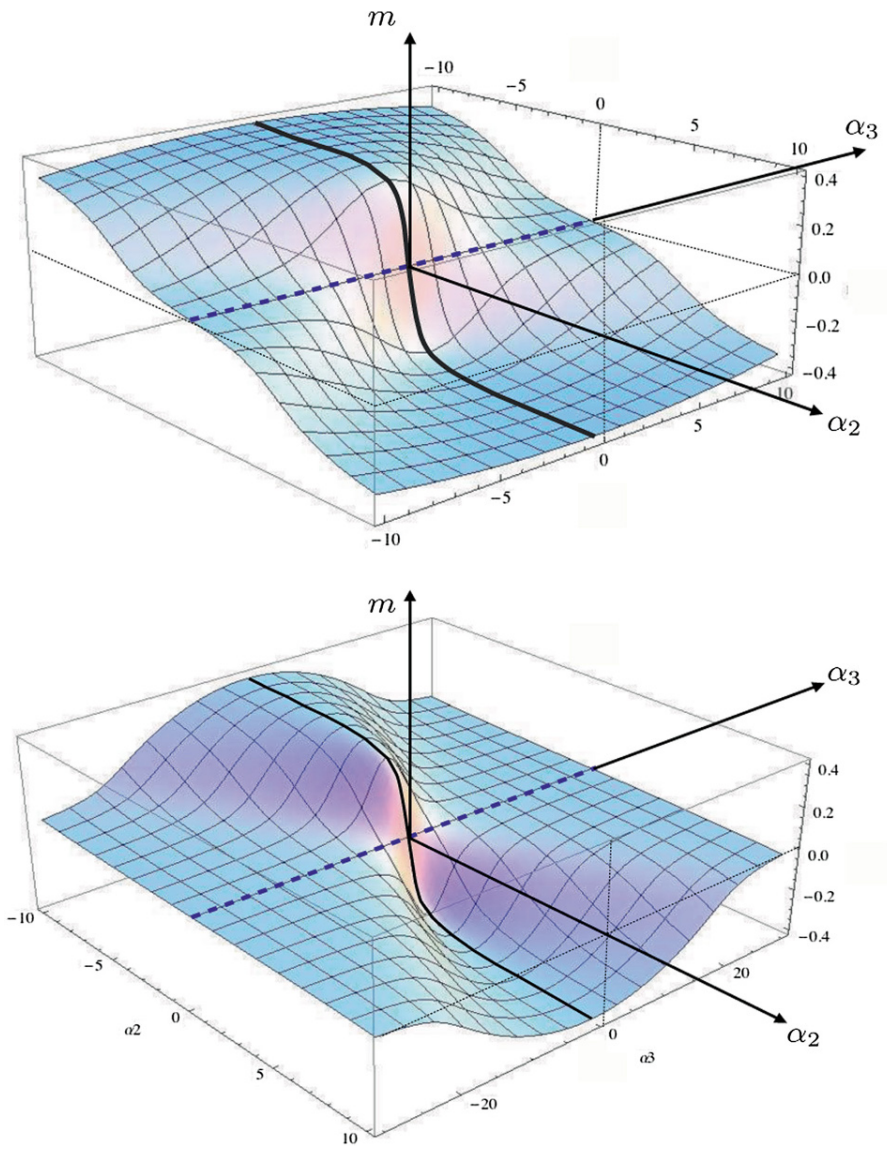

Fig. 8 Plot of the kink angle $\theta_{MERR} = m\,\pi$ as a function of the ratios $\alpha_2 = K_2/K_1$ and $\alpha_3 = K_3/K_1$. At $K_2 = 0$ (mode 1+3) the kinking angle is vanishing. At $K_3 = 0$ the plane case is recovered as well as the well known curve θ versus α: it is here highlighted with a *thick curve*. As well documented, θ_{MERR} and θ_{LS} differ from very small amounts in the plane case. Accordingly, a visualization of θ_{LS} cannot be distinguished from θ_{MERR} and thus the thick curve recovers θ_{LS} as well. Noteworthy, θ_{LS} is independent upon α_3. It is then easy to envisage the surface plot of angle θ_{LS} as the cylindrical envelope of the thick curve along axis α_3. The higher the mode 3 contribution α_3 the higher the difference in the angle of propagation between LS and MERR. This evidence is clear in **b**, whose axis are not in scale

a crack configuration $l(s,\tau)$ which eventually evolves quasi-statically, keeping the SIFs at the onset of propagation $\mathbf{K}^*(\tau) \in \partial\mathbb{E}$. Assume "time" τ in an interval $]t, t^*]$ in which crack grows steadily with "velocity" $0 < \dot{l}(s,\tau) < \infty$. One has in view of (8):

$$\varphi(\tau)-\varphi(t) = A\left(K_1^{*2}(\tau)-K_1^{*2}(t)+K_2^{*2}(\tau)-K_2^{*2}(t)\right) + B\left(K_3^{*2}(\tau)-K_3^{*2}(t)\right) = 0 \qquad (47)$$

with $A = \frac{1-\nu^2}{E}$ and $B = \frac{1+\nu}{E}$. If the crack path in the normal plane at any $s \in \mathcal{F}$ is taken to be smooth in the "time" interval $]t, t^*]$, in other words along a curve at least of class C^1, then the kinking angle $\theta(s,\tau) = 0$ and $\mathbf{K}^*(\tau) = \mathbf{K}(\tau)$ for being $\mathbf{F} = \mathbf{1}$ in expansion (2). If furthermore one selects the local symmetry (Goldstein and Salganik 1974) as a kinking angle criterion, then $K_2^* = 0$ and therefore:

$$\varphi(\tau)-\varphi(t) = A[K_1^*(\tau)+K_1^*(t)][K_1^*(\tau)-K_1^*(t)] + B[K_3^*(\tau)+K_3^*(t)][K_3^*(\tau)-K_3^*(t)] = 0 \qquad (48)$$

Using SIFs expansion (18) one has:

$$[K_1^*(\tau)-K_1^*(t)] = \mathbf{e_1}\cdot\left\{\mathbf{K}^*(t)\frac{k(\tau)}{k(t)} + \mathbf{K}^{(1/2)}\frac{k(\tau)}{k(t)}\sqrt{l(s,\tau)} + \mathbf{K}_0^{(1)}\frac{k(\tau)}{k(t)}l(s,\tau)\right.$$

$$
\begin{aligned}
&+\mathbf{K}_1^{(1)}\frac{k(\tau)}{k(t)}\frac{\partial l}{\partial s}(s,\tau)+\mathbf{K}_{nl}^{(1)}\frac{k(\tau)}{k(t)}[(l(s',\tau)-l(s,\tau))]\\
&-\mathbf{K}^*(t)\Big\}+o(l)\\
=&\mathbf{e_1}\cdot\Big\{\mathbf{K}^*(t)\frac{\delta k}{k(t)}+\mathbf{K}^{(1/2)}\sqrt{l(s,\tau)}+\mathbf{K}_0^{(1)}l(s,\tau)+\\
&+\mathbf{K}_1^{(1)}\frac{\partial l}{\partial s}(s,\tau)+\mathbf{K}_{nl}^{(1)}[l(s',\tau)-l(s,\tau)]\Big\}+o(\delta k\cdot l)
\end{aligned}
\tag{49}
$$

where $\mathbf{e}_1$ denotes the unit vector in direction 1. Analogously:

$$
\begin{aligned}
&[K_3^*(\tau)-K_3^*(t)]\\
&=\mathbf{e_3}\cdot\Big\{\mathbf{K}^*(t)\frac{\delta k}{k(t)}+\mathbf{K}^{(1/2)}\sqrt{l(s,\tau)}+\mathbf{K}_0^{(1)}l(s,\tau)\\
&+\mathbf{K}_1^{(1)}\frac{\partial l}{\partial s}(s,\tau)+\mathbf{K}_{nl}^{(1)}[\,l(s',\tau)-l(s,\tau)\,]\Big\}+o(\delta k\cdot l)
\end{aligned}
\tag{50}
$$

where $\mathbf{e}_3$ denotes the unit vector in direction 3.

By noting that $l(s,0)=0$ by definition of $l(s,t)$ and assuming that it exists a bounded quasi-static velocity $\dot{l}(s,t)$ so that $l(s,\tau)=\dot{l}(s,t)(\tau-t)$ and a load variation velocity so that $\delta k=\dot{k}(t)(\tau-t)$ one has for $\tau\to t^+$:

$$
\begin{aligned}
0=&2(A\,K_1^*(t)\,\mathbf{e}_1+B\,K_3^*(t)\,\mathbf{e}_3)\\
&\cdot\mathbf{K}^{(1/2)}\sqrt{\dot{l}(s,t)}\,\sqrt{\tau-t}\\
&+2(A\,K_1^*(t)\,\mathbf{e}_1+B\,K_3^*(t)\,\mathbf{e}_3)\\
&\cdot\Big\{\mathbf{K}^*\frac{\dot{k}(t)}{k(t)}+\mathbf{K}_0^{(1)}\dot{l}(s,t)+\mathbf{K}_1^{(1)}\dot{l}'(s,t)\\
&+\mathbf{K}_{nl}^{(1)}[\,\dot{l}(s',t)-\dot{l}(s,t)\,]\Big\}\;(\tau-t)+o(\tau-t)
\end{aligned}
\tag{51}
$$

whence the conditions:

$$
\frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot\mathbf{K}^{(1/2)}=0 \tag{52}
$$

$$
\begin{aligned}
&\frac{\partial\varphi}{\partial\mathbf{K}^*}\cdot[\mathbf{K}^*\frac{\dot{k}(t)}{k(t)}+\mathbf{K}_0^{(1)}\dot{l}(s,t)+\mathbf{K}_1^{(1)}\dot{l}'(s,t)\\
&+\mathbf{K}_{nl}^{(1)}[(\dot{l}(s',t)-\dot{l}(s,t))]=0
\end{aligned}
\tag{53}
$$

at $\mathbf{K}_2^*=0$.

Appendix B: MERR and LS

Maximum energy release rate (MERR) *onset of propagation* uses as a magnitude ϑ the energy released during crack advance at any point along the crack front. Such a magnitude is related to stress intensity factors after a kink via Irwin's formula, recently revised at a kink by several authors (Ichikawa and Tanaka (1982); Chambolle et al. (2010)) in two-dimensions, whence the onset of propagation reads

$$
\varphi=\frac{1-\nu^2}{E}(K_1^{*2}+K_2^{*2})+\frac{1+\nu}{E}K_3^{*2}-G_C \tag{54}
$$

It seems extremely desirable, although probably quite involved, an extension of Chambolle et al. (2010) to the three dimensional case. As its formal derivation in the presence of kinking seems not to be available, the validity of Irwin's formula for 3D, widely accepted in the fracture mechanics community, is here assumed.

As stated already several times in this note, the principle of Local Symmetry (LS) and MERR share the same onset of propagation. They differ on the *criteria for kinking angle prediction*. The kink angle predicted by the MERR descends from the general form (7). It reads:

$$
\begin{aligned}
\frac{\partial\varphi}{\partial\theta}=&\frac{1-\nu^2}{E}\left(2K_1^*\frac{\partial K_1^*}{\partial\theta}+2K_2^*\frac{\partial K_2^*}{\partial\theta}\right)\\
&+\frac{1+\nu}{E}2K_3^*\frac{\partial K_3^*}{\partial\theta}=0
\end{aligned}
\tag{55}
$$

Matrix $\mathbb{F}$ has been defined in terms of the ratio $m=\theta/\pi$ in Leblond (1999), Leblond et al. (1999) as

$$
\begin{aligned}
F_{11}(m)=&\,4.1m^{20}+1.63m^{18}-4.059m^{16}+2.996m^{14}\\
&-0.0925m^{12}-2.88312m^{10}+5.0779m^8\\
&+\left(\frac{\pi^2}{9}-\frac{11\pi^4}{72}+\frac{119\pi^6}{15360}\right)m^6\\
&+\left(\pi^2-\frac{5\pi^4}{128}\right)m^4-\frac{3\pi^2m^2}{8}+1\\
F_{12}(m)=&\,4.56m^{19}+4.21m^{17}-6.915m^{15}+4.0216m^{13}\\
&+1.5793m^{11}-7.32433m^9+12.313906m^7\\
&+\left(-2\pi-\frac{133\pi^3}{180}+\frac{59\pi^5}{1280}\right)m^5\\
&+\left(\frac{10\pi}{3}+\frac{\pi^3}{16}\right)m^3-\frac{3\pi m}{2}\\
F_{21}(m)=&-1.32m^{19}-3.95m^{17}+4.684m^{15}-2.07m^{13}\\
&-1.534m^{11}+4.44112m^9-6.176023m^7\\
&+\left(-\frac{2\pi}{3}+\frac{13\pi^3}{30}-\frac{59\pi^5}{3840}\right)m^5\\
&-\left(\frac{4\pi}{3}+\frac{\pi^3}{48}\right)m^3+\frac{\pi m}{2}\\
F_{22}(m)=&\,12.5m^{20}+0.25m^{18}-7.591m^{16}+7.28m^{14}\\
&-1.8804m^{12}-4.78511m^{10}+10.58254m^8\\
&+\left(-\frac{32}{15}-\frac{4\pi^2}{9}-\frac{1159\pi^4}{7200}+\frac{119\pi^6}{15360}\right)m^6
\end{aligned}
$$

$$+\left(\frac{8}{3}+\frac{29\pi^2}{18}-\frac{5\pi^4}{128}\right)m^4-\left(4+\frac{3\pi^2}{8}\right)m^2+1$$

$$F_{33}(m)=\left(\frac{1-m}{1+m}\right)^{m/2}$$

$$F_{13}(m)=F_{31}(m)=F_{32}(m)=F_{23}(m)=0$$

It is straightforward to show that at $K_1 \neq 0$ Eq. (55) is equivalent to

$$\left[(F_{11}+\alpha_2 F_{12})\left(\frac{\partial F_{11}}{\partial\theta}+\alpha_2\frac{\partial F_{12}}{\partial\theta}\right)\right.$$
$$\left.+(F_{21}+\alpha_2 F_{22})\left(\frac{\partial F_{21}}{\partial\theta}+\alpha_2\frac{\partial F_{22}}{\partial\theta}\right)\right]$$
$$+\frac{1}{1-\nu}\alpha_3^2\frac{\partial F_{33}}{\partial\theta}=0 \tag{56}$$

where $\alpha_2 = K_2/K_1$ and $\alpha_3 = K_3/K_1$. For a given material (i.e. a given Poisson ratio) at any couple α_2, α_3 the corresponding kink angle θ_{MERR} solves Eq. (56). At $K_1 = 0$ angle θ_{MERR} is plot as a function of ratio $\alpha_{32} = K_3/K_2$ and of Poisson ratio in Fig. 7.

The local symmetry criterion is the only notable exception to the mathematical representation (7). It provides the kink angle θ_{LS} through the equation $K_2^* = 0$:

$$F_{21}+\alpha_2 F_{22}=0 \tag{57}$$

where $\alpha_2 = K_2/K_1$. For any α_2 Eq. (57) provides the kink angle θ_{LS} which turns out to be independent on the mode 3 stress intensity factor.

Whereas thus in 2D the two angles θ_{MERR} and θ_{LS} differ from very small amounts, in 3D the scenario changes completely as it can be readily seen in Fig. 8. This fact may allow experimental campaigns of investigation to provide conclusive statements on which criteria better describes crack kinking in brittle materials.

References

Amestoy M, Leblond JB (1992) Crack paths in plane situations—ii. Detailed form of the expansion of the stress intensity factors. Int J Solids Struct 29:465–501

Barenblatt GI (1959) On equilibrium cracks forming during brittle fracture (in Russian). Prikladnaya Matematika i Mekhanika 23:434–444; [see also, The mathematical theory of equilibrium cracks in brittle fracture. Adv Appl Mech 7:55–129 (1962)]

Bourdin B, Francfort G, Marigo JJ (2008) The variational approach to fracture. Springer, Berlin

Ceradini G (1965) Un principio di massimo per il calcolo dei sistemi elasto-plastici. Rendiconti Istituto Lombardo di Scienze e Lettere A99

Ceradini G (1966) A maximum principle for the analysis of elastic-plastic systems. Meccanica 1:77–82

Chambolle A, Francfort GA, Marigo JJ (2009) When and how do cracks propagate? J Mech Phys Solids 57(9):1614–1622

Chambolle A, Francfort GA, Marigo JJ (2010) Revisiting energy release rates in brittle fracture. J Nonlinear Sci 20:395–424

Colonnetti G (1918) Sul problema delle coazione elastiche. Rend Accad Lincei 27: NotaI: 257–270, NotaII:331–335

Colonnetti G (1950) Elastic equilibrium in the presence of permanent set. Quart Appl Math 7:353–362

Francfort GA, Marigo JJ (1998) Revisiting brittle fracture as an energy minimization problem. J Mech Phys Solids 46:1319–1342

Goldstein RV, Salganik RL (1974) Brittle fracture of solids with arbitrary cracks. Int J Fract 10:507–523

Griffith AA (1921) The phenomena of rupture and flow in solids. Phil Trans R Soc 221:163–198

Han W, Reddy BD (1999) Plasticity. Springer, New York

Hill R (1958) A general theory of uniqueness and stability in elastic-plastic solids. J Mech Phys Solids 6:236–249

Ichikawa M, Tanaka S (1982) A critical analysis of the relationship between the energy release rate and the SIFs for non-coplanar crack extension under combined mode loading. Int J Fract 18:19–28

Irwin G (1958) Fracture. In: Fluegge S (ed) Handbuch der Physik, Bd. 6. Elastizitaet und Plastizitaet. Springer, Berlin, pp 551–590

Kassir MK, Sih GC (1975) Mechanics of fracture, vol 2. Three-dimensional crack problems. Noordhoff Int. Publ, Leyden

Lazarus V (2011) Perturbation approaches of a planar crack in linear elastic fracture mechanics. J Mech Phys Solids 59:121–144

Leblond JB (1999) Crack paths in three dimensional elastic solids—i. Two term expansion of the stress intensity factors—application to crack path stability in hydraulic fracturing. Int J Solids Struct 36:79–103

Leblond JB, Lazarus V, Mouchrif S (1999) Crack paths in three dimensional elastic solids—ii. Three term expansion of the stress intensity factors—applications and perspectives. Int J Solids Struct 36:105–142

Lemaitre J, Chaboche JL (2000) Mechanics of solid materials. Cambridge University Press, Cambridge

Mielke A (2005) Evolution in rate-independent systems. In: Dafermos C, Feireisl E (eds) Handbook of differential equations, evolutionary equations, vol 2. Elsevier, Amsterdam, pp 461–559

Nguyen QS (1994) Bifurcation and stability in dissipative media (plasticity, friction, fracture). Appl Mech Rev 47(1):1–30

Nguyen QS (2000) Stability and nonlinear solid mechanics. Wiley, New York

Rice J (1989) Weight function theory for three-dimensional elastic crack analysis. In: Wei RP, Gangloff RP (eds) Fracture mechanics: perspectives and directions (20th symposium). ASTM STP 1020, American Society for Testing and Materials, Philadelphia, pp 29–57

Riks E (1979) An incremental approach to the solution of snapping and buckling problems. Int J Solid Struct 15:529–551

Salvadori A (2008) A plasticity framework for (linear elastic) fracture mechanics. J Mech Phys Solids 56:2092–2116

Salvadori A (2010) Crack kinking in brittle materials. J Mech Phys Solids 58:1835–1846

Salvadori A, Carini A (2011) Minimum theorems in incremental linear elastic fracture mechanics. Int J Solids Struct 48:1362–1369

Salvadori A, Giacomini A (2012) The most dangerous flaw orientation in brittle materials and structures. J Mech Phys Solids (submitted for publication)

Salvadori A, Gray LJ (2007) Analytical integrations and SIFs computation in 2D fracture mechanics. Int J Numer Methods Eng 70:445–495

Tadmor EB, Miller RE, Elliott RS (2011) Continuum mechanics and thermodynamics: from fundamental concepts to governing equations. Cambridge University Press, Cambridge

Tonti E (1984) Variational formulation for every nonlinear problem. Int J Eng Sci 22:1343–1371

DOI 10.1007/978-3-319-04397-5_7
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9824-5

ORIGINAL PAPER

Crack patterns obtained by unidirectional drying of a colloidal suspension in a capillary tube: experiments and numerical simulations using a two-dimensional variational approach

C. Maurini · B. Bourdin · G. Gauthier · V. Lazarus

Received: 7 August 2012 / Accepted: 16 February 2013 / Published online: 5 March 2013

Abstract Basalt columns, septarias, and mud cracks possess beautiful and intriguing crack patterns that are hard to predict because of the presence of cracks intersections and branches. The variational approach to brittle fracture provides a mathematically sound model based on minimization of the sum of bulk and fracture energies. It does not require any *a priori* assumption on fracture patterns and can therefore deal naturally with complex geometries. Here, we consider shrinkage cracks obtained during unidirectional drying of a colloidal suspension confined in a capillary tube. We focus on a portion of the tube where the cross-sectional shape cracks does not change as they propagate. We apply the variational approach to fracture to a tube cross-section and look for two-dimensional crack configurations minimizing the energy for a given loading level. We achieve qualitative and quantitative agreement between experiments and numerical simulations using a regularized energy (without any assumption on the cracks shape) or solutions obtained with traditional techniques (fixing the overall crack shape a priori). The results prove the efficiency of the variational approach when dealing with crack intersections and its ability to predict complex crack morphologies without any a priori assumption on their shape.

Keywords Brittle fracture mechanics · Griffith's fracture energy · Variational approach to fracture · Free-discontinuity problems · Drying of a colloidal suspension · Shrinkage cracks

C. Maurini
UPMC Univ Paris 6, UMR 7190, Institut Jean Le Rond d'Alembert, Boite courrier 161-2, 4 Place Jussieu, 75005 Paris, France
e-mail: corrado.maurini@upmc.fr

C. Maurini
CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, Boite courrier 161-2, 4 Place Jussieu, 75005 Paris, France

B. Bourdin
Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
e-mail: bourdin@lsu.edu

G. Gauthier · V. Lazarus (✉)
UPMC Univ Paris 6, UMR 7608, FAST, Bat 502, Campus Univ, 91405 Orsay, France
e-mail: veronique.lazarus@upmc.fr

G. Gauthier · V. Lazarus
Univ Paris-Sud, UMR 7608, FAST, Bat 502, Campus Univ, 91405 Orsay, France
e-mail: gauthier@fast.u-psud.fr

G. Gauthier · V. Lazarus
CNRS, UMR 7608, FAST, Bat 502, Campus Univ, 91405 Orsay, France

1 Introduction

Complex fracture networks involving crack interactions and intersections are observed in a wide variety of situations associated with shrinkage loadings. Giant's Causeway (DeGraff and Aydin 1987), Port Arthur

tessellated pavement (Branagan and Cairns 1993), Bimini Road (Shinn 2009), Mars polygons (Mangold 2005), septarias (Seilacher 2001), fracture networks in permafrost (Plug and Werner 2001), and mud-cracks in coatings or paintings (Colina and Roux 2000) are a few examples that have intrigued people throughout history. Significant efforts have been devoted to reproducing similar phenomena in controlled experiments using either shrinkage induced by cooling (Yuse and Sano 1993; Ronsin et al. 1995; Yang and Ravi-Chandar 2001; Muller 1998), or drying (Hofmann et al. 2006; Goehring et al. 2006; Toramaru and Matsumoto 2004; Gauthier et al. 2010; Lazarus and Pauchard 2011). A large body of theoretical and numerical work has also be devoted to the propagation of preexisting cracks using classical tools of linear fracture mechanics (see for instance Bažant et al. 1979; Nemat-Nasser et al. 1980; Adda-Bedia and Pomeau 1995; Hofmann et al. 2006; Jenkins 2009; Bahr et al. 2010) or phase-field approaches (Corson et al. 2009). Despite all this work, many features of the complex crack networks observed in experiments are still poorly understood. In particular, a common limitation of all the references above is their inability to deal with cracks branching or intersecting (see Saliba and Jagla (2003) or Bahr et al. (2009) for some attempts at dealing with these issues.)

In contrast, the variational approach to fracture mechanics proposed by Francfort and Marigo (1998) provides a solution to these issues by treating the crack shape as a genuine unknown. In the context of brittle fracture, it can be seen as a natural extension of Griffith's energetic formulation Griffith (1920). It is based on the minimization amongst all admissible crack sets and possibly discontinuous displacement fields of a total energy functional consisting of the sum of a bulk (elastic) and a surface term proportional to the surface of the cracks (or their length in two dimensions). The minimization problem is challenging because it is in general technically not possible to test all crack configurations. Instead, a regularized energy is used to the numerical prediction of complex crack patterns without any preliminary assumption on the overall cracks shape. In recent years, efforts have mainly focused on its theoretical and numerical developments (Bourdin et al. 2000, 2008) . However, applications to the prediction of complex crack patterns with a close qualitative but also quantitative comparison to experimental results are still lacking, in particular when crack intersections are involved. This paper constitute a first effort aiming at filling this gap. We apply the variational approach to the controlled drying of a colloidal suspension in a capillary tube for which fascinating experimental results have been obtained when the colloidal suspension gradually turns into a drained porous solid matrix.

During drying. the natural shrinkage of the solid matrix which is prevented by the strong adhesion to the wall of the tube, gives rise to high tensile stresses and to a large number of disordered cracks which gradually self-organize and propagate with a constant cross-sectional geometry. The cross-sectional shape depends on the geometry of the tube and the drying conditions (e.g. Allain and Limat 1995; Dufresne et al. 2003, 2006; Gauthier et al. 2007) and looks mostly like stars where several straight cracks intersect at or near a single point. In this paper, we focus on the two-dimensional problem of the prediction of these cross-sectional crack shapes, but we refrain to solve the entire, far more complex, three-dimensional propagation problem.

Recently, Gauthier et al. (2010) modeled the drying porous solid by an elastic material loaded by a tensile prestress and showed that the observed cross-sectional crack patterns can be correctly predicted by energy minimization amongst a given family of cracks, namely arrays of parallel cracks for flat tubes or star-shaped cracks for cylindrical ones. This demonstrated that their simple model captures the key physical ingredients. However, in this first study the overall crack shape was fixed a priori. Here, we extent the experiments of Gauthier et al. (2010) to non axisymmetric square geometry, for which the crack shape is more difficult to guess, and revisit them through the variational method developed by Bourdin et al. (2000) using the physical model demonstrated previously. We show that this method provides a qualitative and quantitative description of the different cross-sectional crack patterns observed experimentally without any preliminary shape assumptions.

The outline of the article is the following. The experimental setup is described in details in Section 2, illustrating the results obtained by changing the suspension and the drying velocities. In addition to the experiments on circular tubes previous reported in Gauthier et al. (2010), new experiments on capillary tubes with square cross-section highlight the relation between the tube cross-sectional shape and the crack pattern. In Sect. 3, we focus on the regime where the cross-sectional geometry of the crack does not evolve. We describe our

two-dimensional model assuming a linear elastic brittle solid under plane-strain conditions and a drying loading introduced through a tensile isotropic inelastic strain. In Sect. 4, we recall the results of Gauthier et al. (2010) and search for least-energy solutions amongst the class of star-shaped cracks. This allows us to identify some semi-analytical solutions to be used as reference for the verification of the numerical implementation. We then seek to fully identify the crack geometry by numerical minimization of a regularized form of the total energy functional (Sect. 5). There, we briefly describe the details of the approach then introduce a selection principle. We illustrate our approach by providing a qualitative and quantitative comparison with experimental and semi-analytical results in Sect. 7.

2 Experiments

Experiments are conducted in circular or square glass capillary tubes. The height of the tubes is typically 10 cm and the radius of the cross-section of the circular tubes are $R = 0.05$ mm, $R = 0.15$ mm, $R = 0.5$ mm or $R = 0.75$ mm while the edge length of the square tubes is $2R = 1$ mm. According to the manufacturer, the precision on the radii is 10 %. The tubes are positioned in vertical orientation and colloidal suspension is sucked up from their bottom until they are 3/4 full (see Fig. 1a). The top of the tubes is closed and maintained at a pressure slightly below that of the surrounding. Because of the capillary effect, this prevents the fluid from flowing under gravity. Two type of colloids have been used, both made of water silica dispersions: Ludox™HS40 and SM30 designated as HS40 and SM30 in the following. HS40 and SM30 differ in mass fraction (30 and 40 % mass fraction of silica particles for SM30 and HS40 respectively) and in particle size ($r \simeq 3.5$ nm and $r \simeq 6$ nm for SM30 and HS40 respectively). As soon as capillary tubes are filled, water contained in the suspension evaporates from the bottom of the tube. In the early stages of drying, particles aggregate at the wall close to the open edge until they fill a section of the tube. Once this porous layer is formed, the liquid interface is composed of a high number of meniscus with a typical radius of curvature of the order of the particle size. These meniscus generate high negative pressure that cause the water to flow. The water flow drives the silica particles to accumulate at the porous upper surface and a porous plug further develops. Drying is then governed by Darcy's law (Dufresne et al. 2003, 2006). Since the pressure gradient decreases as the plug extends, drying velocity decreases. However for plugs long enough (∼20 tube diameters) drying velocity is almost constant over few diameters (∼5). It is this region that is studied in the sequel. Once the plug height is about twenty times the diameter, the tubes are placed in controlled environment maintained at a constant relative humidity (RH) and temperature (T). Experiments are performed at three different conditions: (i) at a room temperature $T \simeq 20\,^{\circ}C$ and RH maintained below 10% by using desiccant, (ii) at $T \simeq 20\,^{\circ}C$ and RH maintained over 90 % by introducing water in the chamber, (iii) at $T \simeq 3\,^{\circ}C$ and RH $< 10\,\%$.

The high negative capillary pressure, imposed by the interface, in the draining fluid generates high tensile stresses in the gel (Dufresne et al. 2003) that cause it to crack. The cracks pattern can be visualized either by transparency, which allows to follow the crack propagation during the drying, or by cutting the tube at the desired height, to observe the cross-sectional crack shape. Disordered cracks are first observed in the bottom. As the plug extends, the cracks self-organize, become vertical and grow along the drying direction preserving their cross-sectional shape for a length corresponding to several tube diameters (Gauthier et al. (2007) for a more detail description of the evolution). Examples are given in Fig. 1b. For square tubes, two perpendicular diagonal cracks appear. For circular tubes, star-shaped cracks where the tube cross section is divided into n sectors (with a central angle $2\pi/n$) are observed. Once these cracks have appeared, the porous medium continues to dry and to undergo high tensile stresses leading to secondary cracks appearing later on along the cross section of the tube.

In the sequel, we focus only on (i) the tube region where the crack cross-sectional geometry is independent of the depth and on (ii) the time period after the disordered cracks have self-organized and before the appearance of secondary cracks. We observe that the cross-sectional crack shape for a given suspension and tube geometry depends on the drying conditions through the drying velocity only. Indeed drying at $T \simeq 3\,^{\circ}C/RH \leq 10\,\%$ or at $T \simeq 20\,^{\circ}C/RH \geq 90\,\%$ gives the same crack tip velocities and the same crack patterns. Thus we refer to experiments conducted at $T \simeq 3\,^{\circ}C/RH \leq 10\,\%$ or at $T \simeq 20\,^{\circ}C/RH \geq 90\,\%$ as slow velocity (SV) experiments, and those conducted at $T \simeq 20\,^{\circ}C/RH \leq 10\,\%$ as high velocity (HV)

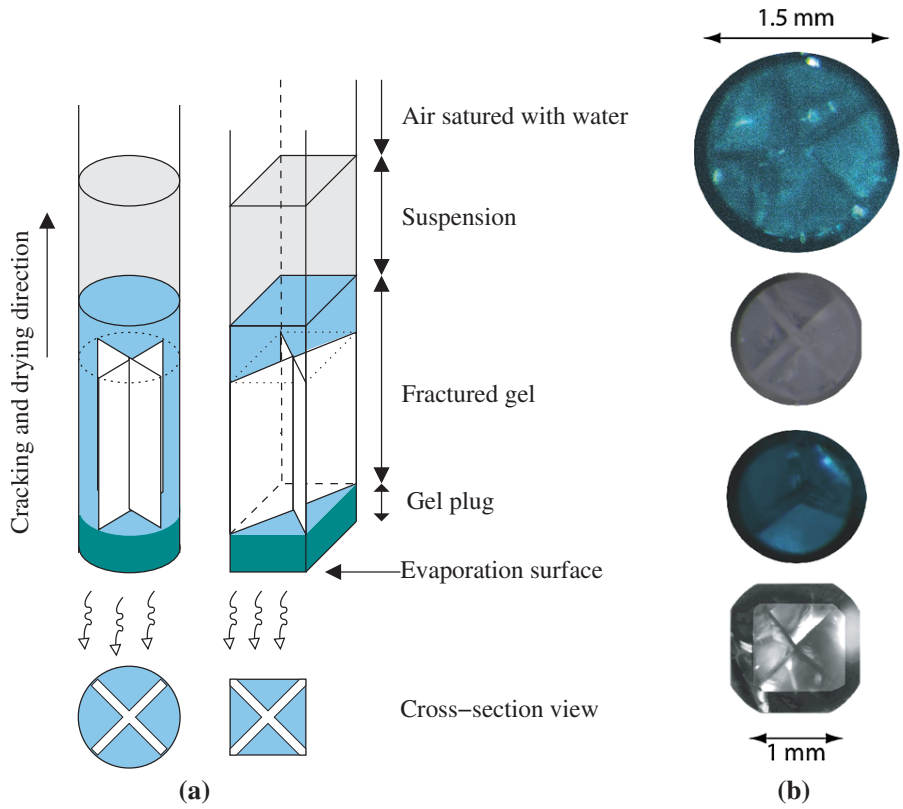

Fig. 1 A glass capillary tube oriented vertically is filled with a colloidal suspension; the *opened bottom* edge allows for evaporation of the water in an environment maintained at a constant relative humidity (RH) and temperature (T). The cross-sectional shape of the cracks depend on the tube shape and size and on the drying conditions. **a** Experimental setup and sketch of the self-organized star-shaped cracks. **b** Pictures of some cross section cuts (the colors depend on the light used)

experiments. A summary of experiments made can be found in Table 3 (Sect. 7). For small enough diameters, the number of cracks observed increases with the diameter. For larger containers, it is obvious that the crack pattern in the center becomes independent of the walls of the tube. We also observe that changing the drying conditions from SV to HV or HV to SV during the same experiment produces a transition between the two cross-sectional crack patterns obtained in pure HV and SV experiments (Gauthier et al. 2010) and that this rearrangement takes places over a distance approximately equal to the tube diameter. This suggests that history effects can be neglected.

3 Model

3.1 Basic hypotheses

In this work, we focus our attention on the steady-state regime where the cross-sectional geometry of the cracks remains constant. We model the problem as a two-dimensional plane-strain one on the cross-section of the tube. We assume that (i) the only effect of drying is to introduce an inelastic strain $\boldsymbol{\epsilon}_0 = \epsilon_0 \mathbf{1}$, where $\mathbf{1}$ denotes the identity matrix and $\epsilon_0 < 0$ is assumed constant throughout the cross section of the domain and that (ii) the solid adheres perfectly to the tube walls. Following the variational approach to brittle fracture (Francfort and Marigo 1998), for a given loading ϵ_0, we search for the deformation and crack configuration corresponding to a minimum of the total energy defined as the sum of the bulk elastic energy and the crack surface energy. The latter is of Griffith type, that is: the energy $\mathcal{S}(\gamma)$ per height unit, associated to a crack set γ is proportional to its length and given by

$$\mathcal{S}(\gamma) := G_c \mathcal{L}(\gamma),$$

where G_c is the specific fracture energy of the material, and $\mathcal{L}$ denotes the length of the crack. For the elastic energy, we suppose that the material is perfectly elastic prior to failure, and we adopt the linearized theory under the small displacement approximation. Finally, we assume that the crack surfaces are stress-free.

In the above formulation of the fracture mechanics problem, we simply look for the optimal cross-sectional crack pattern, according the energy minimality criterion, for given values of the loading ϵ_0. We do not introduce any concept of history dependence of the crack patterns. We solve a two-dimensional *static* problem and not the three-dimensional *quasi-static* evolution problem, in which one looks for the initiation and propagation of cracks as a function of time starting from a well-defined initial condition. Omitting history effects is suggested by the reversibility of the cross-sectional experimental patterns, using a two-dimensional approximation by the experimental observations of depth-independence of the cross-sectional crack patterns. The good agreement between numerical and experimental findings presented in Sect. 7 also justifies this hypothesis, *a posteriori*.

3.2 Variational fracture model

Let us introduce the following equivalent two-dimensional inelastic strain defined by $\boldsymbol{\epsilon}_0^{2d} = (1+\nu)\epsilon_0\,\mathbf{1}_2$, where $\mathbf{1}_2$ is the 2×2 identity matrix. With this notation and the aforementioned hypotheses, the strain energy can be written as:

$$w(\boldsymbol{\epsilon},\boldsymbol{\epsilon}_0) := \frac{E}{2(1+\nu)}\left(\frac{\nu}{(1-2\nu)}\mathrm{tr}^2(\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_0^{2d}) + (\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_0^{2d})\cdot(\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_0^{2d})\right) \quad (1)$$

where E and ν are the Young modulus and the Poisson ratio of the material, $\boldsymbol{\epsilon}$ is the symmetric second-order 2×2 matrix representing the linearized plane strain, tr denote the trace operator, and the dot is used for the scalar product. In linear elasticity, kinematic compatibility implies that $\boldsymbol{\epsilon}(u) = \frac{1}{2}(\nabla u + \nabla u^T)$, where u is the displacement field, ∇ the gradient operator, and the superscript T denotes the transpose operator. We parameterize the inelastic strain $\boldsymbol{\epsilon}_0^{2d}$ representing the drying loading by a non dimensional drying intensity $\bar{\epsilon}$, defined by (see the first remark at the end of this section)

$$\boldsymbol{\epsilon}_0^{2d} := -\bar{\epsilon}\sqrt{\frac{G_c}{ER}}\,\mathbf{1}_2 \quad (2)$$

where R is a characteristic length associated with the cross-section, typically its radius. Hence, the potential energy $\mathcal{P}_{\bar{\epsilon}}$ of the cross-section occupying the open set $\mathcal{C}$ and associated to a displacement field u and a crack set γ for a loading parameter $\bar{\epsilon}$ is given by

$$\mathcal{P}_{\bar{\epsilon}}(u,\gamma) := \int_{\mathcal{C}/\gamma} w_{\bar{\epsilon}}(\boldsymbol{\epsilon}(u))\mathrm{d}S, \quad \text{with}$$
$$w_{\bar{\epsilon}}(\boldsymbol{\epsilon}) := w\left(\boldsymbol{\epsilon}, -\bar{\epsilon}\sqrt{G_c/ER}\,\mathbf{1}_2\right). \quad (3)$$

The total energy is defined as the sum of the potential energy and the surface energy required to create the cracks:

$$\mathcal{E}_{\bar{\epsilon}}(u,\gamma) := \mathcal{P}_{\bar{\epsilon}}(u,\gamma) + \mathcal{S}(\gamma). \quad (4)$$

For any given loading parameter $\bar{\epsilon}$, we seek to find the crack set γ and displacement field u as the global minimizer of (4) amongst *any* admissible crack set and kinematically admissible displacement fields. The admissible crack sets consist of all possible curves or sets of curves inside the boundary of $\mathcal{C}$. For any given crack set γ, the space of the admissible displacements is

$$\mathcal{U}(\gamma) := \{u \in H^1(\mathcal{C}\setminus\gamma;\mathbb{R}^2), \quad u = 0 \text{ on } \partial\mathcal{C}\}, \quad (5)$$

i.e. it consists of all vector valued fields satisfying the adhesion boundary condition and sufficiently smooth (square integrable with square integrable first derivatives) on the uncracked domain. More precisely, the global minimality condition can be expressed as:

$$\text{Find } \gamma\subset\mathcal{C},\ u\in\mathcal{U}(\gamma): \quad \mathcal{E}_{\bar{\epsilon}}(u,\gamma) \le \mathcal{E}_{\bar{\epsilon}}(u^*,\gamma^*), \quad \forall\gamma^*\subset\mathcal{C},\ u^*\in\mathcal{U}(\gamma^*). \quad (6)$$

Remark 1 The scaling factor $\sqrt{G_c/ER}$ in (2) renders all the results, presented in terms of $\bar{\epsilon}$ in the rest of the paper, independent of the material constants and cross-sectional dimension. Other choices for the relevant non dimensional parameter are possible. In particular, as in Gauthier et al. (2010), one could also chose to parameterize the loading in terms of the Griffith length $L_c := EG_c/\sigma_0^2$, where σ_0 is a prestress. After some calculations, one can relate $\bar{\epsilon}$ and L_c by

$$L_c = \frac{R}{\bar{\epsilon}^2}(1-2\nu)^2(1+\nu)^2. \quad (7)$$

This relation will be useful in Sect. 7 as it will allow us to estimate the value of $\bar{\epsilon}$ for various experiments.

4 Simple illustration: star-shaped cracks in circular tubes

Before we dwell upon numerical implementation and numerical experiments, it is possible to gain some

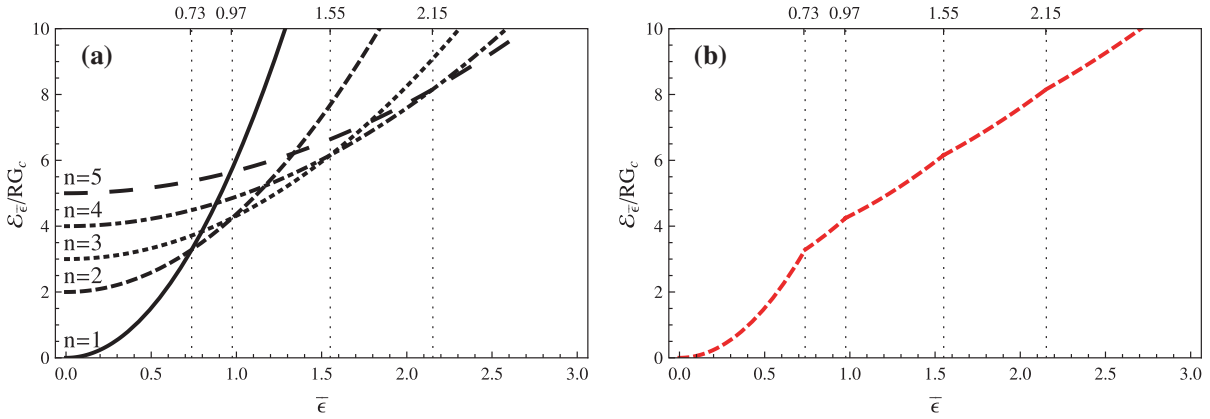

Fig. 2 Energy minimization amongst star-shaped cracks. **a** Total energy associated with γ_n for $n = 1, 2, 3, 4, 5$. **b** The *lower* envelope of the family of *energy curves* gives the energy of the minimizer

valuable insight on the model by restricting the set of admissible cracks to radial ones. It is essentially equivalent to the work of Gauthier et al. (2010) with the difference that the loading parameter we consider here is the inelastic strain instead of the prestress. The construction of the semi-analytic radial crack solution will also be used as a reference point for the analysis of the numerical solutions and the *verification* of the suggested numerical approach.

We consider a circular tube with radius R and star-shaped cracks. By γ_n, $(n > 1)$, we denote a curve consisting of the union of n equi-distributed radial segments partitioning the tube into n polar regions. By analogy, we write $\gamma_1 = \emptyset$. Our motivation for considering such geometries comes from the fact that they are frequently observed in the experiments, at least for small values of n.

Note first that for a given crack pattern γ, the potential energy can be computed by solving a linear elasticity problem, and we write

$$\mathcal{P}_{\bar\epsilon}(\gamma) := \min_{u\in\mathcal{U}(\gamma)} \mathcal{P}_{\bar\epsilon}(u,\gamma), \tag{8}$$

the potential energy of the equilibrium displacement. It is then easy to see that the form of the strain energy density (1) implies that $\mathcal{P}_{\bar\epsilon}(\gamma) = \bar\epsilon^2\mathcal{P}_1(\gamma)$, so that we can rewrite the total energy in the form

$$\mathcal{E}_{\bar\epsilon}(\gamma) = \bar\epsilon^2\mathcal{P}_1(\gamma) + \mathcal{S}(\gamma). \tag{9}$$

Furthermore, for a star-shaped crack γ_n, using eq. (1), one has

$$\mathcal{E}_{\bar\epsilon}(\gamma_n) = \bar\epsilon^2\mathcal{P}_1(\gamma_n) + n\,G_c R. \tag{10}$$

For $n = 1$ the problem can be solved in closed form, the elastic equilibrium is achieved for $u = 0$ and the total energy is $\mathcal{E}_{\bar\epsilon}(\emptyset) = \mathcal{P}_{\bar\epsilon}(\emptyset) = \bar\epsilon^2\pi G_c R/(1+\nu)(1-2\nu)$. For $n > 1$, $\mathcal{P}_1(\gamma_n)$ can be computed by a simple finite element computation. In this setting, for a given loading parameter $\bar\epsilon$, energy minimization reduces to a discrete minimization problem with respect to n. And the total energy of the solution as a function of the loading parameter can be obtained by taking the lower envelope of the family of energy curves associated to each configuration.

Figure 2a represents the total energy associated with γ_n as a function of the loading parameter $\bar\epsilon$ for $n = 1, 2, 3, 4, 5$ and $\nu = 0.3$. Using this graph and the global minimality principle (6), it is easy to identify the optimal crack configuration associated with a given load (the branch of the energy with the smallest value for a given $\bar\epsilon$) as well as the bifurcation points upon which the geometry of the optimal crack set changes (the crossing points upon which the energy branch achieving minimality changes). We obtain that there exists a family $(0 = \bar\epsilon_0, \bar\epsilon_1, \bar\epsilon_2, \dots)$ of critical loadings such that for $\bar\epsilon_{i-1} < \bar\epsilon < \bar\epsilon_i$, $i = 1, 2, \dots$, the optimal crack configuration is *any* curve in the family γ_i. Of course, in the absence of defects or impurities, the solution for a given loading parameter is unique up to a rotation. The numerical values of the critical loadings are $\bar\epsilon_1 \simeq 0.73$, $\bar\epsilon_2 \simeq 0.97$, $\bar\epsilon_3 \simeq 1.55$, $\bar\epsilon_4 \simeq 2.15$, and $\bar\epsilon_5 \simeq 2.76$. Figure 2b shows the energy associated with the optimal configuration, obtained by taking the lower envelope of the family of curves plotted in the left.

Before closing this simple example, we stress again that this analysis is based upon the assumption that the optimal crack geometry is a star-shaped pattern. As we will see in the following sections, relaxing this hypothesis allows to show that the star-shape cracks are indeed optimum for not too high loading and can yield to more complicated but energetically least costly crack patterns for higher loading.

5 General case: minimization over arbitrary crack geometries

In the variational formulation (6), the admissible displacement fields are potentially discontinuous across cracks γ, but the location of the potential discontinuities themselves is not known a priori. This renders the numerical minimization of (4) challenging as most numerical methods such as cohesive, discontinuous or extended finite element methods require at least *some* a priori knowledge of the crack path or of its topology. Indeed, this problem falls into the broader class of *free discontinuity problems* for which a wealth of mathematical and numerical literature now exists (Ambrosio et al. 2000). In the following, we solely focus on the numerical implementation using an extension of that proposed in Bourdin et al. (2000) and Bourdin (2007). It relies on the concept of variational approximation of the total energy functional by a family of regularized elliptic functionals (Ambrosio and Tortorelli 1990, 1992; Braides 1998), depending on a scalar regularization parameter ℓ, and discretized by standard finite elements. In the regularized functional, sharp cracks are represented by a smooth auxiliary variable α in the sense that for small ℓ, the non-zero values of α are localized along thin bands of high strains which can be interpreted as smeared representation of the cracks. From a technical standpoint, the approximation takes place in the sense of Γ-convergence (Braides 2002), i.e. one can prove that as the regularization parameter ℓ goes to 0, the global minimizers and the energy values of the regularized functionals approaches those of the total energy with sharp discontinuities. We briefly recall the numerical approach below and refer the reader to the literature for further details. Note however that all the previous studies on the numerical simulations on fracture mechanics using the variational approach focussed on the *quasi-static* evolution problem. The fact that we consider here the *static* problem entails further issues on the selection of the solutions.

5.1 Regularization by elliptic functionals

The main idea of our approach was originally developed by Ambrosio and Tortorelli (1990, 1992) for an image segmentation (Mumford and Shah 1989) problem, and adapted to fracture mechanics by Bourdin et al. (2000). One introduces a small parameter ℓ with dimension of a length, a secondary variable α taking its values in [0, 1] and representing the crack set, and the regularized functional

$$\mathcal{E}_{\bar{\epsilon}}^{(\ell)}(u,\alpha) := \int_{\mathcal{C}} ((1-\alpha)^2 + k_\ell) w_{\bar{\epsilon}}(\boldsymbol{\epsilon}(u)) \mathrm{d}S + \frac{3G_c}{8} \int_{\mathcal{C}} \left[\frac{\alpha}{\ell} + \ell \, \nabla\alpha \cdot \nabla\alpha \right] \mathrm{d}S. \quad (11)$$

Hence, one approximates the solution of (6) by those of the following minimization problem:

$$\min_{u \in \mathcal{U}, \alpha \in \mathcal{A}} \mathcal{E}_{\bar{\epsilon}}^{(\ell)}(u,\alpha) \quad (12)$$

where $\mathcal{U} = \{u \in H^1(\mathcal{C};\mathbb{R}^2),\ u = 0 \text{ on } \partial\mathcal{C}\}$ and $\mathcal{A} = \{0 \le \alpha \le 1,\ \alpha \in H^1(\mathcal{C};\mathbb{R}),\ \alpha = 0 \text{ on } \partial\mathcal{C}\}$. The main advantages of this regularized formulation is that it eliminates the issue of representing discontinuous fields when their discontinuity set is not known *a priori*. It also reduces energy minimization with respect to any admissible crack geometry to minimization with respect to a smooth field, a much simpler problem. In addition, it can be discretized numerically using standard continuous finite elements.

This regularized functional can be shown to converge in the sense of Γ-convergence to the total energy (4). This implies that for any $\bar{\epsilon}$, the global minimizers of $\mathcal{E}_{\bar{\epsilon}}^{(\ell)}$ converge as $\ell \to 0$ to global minimizers of $\mathcal{E}_{\bar{\epsilon}}$, and that each term in (11) converges to the matching one in (4). The parameter k_ℓ is a small residual stiffness introduced mainly for numerical purposes which is known to have very little impact on the minimizers. The convergence result is valid provided that $k_\ell = o(\ell)$. We refer the interested reader to Braides (2002), Dal Maso (1993) for more details on Γ–convergence and to Braides (1998); Bourdin et al. (2008) for details on the approximation of $\mathcal{E}_{\bar{\epsilon}}$ by $\mathcal{E}_{\bar{\epsilon}}^{(\ell)}$. Formally, as ℓ goes to 0, α remains close to 0 away from the cracks, and approaches 1 along the cracks.

For small but non-zero values of ℓ, both arguments u and α of $\mathcal{E}_{\bar{\epsilon}}^{(\ell)}$ are continuous functions with high gradients (of order $1/\ell$) in bands of width ℓ. The regularized energy we use here is slightly different from the one in Bourdin et al. (2000) and Bourdin et al. (2008). The form used here has some advantages from numerical and theoretical standpoints, which are not discussed here. The interested reader is referred to Pham et al. (2011a) for further details on this point.

Several authors noticed that the regularized form of the energy may be interesting as a model *per se* (Del Piero et al. 2007; Amor et al. 2009; Lancioni and Royer-Carfagni 2009; Freddi and Royer-Carfagni 2010). In particular, it may be interpreted as the energy of a gradient damage model, where α plays the role of the damage field and ℓ of the internal length (Pham et al. 2011a,b). The regularized formulation also resembles phase field approximations of sharp interfaces models (Hakim and Karma 2009; Corson et al. 2009). However, there are significant differences between our numerical approach and the phase-field fracture models. Even in the static or quasi-static case, phase-field models are formulated as a rate-dependent evolution equation and stated as fracture models *per se*.

In the present work, we do not see the regularized formulation as a model in itself, but we consider it as a mere numerical approximation of the total energy of the variational approach of Francfort and Marigo (1998). This approximation is deeply rooted in the mathematical literature on free-discontinuity problems (Braides 1998). In particular the minimization principle for the regularized energy is derived from that of the variational model in the *static* case, which explains our focus on *global* minimizations even though for gradient damage laws, there are *quasi-static* models based on local minimality. This comes with an added difficulty, that of finding *global* minimizers of a non-convex functional. This is a very challenging issue that we do not claim to fully address in this article. Instead, we propose some strategies leading to *local* minima with decreasing energy which are therefore better candidates for *global* optimality.

5.2 Numerical implementation

The numerical minimization of (11) is implemented in a way similar to that described in Bourdin (2007). We discretize the regularized energy by means of linear Lagrange finite elements over an unstructured mesh. As long as the mesh size h is such that $h = o(\ell)$, the Γ-convergence property of (11) to (4) is also true for the discretization of the regularized energy (see Bellettini and Coscia (1994); Bourdin (1999); Burke et al. (2010) for instance). This compatibility condition leads to fine meshes, which are better dealt with using parallel supercomputers. We use PETSc (Balay et al. 1997, 2010, 2011) for data distribution, parallel linear algebra, and TAO (Benson et al. 2010) for the constrained optimization. In order to avoid preferred directions in the mesh, we use the Delaunay-Voronoi mesh algorithm implemented in Cubit, from Sandia National Laboratories (see Negri (1999); Chambolle (1999) for an analysis of the anisotropy induced by structured meshes or grids).

Due to the size of the problems, *global* minimization algorithms are not practical. Instead, we look for local minima by imposing numerically the first-order necessary optimality conditions for (12). We notice that although (11) is not convex, it is convex with respect to each variable individually. We alternate minimizations with respect to u and α, an algorithm akin to a block Newton method or a segregated solver. Note that minimization with respect to u is equivalent to solving a simple linear elasticity problem, but that minimization with respect to $\alpha \in [0, 1]$ requires an actual box-constrained minimization algorithm. Of course, as the total energy is not convex, one cannot expect convergence to a *global* minimizer. However, one can prove that the alternate minimization process is unconditionally stable and globally decreasing and that it leads to a stationary point of (11) which may be a local (or global) minimizer or a saddle point of the energy.

This approach may fall short of our stated goal of achieving *global* minimization of the regularized energy as the outcome of a descent-based algorithms for such a non-convex problem may depend on the regularization parameter ℓ, the mesh size and type, and the starting guess. From a practical standpoint we observe that the algorithm is quite robust with respect to the mesh discretization, provided that the regularization length ℓ is large enough compared to the mesh size. However it can be sensitive with respect to the initial value of α. Different choices of the initial guess α or of the regularization parameter ℓ can lead to convergence to different solutions. This issue is much more troublesome in the present case, in which we attempt to solve

a *static* problem, than when considering a *quasi-static evolution*. In the quasi-static setting, one can follow the evolution of the crack state as the loading increases starting from a well-defined initial state, and formulations based on local minimality of (11) may gain a pertinent physical interpretation, linking the critical load for fracture initiation to the value of the regularization parameter (Pham et al. 2011a). In the present static settings, for the lack of the concepts of history and irreversibility, there is no physically consistent base state and a selection criterion of the solutions obtained by local minimization is necessary. Consistently with the Francfort-Marigo model, we adopt a least energy criterion: among a set of local minima corresponding to a given $\bar{\epsilon}$ and different choices of the initial guesses, we select the one with lowest energy. Again, we make no claim of achieving global minimization of the regularized energy. In our mind, that our numerical simulations match the semi-analytical solutions and the experiments is the best (although still somewhat unsatisfying) argument we can give in favor of this approach.

6 Numerical simulations

We first illustrate our numerical approach on circular tubes (Sect. 6.1). The comparisons of the numerical results with the semi-analytical results for the star-shaped cracks from Sect. 4 gives a partial *verification* of our numerical approach (Sect. 6.2). Results for square tubes are briefly presented in Sect. 6.3.

6.1 Crack shapes for circular tubes

The presentation is organized in three steps: (i) first we report the results obtained as a function of the drying intensity $\bar{\epsilon}$ and the non-dimensional the internal length $\bar{\ell} = \ell/R$, by taking the elastic solution as initial guess for the alternate minimization algorithm; (ii) then we show that a ℓ-refinement technique may reduce the dependence of the results on $\bar{\ell}$ and produce lowest energy solutions; (iii) finally, using the fact that the elastic energy is a 2-homogeneous function of $\bar{\epsilon}$, we identify for each value of the loading the crack shape with the lowest energy among all those obtained after ℓ-refinement.

6.1.1 Results as function of $\bar{\epsilon}$ and $\bar{\ell}$ taking the elastic uncracked state as initial guess

Figure 3 presents the field α obtained by numerical minimization of (11) for various choices of the parameters $\bar{\epsilon}$ and $\bar{\ell}$ with a tube of radius $R = 1$. The material parameters $(E,\ G_c)$ were set to 1 without loss of generality, and the Poisson ratio to 0.3. In each computation, the mesh size was $h = 0.025$ and the residual stiffness was set to $k_\ell = 10^{-6}$. The alternate minimization algorithm was initialized with $\alpha = 0$. The value 1 (corresponding to cracks) of α is encoded in red and the value 0 (the un-cracked material) in blue. A first glance at the table highlights the wide variety of crack geometries obtained, and that the complexity of the fracture pattern increases with the loading parameter. Again we stress that no hypothesis on this geometry is made in the model and that the shape of the crack patterns is purely an outcome of the minimization of the regularized energy. We observe that the width of the transition zone from 1 to 0 decreases as $\bar{\ell}$ goes to 0, which is consistent with the Γ-convergence property stated in Sect. 5. For "large" values of $\bar{\ell}$, when the width of the transition zone is of the order of the diameter of the tube and as $\bar{\epsilon}$ increases, one cannot distinguish between neighboring cracks (see for instance the case $\bar{\ell} = 0.2$ for $\bar{\epsilon} = 5.0$).

A closer look at Fig. 3a highlights the dependency of the crack pattern upon the regularization parameter $\bar{\ell}$. See for instance how for a loading parameter $\bar{\epsilon} = 1.2$, we obtain a triple junction for $\bar{\ell} = 0.2$, but a complex crack made of two triple junctions for $\bar{\ell} = 0.1$, and no cracks at all for $\bar{\ell} = 0.05$. Each of these configuration correspond to a critical point of the energy (11) (likely local minimizers). It may be theoretically shown that below a critical load $\bar{\epsilon}^*(\bar{\ell})$ depending on $\bar{\ell}$, the un-cracked configuration $\alpha = 0$ is a stable critical point of (11). In a simpler 1d setting, it is known that $\bar{\epsilon}^*(\bar{\ell}) = \mathcal{O}(1/\sqrt{\bar{\ell}})$ (Pham et al. 2011a). Here, we observe that $\bar{\epsilon}^*(\bar{\ell})$ increases as $\bar{\ell} \to 0$. When alternate minimizations iterates "escape" the un-cracked solution, they converge to the "nearest" critical point which may or may not be the *global* minimizer of the energy. If the regularized model (11) is seen as a gradient damage model with internal length $\bar{\ell}$ (see Pham et al. 2011a), and if one focusses on criticality instead of global minimality, this behavior is consistent with a *scale effect* linking the critical load and the ratio of the structural dimension and the internal length (Bažant 1999).

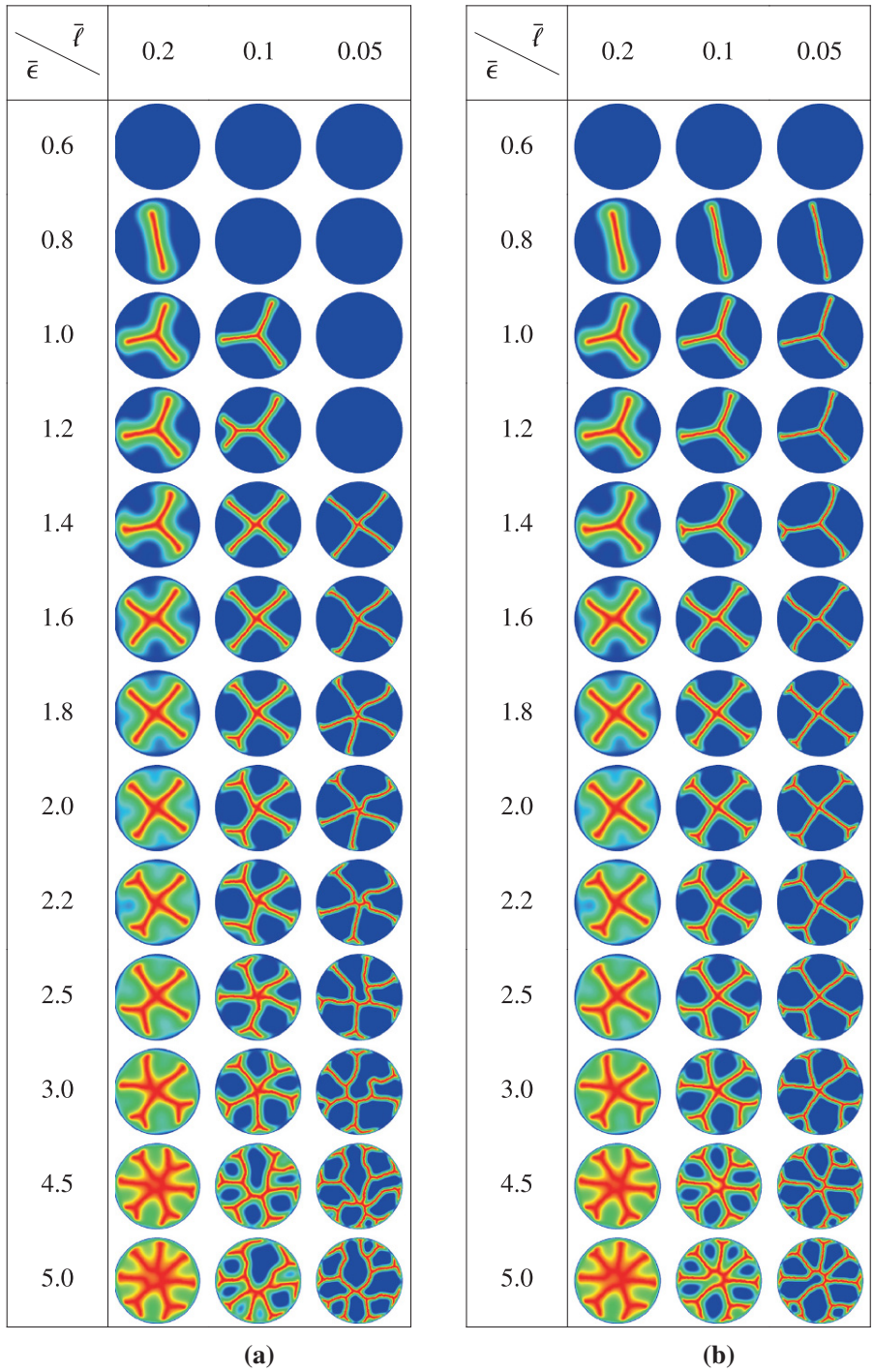

Fig. 3 Numerical results by minimization of functional (11) for circular cross-sections. The results are obtained using uniform Delaunay-Voronoi unstructured meshes with size $h = 0.025$ on disk of radius $R = 1$. The material properties are $E = 1,\ G_c = 1,\ \nu = 0.3$. **a** Direct numerical simulations. Each problem was solved independently initializing the alternate minimizations algorithm with the un-cracked solution $\alpha = 0$. **b** Numerical results obtained using $\bar{\ell}$-refinement. *Each row* corresponds to a set of computations, each taking the one at its *left* as a first guess for u and α

In the case of a quasi-static evolution, Lancioni and Royer-Carfagni (2009) and Pham et al. (2011a) suggested that the internal length can be identified from the critical load at the onset of crack nucleation. In our context, where we only consider a *static* problem and focus on the limiting energy (4), this interpretation is not meaningful. Also, in the experiments, the initial state corresponds to that of the plug with many

existing disordered cracks and the mechanism leading to the selection of the steady state fracture pattern is the self-organization of those cracks and not that of loss of stability of the elastic solution. Hence, we stand by the interpretation that the regularization length is an artificial numerical parameter.

6.1.2 ℓ-refinement

To obtain an accurate resolution for complex crack patterns and, at the same time, avoid the persistence of the purely elastic solution, we implemented an $\bar{\ell}$-*refinement* technique. For a given loading $\bar{\epsilon}$ and mesh size h, we minimize the total energy $\mathcal{E}^{(\bar{\ell})}$ for decreasing values of the regularization length, initializing each computation with the α field obtained at the previous one. Figure 3b represents the outcome of such a series of computations highlighting, how this approach significantly reduces the sensitivity of the numerical results on the regularization length $\bar{\ell}$ and allows us to retrieve an accurate representation of a large family of crack geometries for small and high loadings. The outcome of this set of simulations may still depend on the initial value of $\bar{\ell}$. This value, 0.2 for Fig. 3b, is chosen large enough to avoid the persistence of the purely elastic solution and small enough to allow for localization of α within the domain. In practice, these two criteria do not give a large range for the choice of this parameter. The key properties of the final result of Figs. 3b and of the following Fig. 4b do not sensibly depend on the initial value of $\bar{\ell}$, provided that it fits within the (quite strict) criteria above.

Table 1 provides a quantitative comparison of the energies of the solutions of Fig. 3a, b for $\bar{\ell} = 0.05$. In each case, the configurations obtained using $\bar{\ell}$-refinement (last column of Fig. 3b) have a lesser energy than the one obtained through a direct computation (Fig. 3a). To obtain accurate values of the fracture energies with this choice of $h = \mathcal{O}(\bar{\ell})$, and following the discussion in Bourdin et al. (2008) (Sec 8.1.1 p.103), the reported surface energies are computed using the effective fracture toughness $G_c^{(num)} = G_c/(1 + 3h/8\bar{\ell})$.

6.1.3 Lowest energy crack shapes

For each drying intensity $\bar{\epsilon}$ we select the final cross-sectional crack geometry on the basis of a least energy criterion inspired by the global minimality

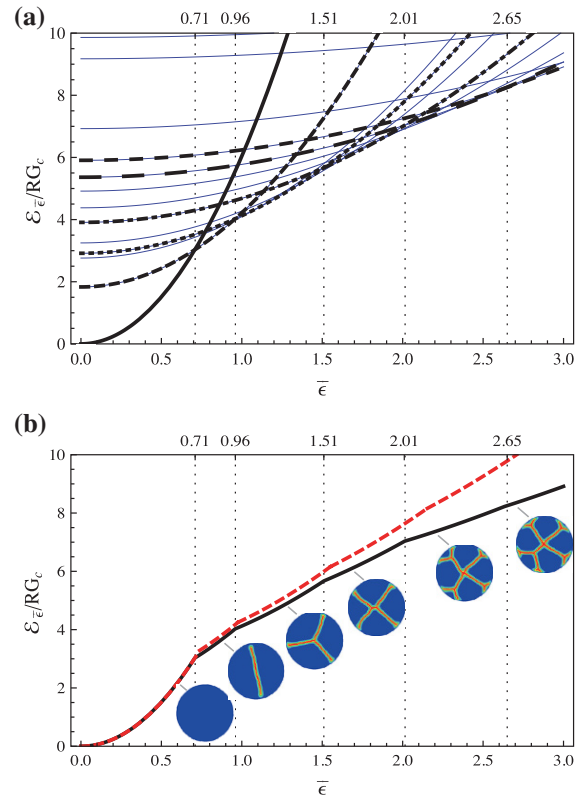

Fig. 4 Minimization over star-shaped cracks vs. minimization of functional (11). As the loading increases, our numerical method identifies crack configurations with much lower energetically than star-shaped cracks. **a** Energies associated with the crack geometries identified in Fig. 3b as a function of the loading parameter. *Thick black lines* distinguishes among all the *energy curves* (*blue lines*), the ones corresponding to configurations attaining the minimal energy for some value of the loading parameters. The *vertical dotted lines* correspond to the critical loading upon which the energy branch for which energy minimality is attained changes. Note that for large $\bar{\epsilon}$ the identification of critical loadings becomes difficult. **b** Comparison of the optimal energy obtained using minimization over star-shaped cracks (*dashed line*) and numerical simulation (*continuous line*). The *continuous line* is the *lower* envelope of the *curves* in Fig. 4a . The pictures represent the optimal crack shapes in each range of the loading parameter delimited by the *vertical dotted lines*

principle (12). To this end, we use a method similar to that in Sect. 4. Noticing that for a given α-field (i.e. crack geometry) the elastic part of the total energy (11) scales quadratically with the loading, one can calculate the total energy that each of the crack patterns obtained numerically for a given loading $\bar{\epsilon}^*$ would have for any $\bar{\epsilon}$. Figure 4a represents the total energy obtained in this way for each of the crack patterns in the last column of Fig. 3b. From this figure, for each $\bar{\epsilon}$, is possible to select the crack pattern with the lowest energy level.

Table 1 Energies of the numerical solutions in Fig. 3 for $\bar{\ell} = 0.05$ without (top) or with (bottom) $\bar{\ell}$-refinement

$\bar{\epsilon}$	Elastic	Surface	Total
0.6	2.2	0	2.2
0.8	3.9	0	3.9
1.0	6.0	0	6.0
1.2	8.7	0	8.7
1.4	1.6	3.9	5.5
1.6	2.0	4.0	6.0
1.8	1.9	4.9	6.8
2.0	2.0	5.3	7.3
2.2	2.1	5.7	7.8
2.5	2.1	6.4	8.5
3.0	2.1	7.3	9.4
4.5	2.3	9.6	11.8
5.0	2.4	9.8	12.2
$\bar{\epsilon}$	Elastic	Surface	Total
0.6	2.2	0	2.2
0.8	1.5	1.8	3.4
1.0	1.3	2.8	4.1
1.2	1.7	2.9	4.6
1.4	2.0	3.3	5.3
1.6	2.0	3.9	5.9
1.8	2.0	4.4	6.4
2.0	2.0	4.9	6.9
2.2	2.0	5.4	7.4
2.5	2.1	5.9	8.0
3.0	2.1	6.9	9.1
4.5	2.3	9.2	11.5
5.0	2.4	9.9	12.3

Amongst all the curves, the ones attaining the lowest energy for some value of the loading parameter are plotted in black and thicker line width. Their lower envelope is the continuous black line reported in Fig. 4b, together with the associated optimal crack geometries. Although it is of course never possible to prove global optimality, the crack geometries depicted in Fig. 4b are the lowest energy configurations we were able to attain, and the ones which we will compare with star-shaped cracks and experiments in the sequel.

6.2 Comparison with star-shaped cracks

Figure 4b compares the total energy of these configurations with the energy of the star-shaped cracks taken from Fig. 2 (dashed line). For small values of the loading parameter, we obtain similar geometries and critical loading. The surface energy obtained is close to the number of branches, and the critical loading upon which we obtain a single straight crack is 0.71 (vs. a theoretical value of $\bar{\epsilon}_2 = 0.73$). Bifurcation between straight and Y-shaped cracks take place at $\bar{\epsilon} \simeq 0.94$ (vs. a theoretical value of $\bar{\epsilon}_2 = 0.97$). More interestingly, for larger values of $\bar{\ell}$, our numerical simulations have identified multiple configurations that are energetically close to each other but always less expensive than star-shaped cracks. In particular, we show that perfect 5-branch stars are never optimal and that configurations consisting of either two triple junctions very close to each others (see for instance $\bar{\epsilon} = 1.6$ in Fig. 3b), a 4-branch star whose branches split in two near the tube boundary (see for instance $\bar{\epsilon} = 1.8, 2.0, 2.2$ in Fig. 3b), or a more complicated patterns like the "stick figure" looking 5 cracks configuration that we obtain for $\bar{\epsilon} = 2.5$ have lesser energy. Of course, that the local geometry near the crack crossing resembles 2 triple junctions near each others rather that an "X" does not really come up as a surprise. As mentioned earlier, the fracture energy (4) resembles the Mumford-Shah energy for edge segmentation (Mumford and Shah 1989). Therefore, it seems natural to expect that if they posses some form of regularity, optimal crack geometries satisfy the Mumford-Shah conjecture which rules out crack crossing, kinks and only allows cracks to meet at 120° triple junctions, locally.

6.3 Square tubes

Finally, we performed another set of numerical simulations on unit square tubes. The results obtained by the same $\bar{\ell}$-refinement method, as in Fig. 3b for circular tubes, are depicted on Fig. 5. The materials parameters and mesh size are unchanged ($E = 1$, $G_c = 1$, $\nu = 1$, $h = 0.025$). Whereas for circular tubes, star-shaped cracks are natural candidates, there were no obvious family of cracks in this case. This geometry also leads to a rich variety of crack patterns and highlights the strength of the proposed method in identifying complex crack patterns without *a priori* hypothesis. Some of the quantitative properties of the optimal cracks highlighted in the case of circular tubes are still observed. Again, cracks seem to split near the

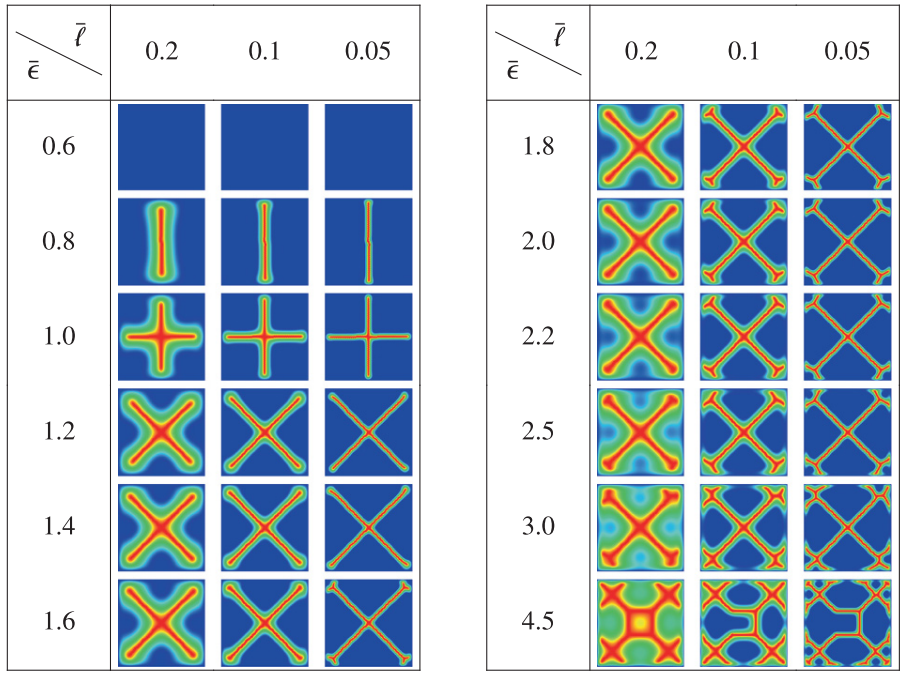

Fig. 5 Numerical results for *square tubes* by minimization of functional (11) using the $\bar{\ell}$-continuation method

edges of the tubes. Again, triple junctions seem to be favored over crack crossing, although in the case of two diagonal or longitudinal cracks, the resolution of our numerical experiments does not allow us to clearly identify the configuration. As for the circular tube, one can further post process the numerical result in order to identify the range of loadings for which each of the identified configuration is optimal. This is presented in Fig. 6. Again, for small values of the loading parameters, simple and somewhat predictable crack geometries are obtained. For larger values of $\bar{\epsilon}$, more complex and less intuitive patterns are energetically more advantageous.

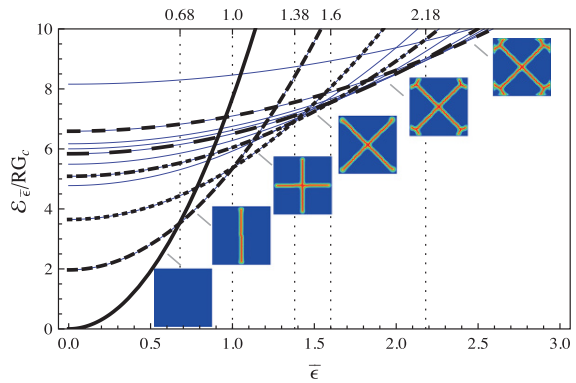

Fig. 6 Range of parameters in which the configuration identified in Fig. 5 are optimal. The *blue lines* correspond to the energy associated with cracks patterns that were identified in Fig. 5 but are never optimal

7 Comparison between experimental and numerical results

7.1 Identification of the loading parameter $\bar{\epsilon}$

Dimensional analysis shows that the model relies on a single parameter, $\bar{\epsilon}$ whose value needs to be estimated in order to perform quantitative comparison between experiments and numerical simulation. As $\bar{\epsilon}$ depends on experimental conditions, colloidal suspension type, and tube geometry, one solution is to try to measure separately ϵ_0, E, and G_c appearing in the definition (2) of $\bar{\epsilon}$. One may obtain the material constants E, G_c by indentation (Malzbender et al. 2002) and the mismatch strain ϵ_0 by beam deflection technics (Tirumkudulu and Russel 2004; Chekchaki et al. 2011) from a thin film drying experiments, for instance. However, such direct measurements are difficult, and transposing the values obtained from one type of experiments (thin film drying) to another (directional drying) is questionable. Indeed, the parameters may depend on the type of

Table 2 Values of Griffith's length L_c (in μm) for several Ludox®suspensions and drying rates (SV = [$T \simeq 3\,^\circ$C and RH $\leq$10 % or at $T \simeq 20\,^\circ$C and RH $\geq$90 %] and HV = [$T \simeq 20\,^\circ$C and RH $\leq$10 %])

L_c	Ludox®SM30 ($r \simeq 3.5$ nm)	Ludox®HS40 ($r \simeq 6$ nm)
HV	34 ± 10	40 ± 10
SV	60 ± 18	45 ± 15

experiments and even evolve in time. For example, the material constants E and G_c of the porous medium may depend on the microstructure, influenced by formation dynamics.

Instead of performing such difficult measurements, whose relevance to our problem may be questionable, we used the method presented in Gauthier et al. (2010), which we briefly summarize. The basis of the method is to consider a directional drying experiment in thin rectangular tubes (Allain and Limat 1995). In this geometry, an array of parallel tunneling cracks is obtained and the cracks spacing λ can be correlated with the Griffith length $L_c := EG_c/\sigma_0^2$ (σ_0 being the prestress induced by the film's drying). Using an energy minimization principle similar to the one in Sect. 4, one can show that the spacing λ is proportional to $\sqrt{L_c t}$, t being the tube's thickness and in particular, for $\nu = 0.3$, one obtains $\lambda \sim 3.1\sqrt{L_c t}$. For a given material and drying parameter, the value of L_c can therefore be deduced from measurements of λ. Table 2 presents the value of L_c for Ludox®SM30 ($r \simeq 3.5$ nm) and Ludox®HS40 ($r \simeq 6$ nm) under high velocity and slow velocity conditions.

We assume that the Griffith length L_c is a well-defined parameter for a given suspension and drying condition, and that it is independent of the cross-sectional geometry of the directional drying experiment. Hence, from the values of L_c in Table 2, we estimate the value of $\bar{\epsilon}$ in the directional drying of circular and square tubes of different diameters using the relation (7), which gives $\bar{\epsilon} = 0.52\sqrt{R/L_c}$ for $\nu = 0.3$.

7.2 Results and analysis

Table 3 reports on the series of experiments on circular tubes described in Sect. 2. From a *qualitative* standpoint we observe that star-shaped appear above a critical load, and that the number of branches increases with the loading, which is consistent with the analysis in Sect. 4 and the numerical simulations of Sect. 6. In order to perform a *quantitative* comparison, we summarize all the results obtained in the case of circular tubes in Fig. 7. The first row corresponds to the outcome of the semi-analytical minimization over star-shaped cracks: the critical values of the loading parameters computed in Sect. 4 upon which bifurcation between different morphologies take place is printed in red letters, and represented by red dashed vertical lines. The second row corresponds to the numerical experiments without *a priori* hypotheses on the crack path. The critical loads extracted from Fig. 4a are printed in black letters, and represented by vertical solid black lines. As highlighted in Fig. 4b, the critical loads obtained in the case of the bifurcation from γ_1 to γ_2, then γ_3 and γ_4 are very close. This part of the table can be seen as a *verification* of the numerical implementation, i.e. as evidences that the computed solutions are indeed solution of the variational fracture model. The third row summarizes the outcome of the experiments from Table 3. For each experiment, the value of the loading parameter is shown together with the accuracy of the measurement. When available, pictures of the cross sections are also displayed. We observe that for every single choice of $\bar{\epsilon}$, the crack geometry predicted by our approach matches the one observed in the experiment. This acts as a *validation* of the variational fracture model as a predictive tool in the setting of drying of colloidal suspension.

We also did a single experiment on a square tube, for an estimated value $\bar{\epsilon} \simeq 1.8$ of the loading parameter for which we obtained two diagonal cracks (see the bottom image in Figure 1b). Again, the numerical simulation in this case matches the experiment (see Figs. 5 and 6).

Despite the modeling simplifying assumption, the complexity of the numerical technique, and the uncertainty of the measurement of the parameters, the agreement between analysis, simulation, and experiments is excellent. Our model correctly captures the essential physics of the crack formation giving credit to the idea that crack growth can be predicted by minimization of the sum of elastic and surface energy over all possible crack path. In order to further justify this idea, one will need to compare experiments and simulations for higher loading parameters $\bar{\epsilon}$ in which case numerical simulations identify complex crack patterns with significantly lower energy than classical star-shaped solutions. For instance, better quality imaging will be required to unambiguously determine if the 5 cracks

Table 3 Experimental results on circular tubes. The value of n corresponds to the number of sectors delimited by the cracks as in Sect. 4

	Ludox®	HV/SV	L_c(μm)	R (μm)	n	$\bar{\epsilon}$
a	SM30	HV	34 ±10	50 ± 5	1	0.6 ± 0.12
b	SM30	SV	60 ±18	150 ± 15	2	0.8 ± 0.16
c	SM30	SV	60 ±18	500 ± 50	3	1.5 ± 0.30
d	HS40	SV	45 ±15	500 ± 50	4	1.7 ± 0.37
e	HS40	HV	40 ±10	500 ± 50	4	1.8 ± 0.27
f	SM30	HV	34 ±10	500 ± 50	4	2.0 ± 0.39
g	HS40	HV	40 ±10	750 ± 75	5	2.3 ± 0.41
h	SM30	HV	34 ±10	750 ± 75	5	2.4 ± 0.48

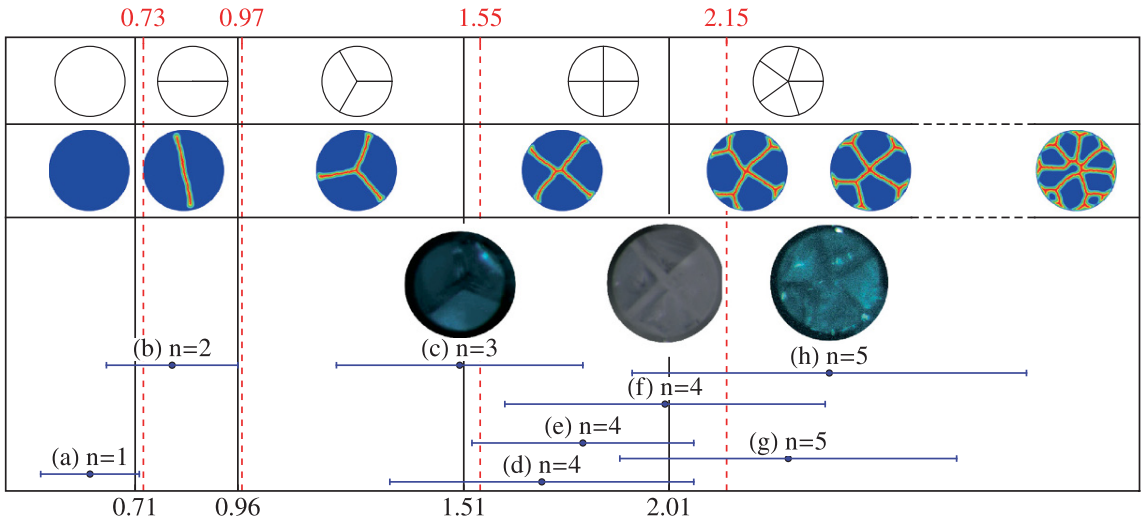

Fig. 7 Comparison between semi-analytical, numerical and experimental results for circular tubes

configurations we observe experimentally (see Fig. 1b or 7) resembles a "stick figure" as in our numerical simulation (cf. Fig. 3 for $\bar{\epsilon} = 2.5$), a regular 5-branch star, or something completely different.

8 Conclusions and future work

In this paper, we have shown that a numerical implementation (Bourdin et al. 2000) of the variational approach to fracture mechanics (Francfort and Marigo 1998) enhanced with a selection criterion is capable to qualitatively and quantitatively predict complex crack shapes starting from an undamaged material. For this, we have studied some unidirectional drying experiments of colloidal suspensions performed in capillary tubes where solvent evaporation leads to the formation of a growing porous solid medium. In these experiments, adhesion on the tubes walls combined with shrinkage lead to high tensile stresses and give rise to cracks whose morphologies depend on the tube geometry and the drying velocities. We verified that the changes in crack geometry can be accounted by a two-dimensional static simple model depending on a single dimensionless parameter $\bar{\epsilon}$ which represents the ratio of the bulk elastic energy (which depends on the intensity of the tensile strain induced by drying) over the cost of fracture. Under the assumption that cracks are star-shaped, the number of branches has been obtained as a function of the loading parameter.

We then presented a numerical method based on the regularization of the total energy introduced in the variational model, and enhanced it with a selection principle and a refinement method. For small enough values of the loading, this allowed us to verify that star-shaped cracks are actually favored in circular tubes, to predict cracks shapes in square tubes and to achieve qualitative and quantitative agreement between numerical simulation, semi-analytical solutions and experiments. This results are summarized in Fig. 7. For higher values of the loading parameters, more complex cracks geometries are observed. These are the situations where the virtue of the variational approach to fracture over more conventional ones requiring at least some *a priori* knowledge of the crack path becomes more striking. But at this point, though, we were not able to compare them with experiments that will require additional work in order to deal with larger tubes for instance. In these situations, the main difficulty is the *post-mortem* analysis of the crack geometry. Microphotography though the sides of the tubes becomes hard to interpret, and cutting the tubes without perturbing the cracks geometry is difficult. Perhaps the solution lies in full three-dimensional imaging of the tubes and post-processing in order to highlight the location of the cracks.

From the modeling perspective, a full three-dimensional linear poroelasticity model (Biot 1941) would be welcome in particular to study the entire propagation from the crack initiation to the crack self-organisation. The complexity of this task mainly lies in the time-dependence of the solid domain and of the material constants. Finally, from a physico-chemical point of view, the link between the drying velocity and the macroscopic $\bar{\epsilon}$ signature will have to be explored.

Acknowledgments The work of VL and GG was partially supported by the ANR Program JC-JC ANR-05-JCJC-0029 Morphologies. CM gratefully acknowledges the funding of the French National Research Council (CNRS) for a PICS bilateral exchange program with BB and a grant of the University Pierre et Marie Curie EMERGENCES-UPMC. BB's work was supported in part by the National Science Foundation under the grant DMS-0909267. Some of the numerical experiments were performed using resources of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575, provided by TACC at the University of Texas under the Resource Allocation TG-DMS060014N and the resources of the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS) under the DARI 2011 allocation 100064.

References

Adda-Bedia M, Pomeau Y (1995) Crack instabilities of a heated glass strip. Phys Rev E 52(4, Part b):4105–4113

Allain C, Limat L (1995) Regular patterns of cracks formed by directional drying of a colloidal suspension. Phys Rev Lett 74:2981–2984

Ambrosio L, Fusco N, Pallara D (2000) Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs, Oxford Science Publications

Ambrosio L, Tortorelli V (1990) Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. Commun Pure Appl Math 43(8):999–1036

Ambrosio L, Tortorelli V (1992) On the approximation of free discontinuity problems. Boll Un Mat Ital B (7) 6(1):105–123

Amor H, Marigo J-J, Maurini C (2009) Regularized formulation of the variational brittle fracture with unilateral contact: Numerical experiments. J Mech Phys Solids 57(8): 1209–1229

Bahr H-A, Hofmann M, Weiss H-J, Bahr U, Fischer G, Balke H (2009) Diameter of basalt columns derived from fracture mechanics bifurcation analysis. Phys Rev E (Stat Nonlinear Soft Matter Phys) 79(5):056103

Bahr H, Weiss H, Bahr U, Hofmann M, Fischer G, Lampenscherf S, Balke H (2010) Scaling behavior of thermal shock crack patterns and tunneling cracks driven by cooling or drying. J Mech Phys Solids 58(9):1411–1421

Balay S, Brown J, Buschelman K, Eijkhout V, Gropp W, Kaushik D, Knepley M, Curfman McInnes L, Smith B, Zhang H (2010) PETSc users manual. Technical report ANL-95/11, revision 3.1, Argonne National Laboratory

Balay S, Brown J, Buschelman K, Gropp W, Kaushik D, Knepley M, McInnes LC, Smith B, Zhang H (2011) PETSc Web page. URL:http://www.mcs.anl.gov/petsc

Balay S, Gropp W, Curfman McInnes L, Smith B (1997) Efficient management of parallelism in object oriented numerical software libraries. In: Arge E, Bruaset AM, Langtangen HP (eds) Modern software tools in scientific computing. Birkhäuser Press, London, pp 163–202

Bažant Z (1999) Size effect on structural strength: a review. Arch Appl Mech 69:703–725

Bažant Z, Ohtsubo H, Aoh K (1979) Stability and post-critical growth of a system of cooling or shrinkage cracks. Int J Fract 15(5):443–456

Bellettini G, Coscia A (1994) Discrete approximation of a free discontinuity problem. Numer Funct Anal Optim 15(3–4): 201–224

Benson S, Curfman McInnes L, Moré J, Munson T, Sarich J (2010) TAO user manual (revision 1.10.1). Technical report ANL/MCS-TM-242, Mathematics and Computer Science Division, Argonne National Laboratory. URL:http://www.mcs.anl.gov/tao

Biot MA (1941) General theory of 3-dimensional consolidation. J Appl Phys 12:155–164

Bourdin B (1999) Image segmentation with a finite element method. M2AN Math Model Numer Anal 33(2):229–244

Bourdin B (2007) Numerical implementation of the variational formulation of quasi-static brittle fracture. Interfaces Free Bound 9:411–430

Bourdin B, Francfort GA, Marigo JJ (2000) Numerical experiments in revisited brittle fracture. J Mech Phys Solids 48(4):797–826

Bourdin B, Francfort GA, Marigo J-J (2008) The variational approach to fracture. J Elast 91(1):5–148

Braides A (1998) Approximation of free-discontinuity problems, vol. 1694 of Lecture Notes in Mathematics, Springer

Braides A (2002) Γ-convergence for beginners, Vol. 22 of Oxford Lecture Series in mathematics and its applications. Oxford University Press, Oxford

Branagan D, Cairns H (1993) Tessalated pavements in the Sydney region, New South Wales. J Proc Roy Soc New South Wales 126(1):63–72

Burke S, Ortner C, Süli E (2010) An adaptive finite element approximation of a variational model of brittle fracture. SIAM J Numer Anal 48(3):980–1012

Chambolle A (1999) Finite-differences discretizations of the Mumford-Shah functional. M2AN Math Model Numer Anal 33(2):261288

Chekchaki M, Frelat J, Lazarus V (2011) Analytical and 3D finite element study of the deflection of an elastic cantilever bilayer plate. J Appl Mech T ASME 78(1):011008

Colina H, Roux S (2000) Experimental model of cracking induced by drying shrinkage. Eur Phys J E 1(2–3):189–194

Corson F, Adda-Bedia M, Henry H, Katzav E (2009) Thermal fracture as a framework for quasi-static crack propagation. Int J Fract 158(1):1–14

Dal Maso G (1993) An introduction to Γ-convergence. Birkhäuser, Boston

DeGraff J, Aydin A (1987) Surface morphology of columnar joints and its significance to mechanics and direction of joint growth. Geol Soc Am Bull 99:600–617

Del Piero G, Lancioni G, March R (2007) A variational model for fracture mechanics: numerical experiments. J Mech Phys Solids 55(12):2513–2537

Dufresne E, Corwin E, Greenblatt N, Ashmore J, Wang D, Dinsmore A, Cheng J, Xie X, Hutchinson J, Weitz D (2003) Flow and fracture in drying nanoparticle suspensions. Phys Rev Lett 91(22):224501

Dufresne E, Stark D, Greenblatt N, Cheng J, Hutchinson J, Mahadevan L, Weitz D (2006) Dynamics of fracture in drying suspensions. Langmuir 22(17):7144–7147

Francfort GA, Marigo J-J (1998) Revisiting brittle fracture as an energy minimization problem. J Mech Phys Solids 46: 1319–1342

Freddi F, Royer-Carfagni G (2010) Regularized variational theories of fracture: a unified approach. J Mech Phys Solids 58(8):1154–1174

Gauthier G, Lazarus V, Pauchard L (2007) Alternating crack propagation during directional drying. Langmuir 23(9): 4715–4718

Gauthier G, Lazarus V, Pauchard L (2010) Shrinkage star-shaped cracks: explaining the transition from 90 degrees to 120 degrees. EPL 89:26002

Goehring L, Morris SW, Lin Z (2006) Experimental investigation of the scaling of columnar joints. Phys Rev E (Stat Nonlinear Soft Matter Phys) 74(3):036115

Griffith A (1920) The phenomena of rupture and flow in solids. Philos Trans R Soc Lond 221:163–198

Hakim V, Karma A (2009) Laws of crack motion and phase-field models of fracture. J Mech Phys Solids 57(2):342–368

Hofmann M, Bahr H, Linse T, Bahr U, Balke H, Weiss H (2006) Self-driven tunneling crack arrays—a 3D-fracture mechanics bifurcation analysis. Int J Fract 141(3–4):345–356

Jenkins D (2009) Determination of crack spacing and penetration due to shrinkage of a solidifying layer. Int J Solids Struct 46(5):1078–1084

Lancioni G, Royer-Carfagni G (2009) The variational approach to fracture mechanics: a practical application to the french Panthéon in Paris. J. Elast 95(1–2):1–30

Lazarus V, Pauchard L (2011) From craquelures to spiral crack patterns: influence of layer thickness on the crack patterns induced by desiccation. Soft Matter 7(6):2552–2559

Malzbender J, den Toonder JMJ, Balkenende AR, de With G (2002) Measuring mechanical properties of coatings: a methodology applied to nano-particle-filled sol-gel coatings on glass. Mater Sci Eng R 36(2–3):47–103

Mangold N (2005) High latitude patterned grounds on mars: classification, distribution and climatic control. Icarus 174(2): 336–359

Muller G (1998) Starch columns: analog model for basalt columns. J Geophys Res 103(B7):15239–15253

Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Commun Pure Appl Math 42:577–685

Nemat-Nasser S, Sumi Y, Keer L (1980) Unstable growth of tension cracks in brittle solids - stable and unstable bifurcations, snap-through, and imperfection sensitivity. Int J Solids Struct 16(11):1017–1035

Negri N (1999) The anisotropy introduced by the mesh in the finite element approximation of the Mumford-Shah functional. Numer Funct Anal Optim 20(9–10):957–982

Pham K, Amor H, Marigo J-J, Maurini C (2011a) Gradient damage models and their use to approximate brittle fracture. Int J Damage Mech 20(4):618–652

Pham K, Marigo J-J, Maurini C (2011b) The issues of the uniqueness and the stability of the homogeneous response in uniaxial tests with gradient damage models. J Mech Phys Solids 59(6):1163–1190

Plug L, Werner B (2001) Fracture networks in frozen ground. J Geophys Res Earth 106(B5):8599–8613

Ronsin O, Heslot F, Perrin B (1995) Experimental study of quasistatic brittle crack propagation. Phys Rev Lett 75(12): 2352–2355

Saliba R, Jagla EA (2003) Analysis of columnar joint patterns from three-dimensional stress modeling. J Geophys Res Solid Earth 108(B10):2476

Seilacher A (2001) Concretion morphologies reflecting diagenetic and epigenetic pathways. Sediment Geol 143(1–2): 41–57

Shinn E (2009) The mystique of beachrock. Special publication of the International Association of Sedimentologists 41(1):19–28

Tirumkudulu M, Russel W (2004) Role of capillary stresses in film formation. Langmuir 20(7):2947–2961

Toramaru A, Matsumoto T (2004) Columnar joint morphology and cooling rate: a starch-water mixture experiment. J Geophys Res 109(B2):B02205

Yang B, Ravi-Chandar K (2001) Crack path instabilities in a quenched glass plate. J Mech Phys Solids 49(1):91–130

Yuse A, Sano M (1993) Transition between crack patterns in quenched glass plates. Nature 362(6418):329–331

DOI 10.1007/978-3-319-04397-5_8
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9839-y

ORIGINAL PAPER

Damage mechanisms in the dynamic fracture of nominally brittle polymers

Davy Dalmas · Claudia Guerra · Julien Scheibert · Daniel Bonamy

Received: 19 November 2012 / Accepted: 19 March 2013 / Published online: 9 April 2013

Abstract Linear elastic fracture mechanics provides a consistent framework to evaluate quantitatively the energy flux released to the tip of a growing crack. Still, the way in which the crack selects its velocity in response to this energy flux remains far from completely understood. To uncover the underlying mechanisms, we experimentally studied damage and dissipation processes that develop during the dynamic failure of polymethylmethacrylate, classically considered as the archetype of brittle amorphous materials. We evidenced a well-defined critical velocity along which failure switches from nominally-brittle to quasi-brittle, where crack propagation goes hand in hand with the nucleation and growth of microcracks. Via post-mortem analysis of the fracture surfaces, we were able to reconstruct the complete spatiotemporal microcracking dynamics with micrometer/nanosecond resolution. We demonstrated that the true local propagation speed of individual crack fronts is limited to a fairly low value, which can be much smaller than the apparent speed measured at the continuum-level scale. By coalescing with the main front, microcracks boost the macroscale velocity through an acceleration factor of geometrical origin. We discuss the key role of damage-related internal variables in the selection of macroscale fracture dynamics.

Keywords Dynamic fracture · Amorphous polymers · Fracture energy · Fractography · High-deformation rate · Microcracks · PMMA

D. Dalmas
Unité Mixte CNRS/Saint-Gobain, Surface du Verre et Interfaces, 39 Quai Lucien Lefranc, 93303 Aubervilliers Cedex, France

C. Guerra · J. Scheibert · D. Bonamy (✉)
CEA, IRAMIS, SPCSI, Group Complex Systems and Fracture, 91191 Gif sur Yvette, France
e-mail: Daniel.Bonamy@cea.fr

J. Scheibert
Laboratoire de Tribologie et Dynamique des Systèmes, CNRS, Ecole Centrale de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully Cedex, France

1 Introduction

Dynamic crack propagation drives catastrophic material failure and is usually described using the linear elastic fracture mechanics (LEFM) framework (Freund 1990; Ravi-Chandar 2004). This theory considers the straight propagation of a single smooth crack and assumes all energy dissipating processes to be entirely localized in a small zone at the crack tip, so called fracture process zone (FPZ). Crack growth velocity v is then selected by the balance between the mechanical energy that flows within the FPZ per time unit and the dissipated energy within the FPZ over the same time unit. This yields (Freund 1990):

$$\Gamma \simeq (1 - v/C_R)\, K^2(c)/E, \tag{1}$$

where C_R and E are the Rayleigh wave speed and the Young modulus of the material, respectively, Γ is the

fracture energy, and $K(c)$ is the Stress Intensity Factor (SIF) for a quasi-static crack of length c. K only depends on the applied loading and the cracked specimen geometry, and entirely characterizes the stress field in the vicinity of the crack front.

Linear elastic fracture mechanics predictions agree well with observations as long as the crack growth is sufficiently slow (Bergkvist 1974). However, large discrepancies are reported at high speed (Ravi-Chandar 2004; Bouchbinder et al. 2010). In particular, the maximal crack speeds attained experimentally in amorphous materials are far smaller (typically by a factor of two) than the limiting speed C_R predicted by Eq. 1.

The existence of a micro-branching instability (Fineberg et al. 1992) at a critical velocity v_b (typically of the order of 0.35–$0.4C_R$) permits to explain *part* of this discrepancy: Beyond v_b, the crack front splits into a multiple crack state that cannot be described by Eq. 1 anymore. However, it is commonly stated (Sharon and Fineberg 1999) that LEFM works well below v_b, or more generally when microbranching instabilities are absent. As a matter of fact, most recent works focused on failure regimes beyond v_b (see e.g. Gumbsch et al. 1997; Adda-Bedia et al. 1999; Henry and Levine 2004; Bouchbinder et al. 2005; Spatschek et al. 2006; Henry 2008). Still, several observations reported at lower speeds remain puzzling. In particular, even for velocities much lower than v_b, the measured dynamic fracture energy is generally found to be much higher than that at crack initiation (Sharon and Fineberg 1999; Kalthoff et al. 1976; Rosakis et al. 1984; Fond and Schirrer 2001).

The experiments reported here were designed to better understand the mechanisms that select crack velocity in dynamic fracture. In this context, we developed an experimental setup (presented in Sect. 2) that permits to characterize over a wide range of crack speeds, in particular speeds below v_b, the dissipative and damage processes that develop in polymethylmethacrylate (PMMA), classically considered as the archetype of brittle amorphous materials. This setup allowed us to evidence a novel critical velocity, smaller than v_b, beyond which crack propagation goes hand in hand with the nucleation and growth of microcracks ahead of the main crack front (Sect. 3). Via accurate *post-mortem* analysis of the patterns let on fracture surfaces, we were able to reconstruct the full spatio-temporal dynamics of these microfailure events (Sect. 4). Their statistics and their dependency with crack tip loading have been characterized (Sect. 5). In Sect. 6, we will show how microcracking acts and selects the apparent velocity measured at the continuum-level scale.

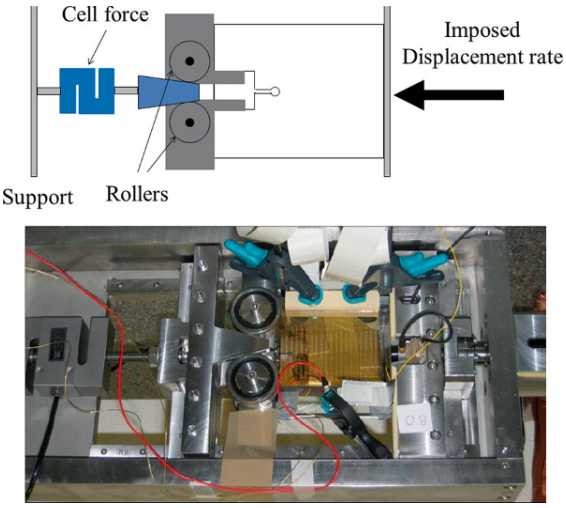

Fig. 1 Sketch (*top*) and photo (*bottom*) of the experimental setup

2 Experimental setup

Fracture tests were performed in the wedge splitting geometry (Bruhwiler and Wittmann 1990; Karihaloo and Xiao 2001) depicted in Fig. 1. Specimens were prepared from PMMA parallelepipeds of size $140 \times 125 \times 15\,\mathrm{mm}^3$ in the propagation (x-axis), loading (y-axis), and thickness (z-axis) directions, respectively. Table 1 gives the main characteristics of the PMMA used in this study. Subsequently, a notch was shaped (1) by cutting a $25 \times 25\,\mathrm{mm}^2$ rectangle from the middle of one of the $125 \times 15\,\mathrm{mm}^2$ edges; and (2) by subsequently adding a 10-mm groove deeper into the specimen. Two steel blocks equipped with rollers were then placed on both sides of this notch and the specimen was loaded by pushing a wedge (semi-angle of 15°) at a small constant velocity ($40\,\mu\mathrm{m/s}$) between these two blocks. This permits (1) to spread the loading force over a large contact area and prevent any plastic deformation of PMMA at the loading contacts; and (2) to suppress friction in the system. As a result, the vicinity of the crack tip can be assumed to be the sole dissipation source for mechanical energy in the system.

In such a wedge splitting geometry, the SIF decreases with crack length. To obtain dynamic failure, we then introduced a circular hole of tunable diameter from 2 to 8 mm at the tip of the seed crack. This hole blunts

Table 1 Mechanical properties of the PMMA used in our fracture experiments

	E(GPa)	ν	Density (kg/dm^3)	C_d (m/s)	C_s (m/s)	C_R (m/s)
PMMA	2.8 ± 0.2	0.36 ± 0.08	1.18 ± 0.02	2010 ± 60	950 ± 10	880 ± 30

E and ν denote the Young modulus and the Poisson ratio, respectively. C_D, C_S, and C_R denote the speeds of dilational, Rayleigh waves respectively

the seed crack and delays the propagation onset. This increases the mechanical energy stored in the sample at the onset of crack propagation. It allowed us to widen the range of SIF (from 0.9 to 4.5 MPa $\sqrt{\text{m}}$) and that of crack velocity (from 75 m/s $\approx 0.08C_R$ to 500 m/s $\approx 0.57C_R$) accessible in the experiments.

The apparatus used to carry out the experiments is homemade (Fig. 1). One side of the machine is fixed while the other moves via a stepper motor (Oriental motor EMP400 Series) allowing incremental displacements as small as 40 nm. The compressive force is measured, up to 20 kN via a S-type Vishay load cell 363 Series. In all performed tests, both the displacement and the force are recorded via a computer with a time resolution of 1 s. In particular, the value of the applied force at propagation onset is recorded and later used in the finite element analyses described thereafter.

To measure the instantaneous crack velocity, we used a modified version of the electrical resistance grid technique (Cotterell 1968; Anthony et al. 1970; Boudet et al. 1996; Fineberg et al. 1992). A series of 90 parallel conductive lines (2.4-nm-thick Cr layer covered with 23-nm-thick Au layer) were deposited on one of the two $140 \times 125\,\text{mm}^2$ sides of the specimen using magnetron sputtering deposition through a polymer mask glued on the sample surface. Stripes width and length are 500 μm and 100 mm, respectively, and they are separated by gaps of 500 μm. Once the sample is installed in the compression machine, each line is connected in series with a resistor of 30 kΩ—this value was chosen to be extremely high with respect to that of the strip (about few ohms) so that the latter is negligible. As the crack propagates throughout the specimen, it successively cuts the conductive lines, which, each time, makes the global system resistance increase by a constant increment. The time locations of these jumps are detected via an oscilloscope (acquisition rate: 10 MHz) through a voltage divider circuitry. In order to increase the accuracy of jump detection, we used the four channels of the oscilloscope in cascade with different offsets, so that each channel is triggered as the previous one gets out of its range (Fig. 2a, b). As a result, we obtained the crack length $c(t)$ as a function of time with space and time accuracies of 40 μm and 0.1 μs, respectively (Fig. 2c). The profile of instantaneous velocity $v(c)$ as a function of crack length can then be easily deduced (Fig. 2d).

Finite element analysis was used to estimate the SIF evolution during failure. Figure 3:Top represents the meshing of the complete system. The average mesh size is 1 mm and reduces logarithmically down to 1 nm at the crack tip (Fig. 3:Bottom), in order to resolve in space the crack tip opening. Only half of the system is required for symmetry reasons. Young modulus and Poisson ratio are set to 2.8 GPa and 0.36, respectively (see Table 1). The boundary conditions are the following: (1) The left edge is perfectly adhesive to a polymeric layer of thickness 5 mm, Young's modulus 3 GPa and Poisson ratio 0.41; (2) The left edge of the layer has a no-displacement condition; (3) On the right, the notch is loaded through a L-shaped block, of thickness 5 mm, Young modulus 300 GPa and Poisson ratio 0.4, perfectly adhering to the sample; (4) The L-shaped block is bounded to a virtual roller, the center of which can only move along a line parallel to the wedge side, at a distance equal to the roller radius; (5) The crack edge is stress free; and (6) Vertical displacement is forbidden along the uncracked part of the symmetry plane (top line of Fig. 3:Top). For each sample, the equilibrium position of the wedge yielding the measured applied load at the onset of crack propagation was determined using a plane stress static finite element code (Cast3M 2007). Quasi-static crack propagation is then simulated by increasing the crack length while imposing the wedge position to remain constant. This latter assumption is experimentally justified. Indeed, it has been observed that the load cell signal started to be modified only after a few hundreds of microseconds (typical time of an experiment) after crack initiation[1]. This is attributed to the time required

[1] To measure this time shift, we directly connected the Wheatstone bridge of the load cell to an oscilloscope, without passing

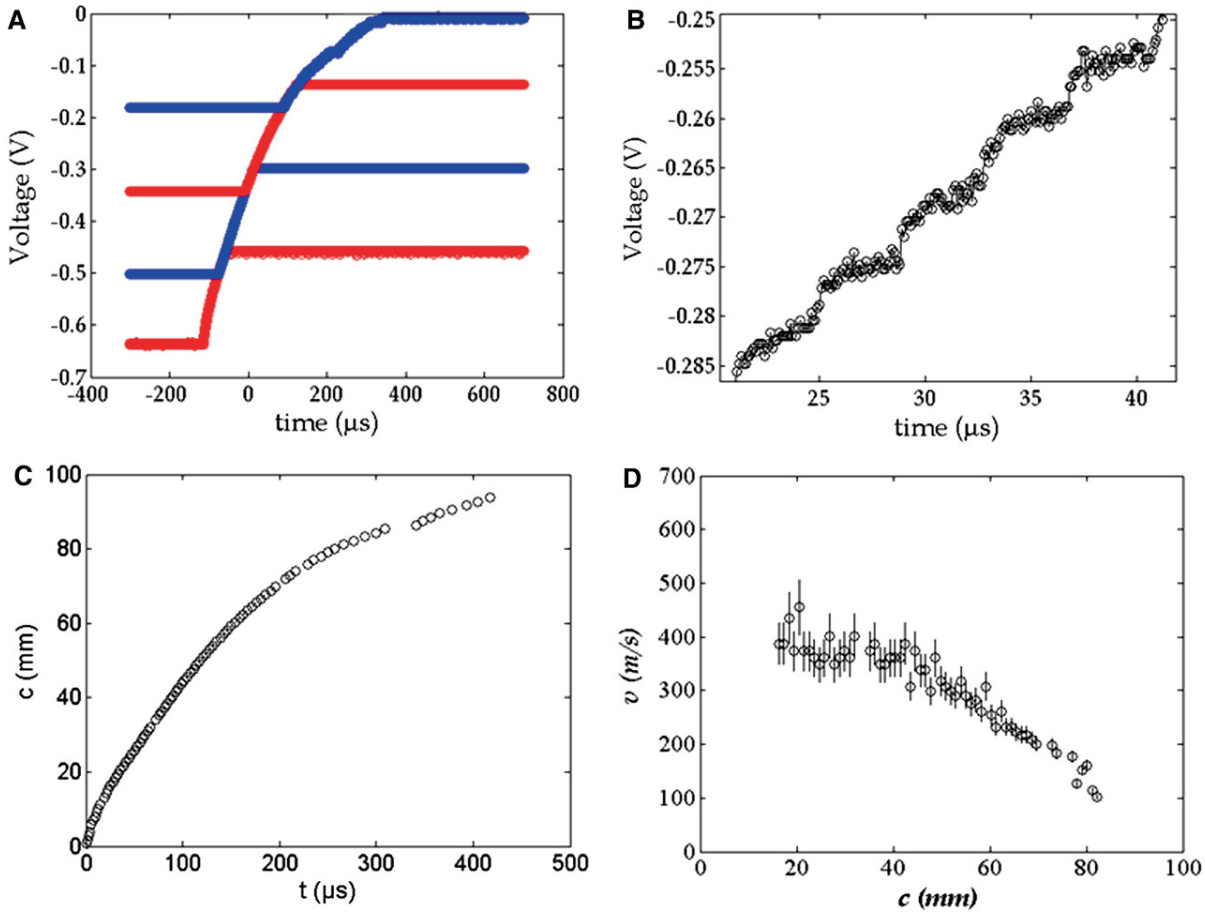

Fig. 2 Measurement of instantaneous crack velocity. **a** Voltage as a function of time as recorded via the oscilloscope. **b** Zoom that permits to distinguish the individual jumps yielded by the successive cuts of the conductive lines as crack advances. **c** Resulting position of crack tip as a function of time. **d** Final *curve* showing the variations of crack velocity with crack length

for the sound waves to travel from the crack tip to the load cell. As a result, the wedge position can be assumed to be practically constant during crack propagation. Static SIF is then determined, for each value c of crack length, using the J-integral method (Rice 1968). The validity of our SIF calculation was benchmarked against literature results for wedge-splitting configurations: Our code agreed to better than 2 % with the results of Karihaloo and Xiao (2001) when using their system parameters.

To image the post-mortem fracture surfaces, we used a Leica DM2500 microscope. Most of the images were taken with a 5× objective under polarized light. The resulting images are $1.4 \times 1\,\mathrm{mm}^2$ in area, with a pixel size of 677 nm. We also imaged the fracture surfaces with an interferometry profilometer (FOGALE Nanotech) which allows one to gather topographical information. These 3D images were taken with a 5× objective. Their area is $1.4 \times 1\,\mathrm{mm}^2$ with a pixel size of 1.86 μm.

3 Selection of fracture energy with velocity and evidence of a brittle/quasibrittle transition

In each of the experiments performed, both the SIF profile $K(c)$ and the instantaneous velocity profile $v(c)$ have been determined (Fig. 4:Left and center). From these curves, one can derive the profiles of fracture energy, $\Gamma(c)$, using Eq. 1 (Fig. 4:Right). We

Footnote 1 continued
through the signal conditioner. This latter, indeed, imposes a time resolution of 1 s.

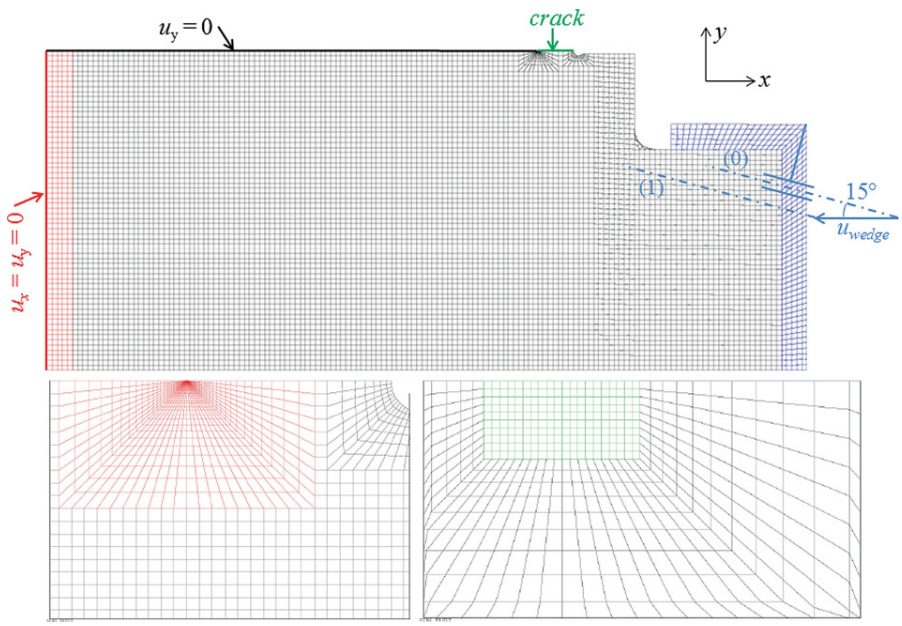

Fig. 3 Computation of SIF. *Top* typical mesh used for finite elements calculations, in order to access the stress/strain fields in the experiments. *Red* polymeric layer. *Black* sample. *Blue* L-shaped block. *Green line* cracked line. The slide link connected to the L-shaped block is used to model the motion of the contact point between the pushed wedge and the roller. Specimen loading is achieved by translating horizontally the slide link, from (0) to (1), over a distance u_{wedge} selected so that the horizontal force applying on the slide link is half that measured experimentally. *Bottom-Left* zoom on the meshing in the transition region (*red*) between coarse meshing (1 mm mesh size) in the bulk and fine meshing close to the crack tip. On the right is shown part of the circular hole at the seed crack tip. *Bottom-Right* zoom on the crack-tip region (*green*), meshed with a size of 1 nm

note here that Eq. 1 is strictly valid in the case of crack propagation in an elastic half-space, i.e. in the absence of waves emitted by the crack and coming back to it after reflection on sample boundaries (Freund 1990; Goldman et al. 2010). We argue that our experiments are essentially unaffected by such reflected wave for at least two reasons: (1) as measured by Boudet et al. (1995), the sound energy emitted by a crack in PMMA lies within the range 1–4 $\mathrm{J/m^2}$, a very small value compared to the energy required to fracture the material (a few $\mathrm{kJ/m^2}$, see below); and (2) this radiated energy is quickly dissipated within the material's bulk—The acoustic attenuation coefficient has been measured to be 0.67 dB/(cm MHz).

Figure 5 superimposes the various curves Γ versus v measured in the various experiments performed at various levels U_0 of the initially stored mechanical energy. Several regimes can be evidenced:

- For small v, Γ roughly remains constant, close to K_c^2/E, as expected within standard LEFM;
- As v increases and reaches a first critical velocity $v_a \approx 165\,\mathrm{m/s} \approx 0.19C_R$, Γ suddenly increases to a value about four times larger than $\Gamma(v=0) = K_c^2/E$;
- Beyond v_a, Γ slowly increases with v;
- As v reaches a second critical velocity $v_b \approx 350\,\mathrm{m/s} \approx 0.36C_R$, Γ starts to increase rapidly again. It seems to diverge at $v \approx 450\,\mathrm{m/s} \approx 0.5C_R$.

Note also the fairly good collapse of the different curves $\Gamma(v)$ below v_b, and the large dispersion above. This suggests that Eq. 1 is relevant for $v \leq v_b$,—provided a suitable velocity dependence $\Gamma(v)$ is ascribed—, but not beyond. The second critical value $v_b \approx 350\,\mathrm{m/s} \approx 0.36C_R$ is found to correspond to the onset of microbranching instability widely discussed in the literature (see e.g. Fineberg et al. 1992; Sharon and Fineberg 1999). The first critical value $v_a \approx 165\,\mathrm{m/s} \approx 0.19C_R$ was observed in our series of experiments and reported in (Scheibert et al. 2010) for the first time. This observation was made possible by the use of the wedge-splitting geometry, which is a decelerating crack configuration. It offers many data points at relatively low velocities, contrary to most other devices used before. The rapid increase of Γ

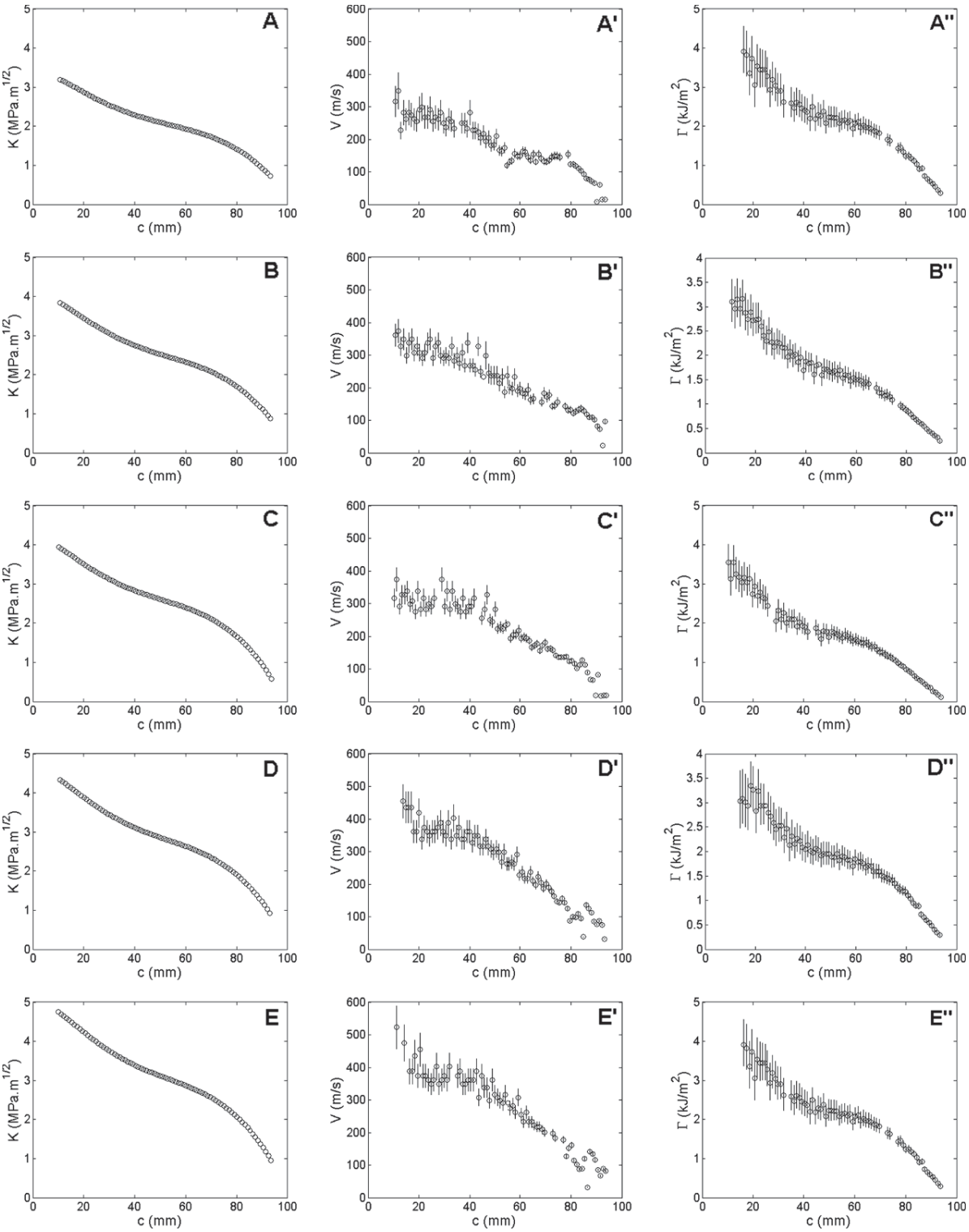

Fig. 4 Variation of SIF K (*left*), instantaneous crack velocity v (*center*) and fracture energy Γ (*right*) as a function of crack length c for five experiments with different stored mechanical energy U_0 at propagation onset: 2.0 J (A, A′ and A″), 2.6 J (B, B′ and B″), 2.9 J (C, C′ and C″), 3.8 J (D, D′ and D″), and 4.2 J (E, E′ and E″)

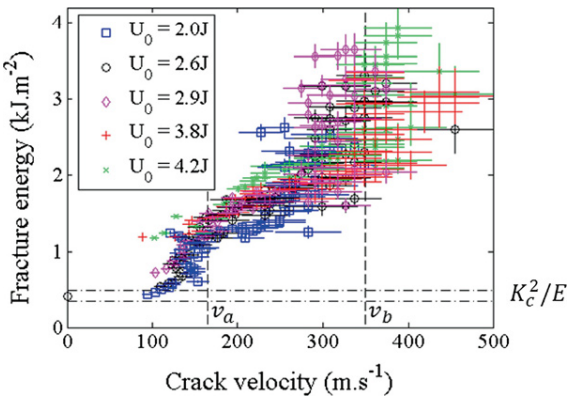

Fig. 5 Fracture energy Γ as a function of crack velocity v for five different experiments with different stored mechanical energies U_0 at crack initiation. The *two vertical dashed lines* correspond to v_a and v_b. The *two horizontal dashed lines* indicate the confidence interval (95 %) for the measured fracture energy K_c^2/E at crack initiation (taken from Scheibert et al. (2010))

with v around v_a provides a direct interpretation for the repeated observations of cracks that span a large range of Γ, while keeping a nearly constant velocity about $0.2C_R$ (see e.g. Ravi-Chandar and Knauss 1984; Ravi-Chandar and Yang 1997).

To shed light on the nature of the transition at $v = v_a$ evidenced on the curve relating Γ to v, we examined the post-mortem fracture surfaces and their evolution as crack speed increases (Fig. 6). For v smaller than v_a, fracture surfaces remain smooth at optical scales. But above this threshold, conic marks start to be observed. These marks remain confined within a roughness scale of few micrometers at the fracture surfaces, contrary to the microbranches that deeply develop into the specimen bulk above v_b (see side views and topographical images in Fig. 7).

Similar conic marks were reported in the fracture of other brittle materials, among which polystyrene (Regel 1951), oxide glasses (Smekal 1953; Holloway 1968), cellulose acetate (Kies et al. 1950), polycrystalline materials (Irwin and Kies 1952), or homalite (Ravi-Chandar and Yang 1997). They are classically associated (Smekal 1953; Ravi-Chandar and Yang 1997; Rabinovitch et al 2000) with the presence of microcracks that nucleate and grow ahead of the main crack front, and subsequently coalesce with it.

Figure 8 shows the surface density ρ of conic marks as a function of crack speed v. Below v_a, no mark is observed, irrespectively of the chosen magnification (from 5× to 50×). Above this value, ρ increases almost linearly with $v - v_a$. The precise correspondence between the critical velocity v_a at which the curve $\Gamma(v)$ exhibits a kink and that at which conic marks start to be observed suggests that both phenomena are the signature of the same transition. Above v_a, PMMA failure switches from nominally brittle to quasi-brittle and goes hand in hand with microcracking that develop ahead of the main front. In the following, we will use the term "Damage Zone" (DZ) to refer to the zone where microcracks develop in the vicinity of the main crack tip. We will distinguish this zone from the FPZ (smaller than the DZ) that embeds dissipative mechanisms (crazing for instance) at the tip of each (micro)crack front.

Note that the region of the fracture surface that bears conic markings also has a characteristic aspect to the naked eye, as it scatters ligth very efficiently. This is markedly different from the high reflectivity of the region that bears no microcrack. In fractography these regions are classically referred to as the mist and mirror zones, respectively (see e.g. Hull 1999; Rabinovitch

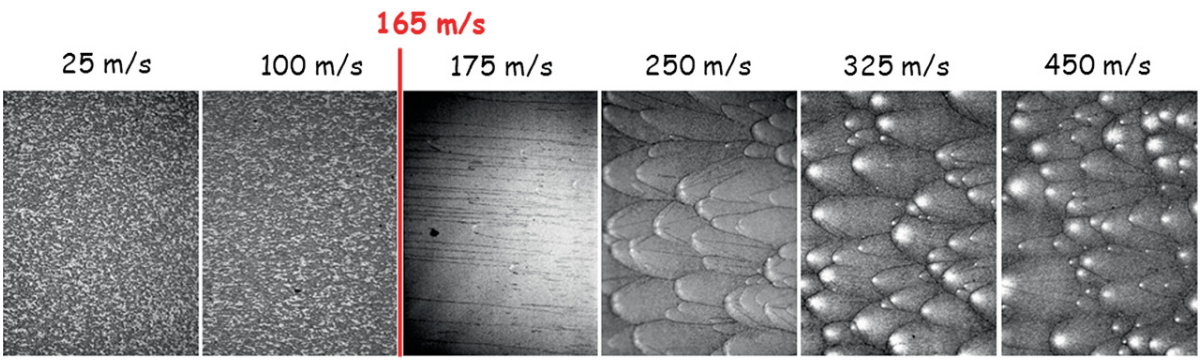

Fig. 6 $1 \times 1.4\text{mm}^2$ microscope images of fracture surfaces created at various speed v

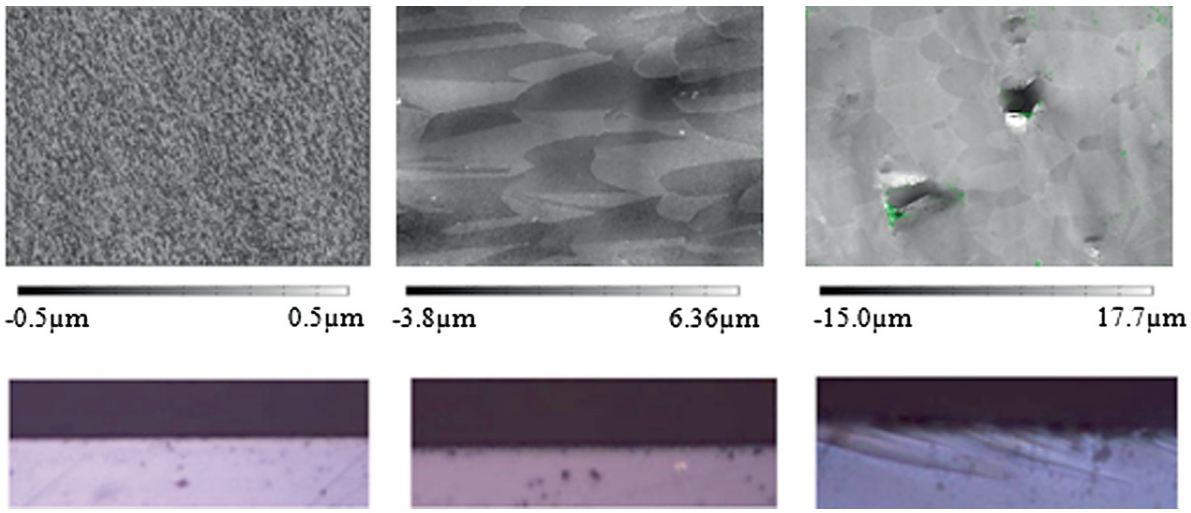

Fig. 7 Profilometric $1.4 \times 1\,\text{mm}^2$ surface (*Top*) and optical $0.7 \times 0.25\,\text{mm}^2$ side-view (*Bottom*) images taken at velocity smaller than v_a (*left*), between v_a and v_b (*center*), and above v_b (*right*). Crack has propagated from *left* to *right*

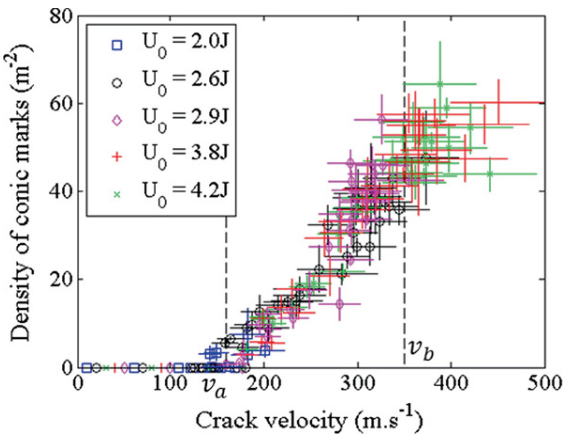

Fig. 8 Surface density ρ of conic marks as a function of crack velocity for five different experiments with different stored mechanical energies U_0 at crack initiation (taken from Scheibert et al. (2010))

and Bahat 2008). Our observation therefore strongly suggests that the mirror/mist transition is simply the morphological counterpart of the brittle/quasi-brittle transition that occurs at v_a

4 Fractographic reconstruction of individual microfailure events

In order to further characterize and understand the dynamics of microcracking events which develop during fast fracture in PMMA ($v > v_a$), we proposed a numerical reconstruction based on the post-mortem analysis of the traces left by these events (Guerra et al. 2012). The typical time between two successive microcracking events is about 10 ns, which makes them inaccessible using standard techniques as e.g. fast imaging or acoustic emission analysis.

The first step is to identify where microcracks have initiated. On the optical images of post-mortem fracture surfaces (Fig. 9a, b), many well defined points of high optical reflectivity can be seen. These bright/white areas are believed to come from large plastic deformations accompanying the nucleation of microcracks. Thus, they allow us to determine precisely the position of the different nucleation sites. For some microcracks, the precise location of their nucleation site is further constrained by the convergence of fragmentation lines onto it.

The second step is to infer, from the conic patterns, the velocities at which the various microcracks have grown. It is commonly admitted in the literature (Smekal 1953; Ravi-Chandar and Yang 1997; Rabinovitch et al 2000) that the conical marks indicate the intersection points between two interacting microcracks. Let us then consider two microcracks that propagate radially along slightly different planes with velocities c_1 and c_2 (Fig. 9b). The intersection between the initial (micro)crack front and the secondary microcrack leaves a visible trace (i.e. a tiny height difference) on the fracture surfaces. In the coordinate system

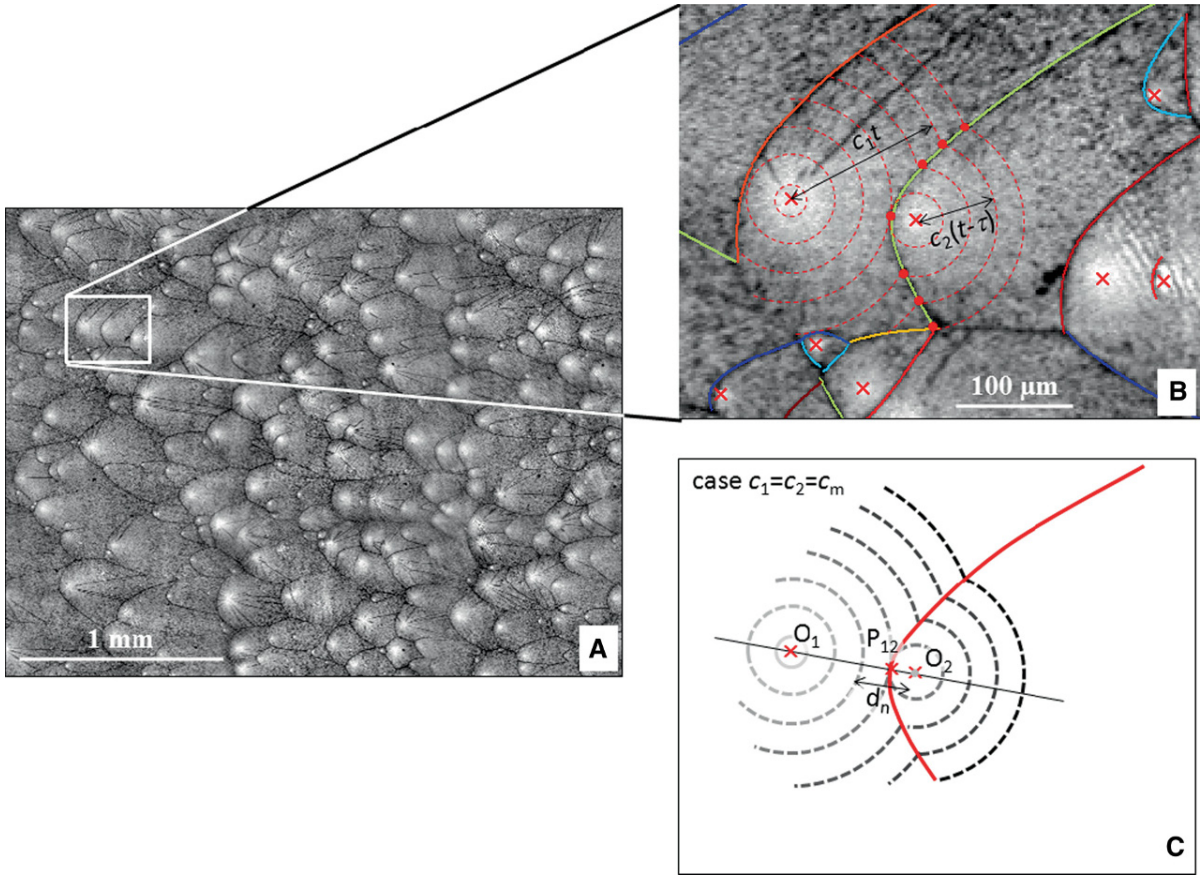

Fig. 9 **a** Typical microscope image of the post-mortem fracture surface. *Bright regions* indicate the nucleation centers. **b** Interpretation of the conic marks: *red dots* sketch the successive positions of the fronts that, by interacting, will be giving rise to the *green branch* of the conic mark. These two microcracks form at $t = 0$ and $t = \tau$, then radially grow at speeds c_1 and c_2. Branch fitting via Eq. 2 permits to measure c_2/c_1. Note that a conic mark can be made of several branches. **c** When $c_2 = c_1$, branches are (mathematical) true conics and the nucleation distance d_n between the front of the incident microcrack (1) and the nucleation point of the forming one (2) is twice the apex-focus distance 0_2P_{12} (taken from Guerra et al. (2012))

$(\mathbf{e}_x, \mathbf{e}_y)$ centered on the site of nucleation of the first microcrack and chosen so that the x-axis passes through the two centers of nucleation, the equation describing the mark left postmortem can be written as (Guerra et al. 2012):

$$\frac{y}{\Delta} = \pm \left[\frac{2c^2}{(c^2-1)^2} \frac{c_1\tau}{\Delta} \sqrt{\left(1-\frac{2x}{\Delta}\right)(c^2-1)+c^2\left(\frac{c_1\tau}{\Delta}\right)^2} - \left(\frac{x}{\Delta}\right)^2 + \frac{1}{(c^2-1)^2} + \frac{c^2+1}{(c^2-1)^2}\left(c^2\left(\frac{c_1\tau}{\Delta}\right)^2 - \frac{2x}{\Delta}\right)\right]^{1/2}, \qquad (2)$$

where $c = c_2/c_1$, Δ is the distance between the two nucleation sites, and τ is the time interval between two microcracking events. The analysis of this equation shows that the number $c_1\tau/\Delta$ sets the aspect ratio of the fractographic trace, while the ratio $c = c_2/c_1$ fixes its form (see Guerra et al. 2012 for details).

We used Eq. 2 to directly extract the ratio $c = c_2/c_1$ for each couple of interacting microcracks ($\sim$ 400 couples per image) on different optical pictures obtained at different speeds (Fig. 10). Regardless of the macroscopic crack velocity, we obtain for c a Gaussian distribution centered on 1 with a standard deviation of about 0.03 (Fig. 10). This analysis permits to demonstrate that all microcracks grow at the same velocity c_m inside the DZ. Note that, up to now, nothing prevents c_m to vary with v or K. The mechanism that fixes the value for c_m will be discussed later in this paper (see Sect. 6).

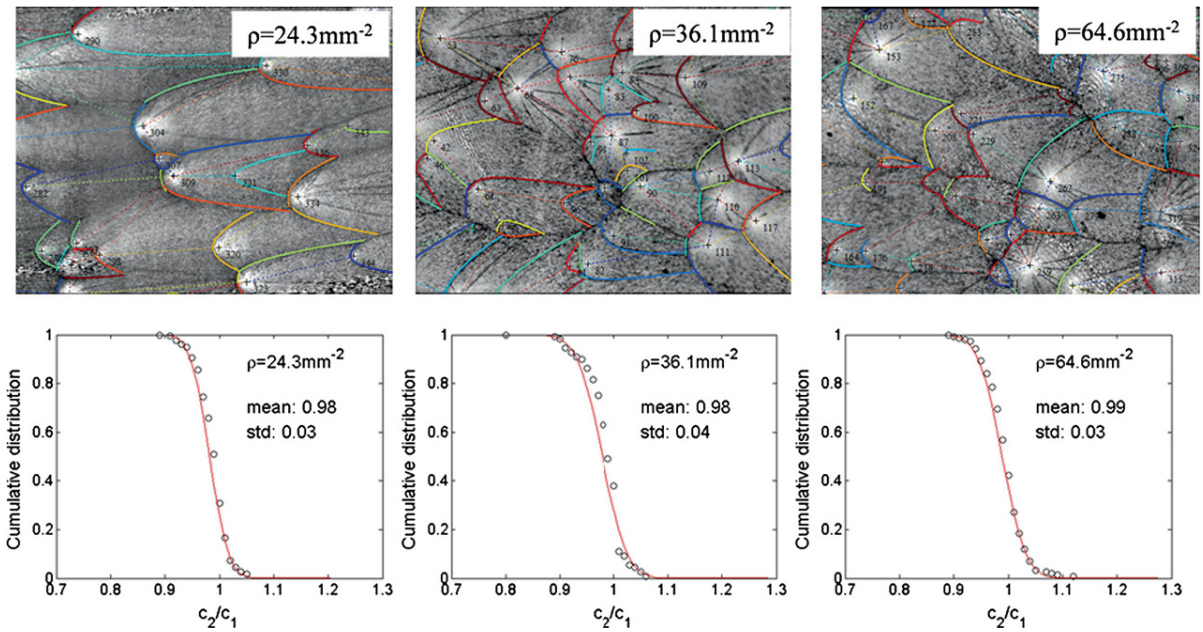

Fig. 10 Direct extraction of the relative speed between two interacting microcracks at three different microcrack densities. (*Top*) Zones of investigation. Each conic branch has been attributed a given *color* and the nuclei of the two corresponding interacting microcracks has been joined by a *dotted segment* of the *same color*. Note that a conic-like mark is often made of several of these conic branches. The ratio c_2/c_1 is the only adjustable parameter in Eq. 2 to determine the branch geometry once the nuclei position and the branch apex are set. (*Bottom*) Corresponding distributions for c_2/c_1: in the three cases, the distributions are found to fit normal distributions of mean value 0.98–0.99 and standard deviation 0.03–0.04, irrespective of ρ (taken from the supporting information of Guerra et al. (2012))

By using $c_2 = c_1 = c_m$ in Eq. 2, the following relation can be obtained:

$$\frac{y}{\Delta} = \pm\left[4\frac{d_n(2\Delta-d_n)}{(\Delta-2d_n)^2}\left(\frac{x}{\Delta}\right)^2 - 4\frac{d_n(2\Delta-d_n)}{(\Delta-2d_n)^2}\left(\frac{x}{\Delta}\right) + \frac{d_n(2\Delta-d_n)}{(\Delta-2d_n)^2} - 4\frac{d_n}{\Delta} + 4\left(\frac{d_n}{\Delta}\right)^2\right]^{1/2}, \quad (3)$$

where $d_n = (\Delta - c_m\tau)$ represents the critical distance between the front of the incident microcrack and the center of the nucleated microcrack at the instant of its nucleation. This distance is twice the distance between the apex and the focus of the considered microcrack and, as such, d_n can be directly extracted from fractographic images (Fig. 9c).

We are now in a position to reconstruct the full dynamics of microcracking events from the knowledge of two series of parameters that can be directly extracted from fractography: The position of nucleation sites and the critical nucleation distance d_n associated to each conic mark. We analyzed several areas of the fracture surfaces, corresponding to various macroscopic velocities above the microcracking onset ($v > v_a$). In all cases, nine optical images with partial overlap were recorded and gathered into a single large image—This ensures adequate statistics in the following analyses. The methodology to extract the data is the following:

- We record the coordinates x and y of the nucleation sites;
- We estimate the nucleation distance, d_n, by inferring, for each nucleation site, who was the parent microcrack from the relative position of the conics apex with respect to the nucleation site (family criterion). Then, d_n is twice the apex-nucleation center distance.

Following Ravi-Chandar and Yang (1997), the reconstruction is initiated with an originally straight crack front positioned at the left of the zone of analysis. This front propagates from left to right, at a constant velocity of 1 pixel/timestep. Once the distance between one of the microcrack nucleation sites and the primary crack front reaches d_n, a secondary microcrack

starts to grow radially at the same velocity. The front is now the combination of both the translating straight front and the radially growing circular front. Beyond a given time, these two fronts intersect and define the position of the conic mark. When the growing front reaches again a distance d_n associated with another site, a new microcrack is nucleated. The procedure is repeated until all the nucleation sites have been triggered.

Figure 11 shows two examples of reconstruction for two different macroscopic velocities and Fig. 12 superimposes the conic marks obtained by the simulation to those really observed in fractography. We note the fairly good agreement between experiment and reconstruction away from the edges (in the center of the images). This permits to validate the method. Thanks to it, we are able to reconstruct deterministically the dynamics of micro-damage during fast fracture in PMMA from the observation and analysis of the patterns observed on the post-mortem fracture surfaces. We emphasize the resulting resolution of $\sim 1\,\mu$m (1 pixel) in space, and of 10 ns in time (pixel size/c_m with $c_m \approx 200$ m/s as it will be demonstrated further in the paper). This is far better than what can be obtained via conventional experimental mechanics methods as e.g. acoustic emission or fast imaging methods.

Note however some discrepancies between experiments and reconstructions on the left, top and bottom edges of the analyzed zones. Such finite size problems are unavoidable. Along the top and bottom sides of the image, the influence of microcracks outside the field of view are naturally ignored. Also, a non-realistic straight vertical front has been used, as we cannot predict the precise instants at which the leftmost centers have turned to microcracks during the reconstruction. In order to test the sensitivity of the reconstruction to the initial front shape, we have performed different simulations with the same inputs except the initial front shape. In Fig. 13, three different cases are tested : (1) a straight vertical front, (2) a sinusoidal front with an amplitude and a period equivalent to the mean value of a wavy front far from the first step in previous reconstruction and (3) the same sinusoidal front vertically shifted by half a period. We can observe that, except of course at the very beginning of the simulation (i.e. on left side of the images), the difference between these three cases are very small. Moreover, these small differences tend to disappear as the front to propagate, i.e. as more and more microcracks are involved. In some cases, we have observed differences propagating over the whole image. Perfect reconstructions would be reached only by analyzing the whole fracture surfaces, and not some partial areas. Knowing all the history of crack propagation is the only

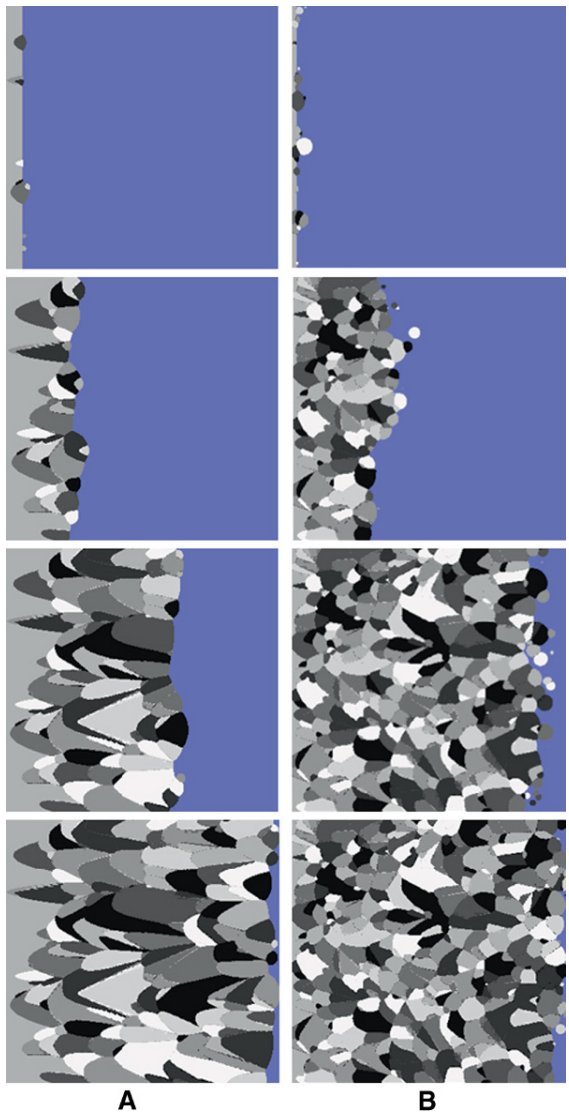

Fig. 11 Reconstructed sequence of crack propagation and microfailure events at the microscale for two different values of K, namely $K = 2.53$ MPa $\sqrt{\text{m}}$ (*left*) and $K = 4.18$ MPa $\sqrt{\text{m}}$ (*right*). The *blue part* corresponds to uncracked material and we ascribed an arbitrary gray level to each of the growing microcrack. From this, one can deduce the intersection points between microcracks that coincide with the observed conics marks on the fracture surfaces. The size of the two zones of analysis is $2.5 \times 2.5\,\text{mm}^2$. The macroscopic crack propagates from *left* to *right*

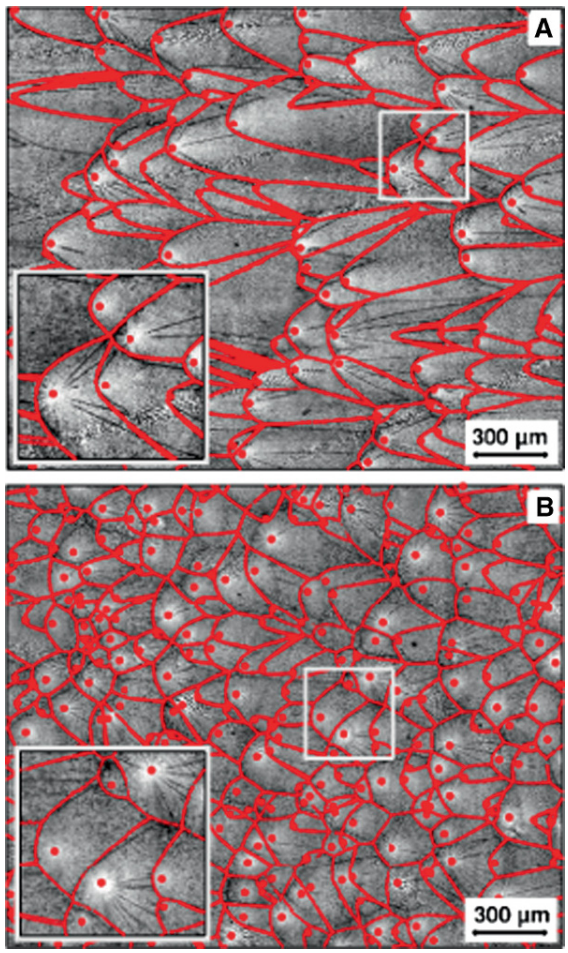

Fig. 12 Qualification of the procedure: superimposition of the conic pattern as obtained from reconstruction onto that observed by fractography. **a** $K = 2.77\,\text{MPa}\sqrt{\text{m}}$. **b** $K = 4.18\,\text{MPa}\sqrt{\text{m}}$. *Red dots* indicate nucleation centers (taken from the supporting information of Guerra et al. (2012))

solution for the determination of the real initial front shape which could allow a perfect reconstruction.

5 Quantitative selection of the microdamaging state by crack loading

The reconstruction method, described in the previous section, shows that the dynamics of micro-damage is entirely determined by the position of nucleation sites, the nucleation distances d_n, and the microscopic velocity of the fronts c_m. To understand what selects these three parameters, they will be explored in more detail.

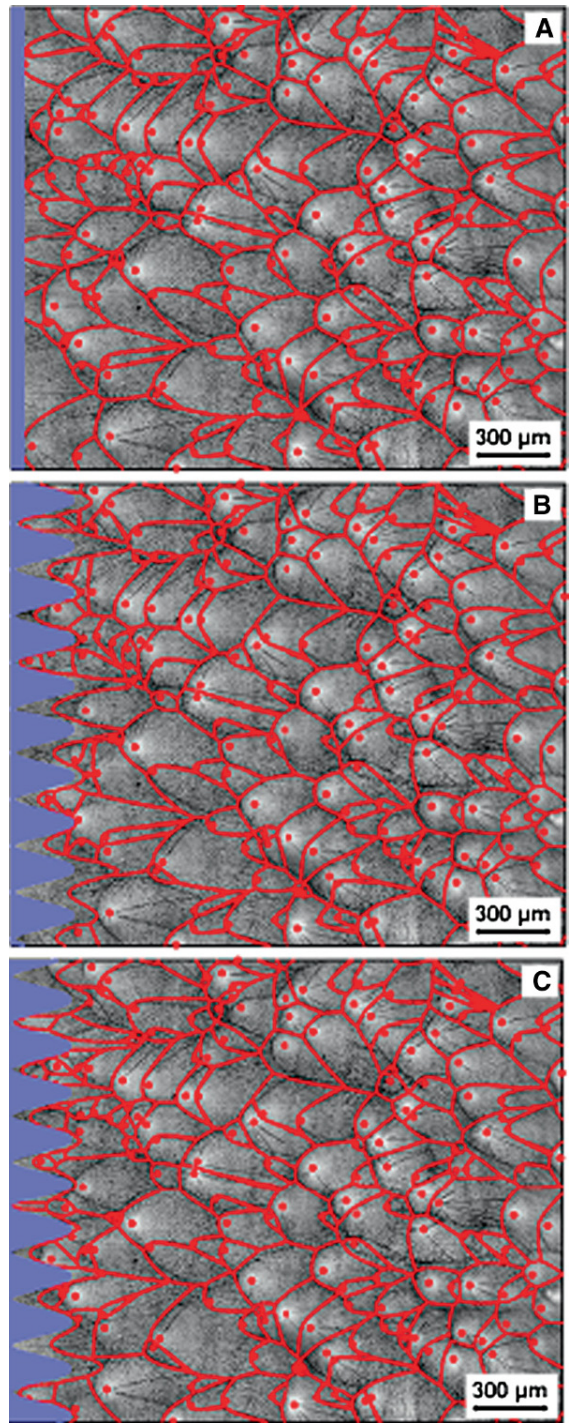

Fig. 13 Influence of initial crack front morphology on the reconstruction for $\rho = 45.0\,\text{mm}^{-2}$ ($K = 3.65\,\text{MPa}\sqrt{m}$). Three different initial conditions were used. **a** *Straight vertical line*. **b** *Vertical sinusoidal shape* with a period of 186 µm and a peak-to-peak amplitude of 242 µm. **c** *Same sinusoidal shape*, but translated vertically over half a period (taken from the supplementary information of Guerra et al. (2012))

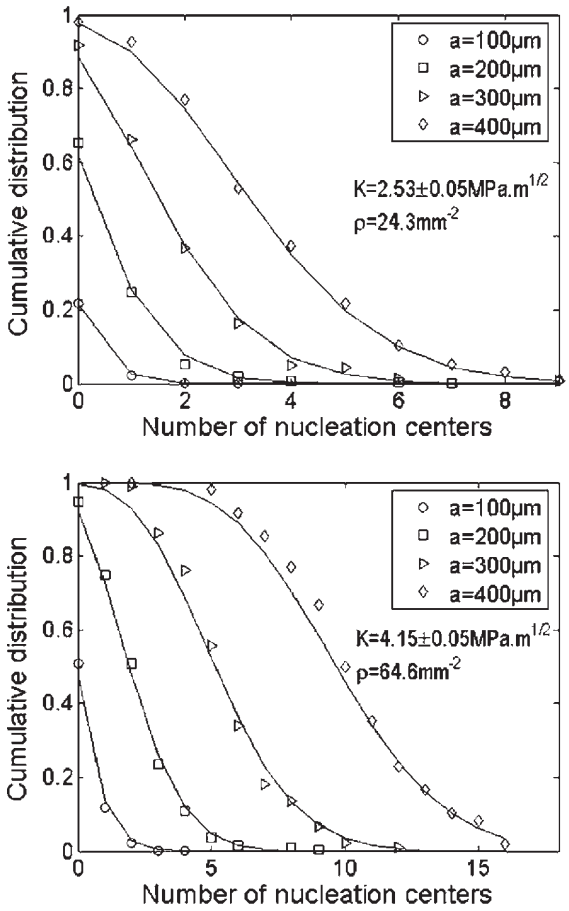

Fig. 14 Cumulative distribution of the number of nucleation centers contained in square regions of lateral size a, for two different fractographic images. *Solid lines* represent Poisson fits of parameter ρa^2, where the fitted value ρ can be identified with the center density

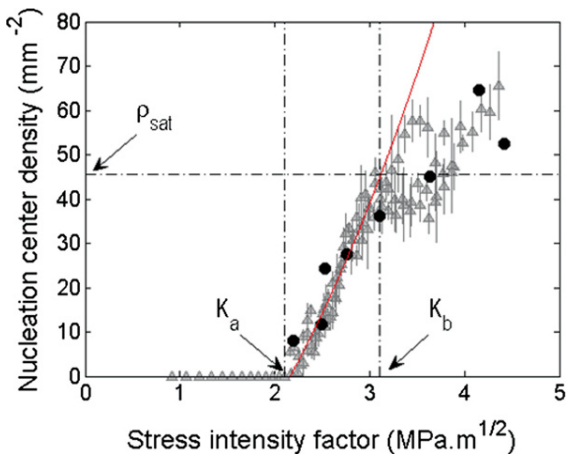

Fig. 15 Variation of the density ρ of nucleation centers with applied SIF. *Black disks* correspond to the eight images on which post-mortem reconstruction was performed. *Red line* is a fit via Eq. 4. K_a and K_b are associated with microcracking and microbranching onset, respectively (taken from Guerra et al. (2012))

5.1 Spatial distribution of nucleation centers

To characterize the spatial distribution of nucleation sites, we computed the distribution of the number of sites that can be observed on a square of a given size a. Figure 14 shows these distributions for two different fractographic pictures, i.e. for two values of K. These distributions follow a Poisson distribution with a single parameter ρa^2 where ρ is constant over a given fractographic picture but varies with K. This reflects a homogeneous and uncorrelated distribution of the sites all along the post-mortem fracture surface. Then, ρ can be identified with the surface density of nucleation sites and fully characterizes their spatial distribution.

Figure 15 presents the evolution of ρ with loading K. The density is naturally zero for K smaller than K_a (SIF value at microcracking onset, *i.e.* when $v = v_a$) as there is no microcracking. Beyond this value, it increases with K. Also, this curve exhibits large fluctuations and seems to saturate when the density reaches a value $\rho_{sat} \approx 45.5\,\mathrm{mm}^2$, beyond K_b. Note that this latter value coincide with microbranching onset and, therefore the $\rho(K)$ curve does not really make sense above K_b since the front has splitted into various microbranches, which prevents us from correlating the macroscopic K value to the number of conic marks on the main crack only.

To account for the curve $\rho(K)$, let us consider (Scheibert et al. 2010) that a discrete population of weak localized zones is present within the material with a bulk density ρ_v. Let us then assume that a microcrack can nucleate in these areas provided the two following conditions are fulfilled:

- The local stress at the considered zone is large enough, *i.e.* larger than a given threshold value σ^* (smaller than the yield limit σ_Y of the material);
- The considered zone is far enough from the main crack front to allow the microcrack to grow, i.e. at a distance greater than d_a.

The density ρ of nucleated microcracks per unit area of fracture is then given by $\rho_v\{h_\perp - d_a\}$ where $h_\perp$ is the thickness of the layer where the stress has exceeded σ^*($h_\perp$ is measured perpendicular to the mean plane of fracture). The singular form of the stress field at crack tip leads to write $h_\perp \propto K^2/\sigma^{*2}$, which yields

(Scheibert et al. 2010):

$$\begin{array}{ll} \rho = 0 & \text{for} \quad K \leq K_a \\ \rho \propto K^2 - K_a^2 & \text{for} \quad K \geq K_a, \end{array} \tag{4}$$

with $K_a \propto \sigma^* \sqrt{d_a}$. This equation, plotted in red on Fig. 15, reproduced quite well the experimental data below K_b.

As the determination of ρ from fracture surface is much more accurate and consensual than the determination of K (based on FEM calculation with inherent approximations), in the following, we will use ρ rather than K as the control parameter as a function of which the various quantities further defined in this paper will be plotted. Indeed, ρ fully characterizes the spatial distribution of nucleation sites and monotonically evolves with K within the relevant range $K_a \leq K \leq K_b$.

5.2 Chronology of microfailure events

Figure 16 shows the cumulative distribution of nucleation distances d_n measured on two typical fracture surfaces. The probability density function (derived from the cumulative distribution) is zero for $d_n \leq d_{\min}$, maximum at $d_{\min}$, subsequently linearly decreases, and is finally zero as d_n exceeds $d_{\max}$. This distribution form was observed for all the fracture surfaces, irrespectively of ρ. $d_{\max}$ is found to increase with ρ, while $d_{\min}$ decreases with ρ and remains very close to zero (Guerra et al. 2012).

The variation of the mean value $\bar{d_n}$ of d_n with ρ exhibits two regimes (Fig. 17:Top): (1) A linear increase with ρ followed by (2) a saturation ($\bar{d_n}^{sat} \approx 50\,\mu\text{m}$) when ρ is greater than the value ρ_{sat} associated in Fig. 15 with the microbranching onset (at $K = K_b$).

To understand the saturation origin, we plotted, on Fig. 17:Bottom, the evolution of the mean nearest-neighbor distance $\langle \Delta r \rangle$ with ρ as it is predicted for a 2D Poissonian spatial distribution ($\langle \Delta r \rangle = 1/(2\sqrt{\rho})$). The plotted errorbars also indicate the associated standard deviation ($\sigma_{\Delta r} = \sqrt{(4-\pi)/(4\pi\rho)}$). When ρ reaches ρ_{sat}, in some cases, the distance between two neighbouring nucleation centers is of the order of the nucleation distance. As a result, two centers can nucleate almost at the same time (avalanche effect). The number of microcracks involved in such avalanches, also plotted in the figure, increases in the vicinity of ρ_{sat}. This strongly suggests that the observed saturation in the d_n evolution (Fig. 17:Top) results from this steric effect.

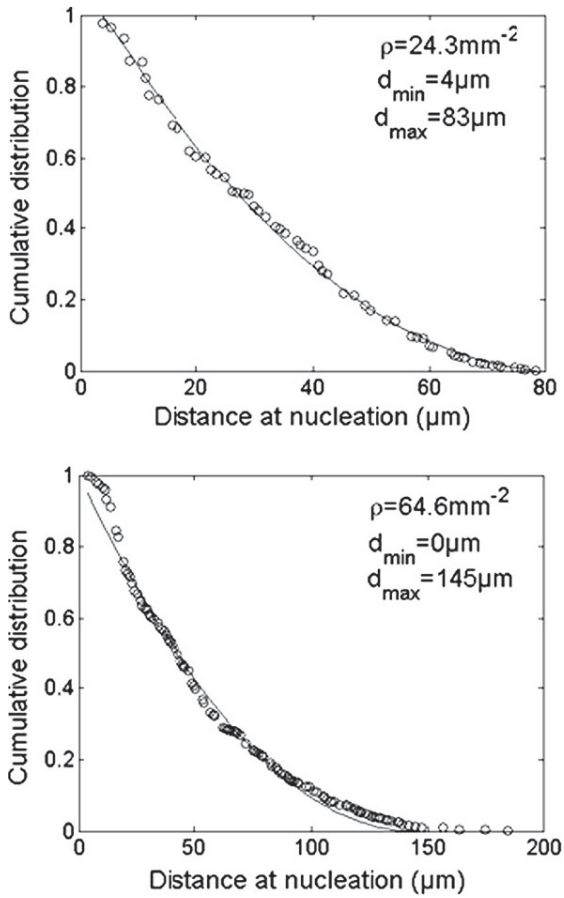

Fig. 16 Cumulative distribution of the nucleation distances d_n, as measured for two different fractographic images. *Solid lines* are fits with $P(d_n) = ((d_{\max} - d_n)/(d_{\max} - d_{\min}))^2$

5.3 Velocity of microcrack growth

The fractographic analysis performed in Sect. 4 showed that, within the DZ, all microcracks grow at the same velocity c_m. Unfortunately, it does not allow us to directly measure the value of c_m and its possible dependence with ρ. Indeed, Eq. 2 shows that the form taken by the conical brands depends only on the position of the nucleation sites and on the nucleation distance but not on the value of c_m.

We propose to use the deterministic reconstructions obtained in Sect. 4 to connect the macroscopic crack speed v to c_m. Figure 18a shows the evolution of the mean front position measured from reconstructions of different fracture surfaces (i.e. for different densities of nucleation centers) as a function of numerical time step normalized by c_m (homogeneous to a distance,

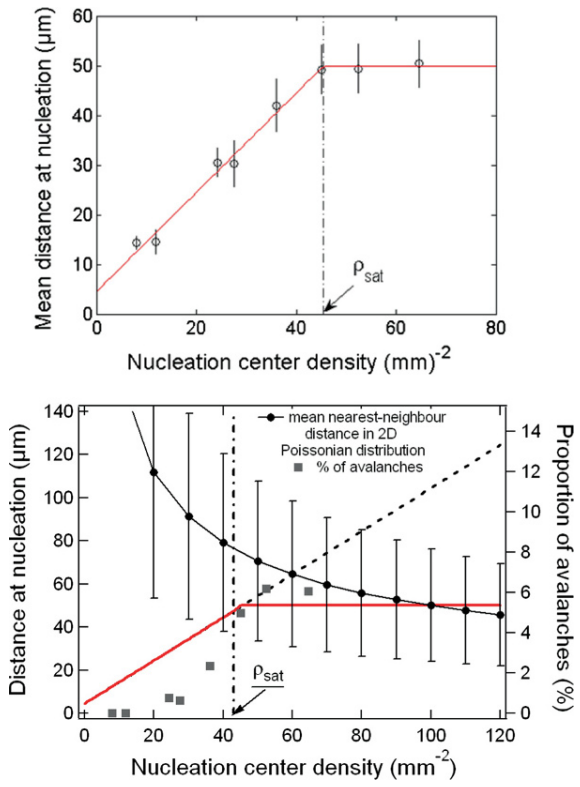

Fig. 17 *Top* mean distance at nucleation $\bar{d}_n$ as a function of ρ. The red curve is a linear fit over $\rho \leq \rho_{sat} = 45.5\,\mathrm{mm}^{-2}$ and a plateau at $\bar{d}_n^{\,sat} = 50\,\mu m$ over ρ_{sat}. *Bottom* saturation of $\bar{d}_n$ and avalanches. The *red line* is the same as in the *top* figure. It is compared to the mean nearest-neighbor distance in a Poissonian distribution (*black solid line*). *Solid squares* indicate the proportion of microcracks involved in avalanches, as computed from the reconstruction

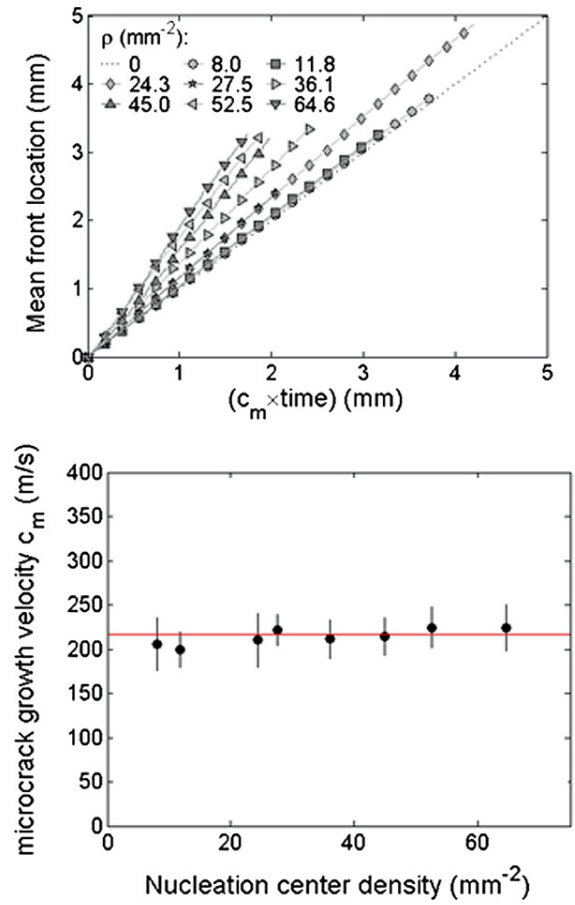

Fig. 18 Determination of the local propagation velocity c_m of individual microcracks. *Left* evolution of the mean crack front position as a function of $c_m \times t$. Different curves correspond to different fractographic images with different microcrack density. Slope of these curves defines the ratio A between continuum-level scale failure velocity v and microcrack velocity c_m. *Right* variation of c_m with ρ. *Horizontal red line* indicates the mean value $c_m \approx 217$m/s

corresponding to the distance travelled by a crack front at a velocity c_m in absence of micracking events). All the curves are linear with a slope A that defines the ratio v/c_m. At this point, one should recall that the reconstruction procedure starts with an unrealistic initial straight crack front. We checked that this does not affect the measured value A (Guerra et al. 2012). Then, since v is known, c_m can be deduced and plotted, on Fig. 18b as a function of ρ. This microscopic velocity is found to be a constant, $c_m \approx 217\,\mathrm{m/s} = 0.24 C_R$, independent of the loading K, the density ρ, and the macroscale velocity v. The implications of this result will be further discussed in the next section.

The fracture energy measurements $\Gamma(v)$ made in Sect. 3 will now allow us to understand why such a constant velocity c_m is selected within the microcracking regime. Figure 19 shows the evolution of Γ with the *dynamic* SIF K_d. In the low speed regime (i.e. for low K_d) where crack propagates without involving any microdamaging or microbranching, Γ scales with K_d^2. We can also expect that the FPZ size R_c scales with K_d^2: $R_c = K_d^2/\alpha\sigma_Y^2$ where α is a constant of order 1, and σ_Y is the material's yield stress. The scaling of Γ with K_d thus indicates a linear variation between Γ and R_c and, following Scheibert et al. (2010), a constant dissipated energy ϵ per volume unit within the FPZ. The fracture energy then writes:

$$\Gamma = \frac{\epsilon}{\alpha\sigma_Y^2} K_d^2 + \left(\frac{1}{E} - \frac{\epsilon}{\alpha\sigma_Y^2}\right) K_c^2 \tag{5}$$

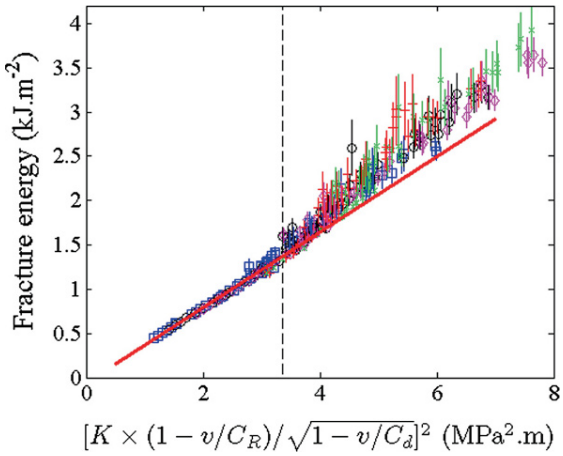

Fig. 19 Variation of the fracture energy Γ with $K_d = K \times (1 - v/C_R)/\sqrt{1 - v/C_D}$. The *vertical dashed line* corresponds to microcracking onset $K_d^a = 1.83\,\text{MPa}\sqrt{\text{m}}$. Note that K_d can *stricto-senso* be identified with the dynamic SIF in absence of microcracking, *i.e.* for $K_d \leq K_d^a$. The *solid red line* is a fit with Eq. 5

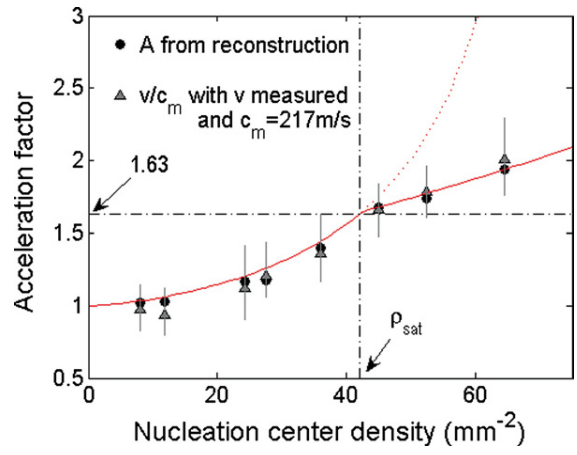

Fig. 20 Acceleration factor $A = v/c_m$ as a function of the microcrack density ρ. *Red line* is a fit by Eq. 9 with $b = 1.19 \pm 0.02$. *Horizontal dot line* defines v_b/c_m and passes through the slope breakdown observed when $\rho = \rho_{sat}$ (taken from Guerra et al. (2012))

As long as $v \leq v_a$ (thus without microdamage), one can relate K_d to K (Freund 1990): $K_d = K \times (1 - v/C_R)/\sqrt{1 - v/C_D}$ where C_D refers to the speed of dilational waves (cf. Table 1). By replacing K with K_d in Eq. 1, and by subsequently injecting the result in Eq. 5, one gets the following expression for Γ versus v:

$$\Gamma = \frac{1 - \frac{\epsilon E}{\alpha \sigma_Y^2}}{1 - \frac{\epsilon E}{\alpha \sigma_Y^2} \frac{1 - v/C_R}{1 - v/C_D}} \frac{K_c^2}{E} \tag{6}$$

This expression yields a divergence of Γ at a finite velocity c_∞ given by:

$$c_\infty = \left(\frac{\epsilon E}{\alpha \sigma_Y^2} - 1 \right) \frac{C_R C_D}{\frac{\epsilon E}{\alpha \sigma_Y^2} C_D - C_R} \tag{7}$$

In this expression, the only unknown quantity is $\epsilon/\alpha\sigma_Y^2$. This can be evaluated by fitting the first linear part (i.e for $K_d \leq K_d^a$) of the curve Γ versus K_d^2/E plotted in Fig. 19 with Eq. 5. One then gets:

$$c_\infty \simeq 204\,\text{m/s} = 0.23 C_R \tag{8}$$

This value sets the maximum crack growth velocity in the absence of microcracking: Beyond c_∞ dissipation diverges within the FPZ. Assuming that the same dissipation mechanisms are involved within the FPZ of each microcrack when $v > v_a$, we expect that the limiting speed of microcracks is also determined by c_∞, so that $c_m \approx c_\infty$.

6 Role of microdamaging in the selection of continuum-level scale failure velocity

The previous section has allowed us to better understand how the different variables that characterize the microdamage developing within the DZ are selected when v exceeds v_a. The most surprising observation concerns the propagation velocity of microcracks c_m that remains constant, of the order of 217 m/s, and significantly lower than the macroscopic velocity v. In other terms, damage does not slow down the macroscopic crack tip as it was commonly believed until now (Ravi-Chandar 2004; Ravi-Chandar and Yang 1997; Washabaugh and Knauss 1994), but on the contrary it accelerates it. And this acceleration factor is all the more important as ρ (or equivalently K) increases (Fig. 20).

The temporal evolution of a given point on the main front, shown in Fig. 21, permits to better understand the effect of microdamage on the dynamics of macroscopic cracking. The trajectory of this point exhibits random and sudden jumps. These jumps actually correspond to the coalescence of a microcrack with the main front. Between these events, the local front velocity remains close to c_m (*i.e.* the slope is equal to 1). Such an acceleration effect can be captured by a simple mean-field model, detailed in Guerra et al. (2012), which simply consists in counting the number of coalescence events as the main front propagates over a unit of length when

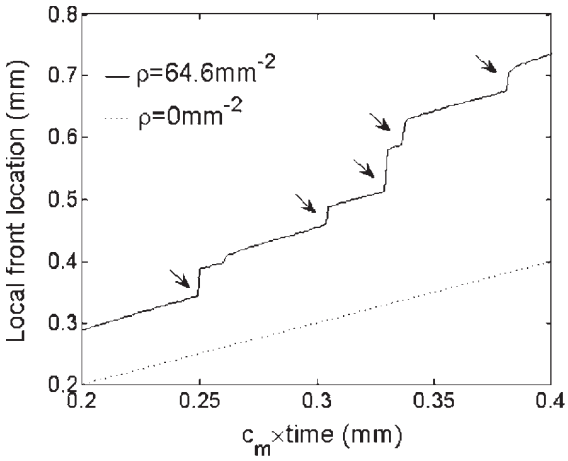

Fig. 21 Position of a given point of the front as a function of "time" $c_m \times t$ as obtained from the reconstructed simulation, applied to a fracture surface with $\rho = 64.6\,\mathrm{mm}^{-2}$. The *dotted straight line* indicates the expected position without microdamage (slope 1). The jumps (*arrows*) correspond to coalescence events with microcracks (taken from Guerra et al. (2012))

the nucleation sites are located at the nodes of a square lattice. One then gets:

$$A = \frac{1}{1 - b\bar{d}_n\sqrt{\rho}} \tag{9}$$

where b is a numerical factor close to 1. This equation, combined with the evolution of $\bar{d}_n$ with ρ (red line in Fig. 17:Top) is plotted in red in Fig. 20. It reproduces quite well both the data coming from experiments (grey triangles) and from the reconstruction (black circle). Note the change in slope observed at ρ_{sat}. At this point, $A = A_b = 1.63$. This corresponds to a ratio $A_b \simeq v_b/c_m$ where $v_b \simeq 0.4C_R$ is associated with the microbranching onset. This strongly suggests that the steric effect at the origin of the saturation of d_n also plays a role (still unsolved) in the micro-branching instability.

7 Conclusion

The series of experiments reported here were designed to shed light on the dissipation mechanisms that develop during fast crack growth in PMMA. They reveal that, above a well-defined critical velocity, PMMA stops to behave *stricto senso* as a nominally brittle material and crack propagation goes hand in hand with microcracks that nucleate and grow ahead of the main front. Those let characteristic marks (referred to as conic marks) on post-mortem fracture surfaces, the morphological analysis of which allowed us to reconstruct the full dynamics of microfailure events. The simultaneous space and time accuracies were of the order of the micrometer and the tens on nanoseconds, *i.e.* far better than what was reachable up to now.

Analysis of the reconstructions demonstrated that the true local propagation velocity of single cracks remains limited to a fairly low value $c_m \simeq 0.23c_R$ in PMMA, while the *apparent* fracture velocity measured at the continuum-level scale (*e.g.* via potential drop method) can be much higher (up to twice larger!). Such an anomalously high measured velocity results in fact from the coalescence of microcracks with each other and with the main front—all of them growing at c_m. In other words, the main effect of microdamaging is not, as commonly believed up to now (Washabaugh and Knauss 1994; Ravi-Chandar 2004; Ravi-Chandar and Yang 1997), to slow down fracture by increasing dissipation within the DZ, but on the contrary to boost the propagation of the main crack.

The value c_m that limits the true local velocity of single crack fronts in PMMA in the dynamic regime is set by the physico-chemical dissipative and nonlinear processes that develop within the FPZ, *e.g.* thermal (Estevez et al. 2000), viscoelastic effects (Boudet et al. 1996; Persson and Brener 2005) or hyperelastic processes (Bouchbinder et al. 2008; Livne et al. 2010; Buehler et al. 2003). Hence, it will depend on the considered material. On the other hand, the boost effect induced by microdamaging takes the form of a purely geometrical factor controlled by two quantities: the density ρ of nucleation sites and the mean distance $\overline{d_n}$ at nucleation. These two variables fully characterize the damaging state within DZ and evolve with K. Once ascribed, they permit to relate the continuum-level fracture speed v to the true local propagation speed c_m. Further work is required to unravel how PMMA internal structure selects the material-dependent quantities c_m, $\rho(K)$, and $\overline{d_n}$.

The presence of microcracks forming ahead of a dynamically growing crack has been evidenced in a variety of materials, *e.g.* in most brittle polymers (Ravi-Chandar 1998; Du et al. 2010), in rocks (Ahrens and Rubin 1993), in some nanophase ceramics and nanocomposites (Rountree et al. 2002), in oxide glasses (Rountree et al 2010), in metallic glasses (Murali et al. 2011), etc. The boost mechanism demonstrated here on PMMA is expected to hold in this whole class of

materials. Note that a similar effect has also been invoked (Prades et al 2005) in oxide glasses under stress corrosion, where ultra-slow crack propagation (down to few tenths of nanometers per second) was found to involves "nano"-cracks forming ahead of the main crack tip (Célarié et al. 2003; Celarié et al. 2003; Bonamy et al. 2006; Ferretti et al. 2011). This yields us to conjecture that failures with anomalously high apparent velocities measured at continuum-level scale may arise in all situations involving propagation-triggered microcracks, including *e.g.* shear fracture in compressed granite (Moore and Lockner 1995) , thermal failure in shale (Kobchenko et al. 2011; Panahi et al. 2012), and more generally failure of so-called quasi-brittle materials. One of the most robust observations in this field is the power-law form followed by the distribution in size and in time interval between two successive microfailure events (see Bonamy 2009; Deschanel et al. 2009 for recent reviews). It would be interesting to see whether or not the microcracking evidenced here in dynamic failure regime share the same scale-free features and how this affects the boost effect. Work in this direction is under way.

Acknowledgments We warmly thank K. Ravi-Chandar (Univ. of Texas, Austin) for many illuminating discussions. We also thank T. Bernard (SPCSI) for technical support, P. Viel and M. Laurent (SPCSI) for gold deposits, and A. Prevost (ENS, Paris) for his help with the profilometry measurements at ENS. We also acknowledge funding from French ANR through Grant No. ANR-05-JCJC-0088 and from Triangle de la Physique through Grant No. 2007-46.

References

Adda-Bedia M, Arias R, Ben-Amar M, Lund F (1999) Dynamic instability of brittle fracture. Phys Rev Lett 82:2314

Ahrens TJ, Rubin AM (1993) Impact-induced tensional failure in rocks. J Geophys Res Planets 98:1185

Anthony SR, Chub JP, Congleton J (1970) The crack branching velocity. Philos Mag 22:1201

Bergkvist H (1974) Some experiments on crack motion and arrest in polymethylmethacrylate. Eng Fract Mech 6:621

Bonamy D (2009) Intermittency and roughening in the failure of brittle heterogeneous materials. J Phys D Appl Phys 42:214014

Bonamy D, Prades S, Rountree CL, Ponson L, Dalmas D, Bouchaud E, Ravi-Chandar K, Guillot C (2006) Nanoscale damage during fracture in silica glass. Int J Fract 140:3

Bouchbinder E, Mathiesen J, Procaccia I (2005) Branching instabilities in rapid fracture: dynamics and geometry. Phys Rev E 71:056118

Bouchbinder E, Livne A, Fineberg J (2008) Weakly nonlinear theory of dynamic fracture. Phys Rev Lett 101:264302

Bouchbinder E, Marder M, Fineberg J (2010) The physics of simple cracks. Annu Rev Condens Matter Phys 1:371

Boudet JF, Ciliberto S, Steinberg V (1995) Experimental study of the instability of crack propagation in brittle materials. Europhys Lett 30:337

Boudet JF, Ciliberto S, Steinberg V (1996) Dynamics of crack propagation in brittle materials. J Phys II Fr 6:1493

Bruhwiler E, Wittmann FH (1990) The wedge splitting test: a method for performing stable fracture mechanics tests. Eng Fract Mech 35:117

Buehler MJ, Abraham FF, Gao HJ (2003) Hyperelasticity governs dynamic fracture at a critical length scale. Nature 426:141

Cast3M finite element code: http://www-Cast3M.cea.fr/

Celarié F, Prades S, Bonamy D, Dickelé A, Bouchaud E, Guillot C, Marlière C (2003) Surface fracture of glassy materials as detected by real-time atomic force microscopy (afm) experiments. Appl Surf Sci 212:92–96

Célarié F, Prades S, Bonamy D, Ferrero L, Bouchaud E, Guillot C, Marlière C (2003) Glass breaks like metal but at the nanometer scale. Phys Rev Lett 90:075504

Cotterell B (1968) Fracture propagation in organic glasses. Int J Fract 4:209

Deschanel S, Vanel L, Godin N, Maire E, Vigier G, Ciliberto S (2009) Mechanical response and fracture dynamics of polymeric foams. J Phys D Appl Phys 42:214001

Du P, Xue B, Song Y, Zuo M, Lu S, Zheng Q, Yu J (2010) Experimental observation and computer simulation of conic markings on fracture surfaces of polymers. J Mater Sci 45:3088

Estevez R, Tijssens MGA, der Giessen EV (2000) Modeling of the competition between shear yielding and crazing brittle polymers. J Mech Phys Solids 48:2585

Ferretti D, Rossi M, Royer-Carfagni G (2011) An espi experimental study on the phenomenon of fracture in glass. Is it brittle or plastic? J Mech Phys Solids 59:1338–1354

Fineberg J, Gross SP, Marder M, Swinney HL (1992) Instability in the propagation of fast cracks. Phys Rev B 45:5146

Fond C, Schirrer R (2001) Dynamic fracture surface energy values and branching instabilities during rapid crack propagation in rubber toughened pmma. C R Acad Sci Paris 329:195

Freund LB (1990) Dynamic fracture mechanics. Cambridge University Press, Cambridge

Goldman T, Livne A, Fineberg J (2010) Acquisition of inertia by a moving crack. Phys Rev Lett 104:114301

Guerra C, Scheibert J, Bonamy D, Dalmas D (2012) Understanding fast macroscale fracture from microscale post-mortem patterns. Proc Natl Acad Sci USA 109:390

Gumbsch P, Zhou SJ, Holian BL (1997) Molecular dynamics investigation of dynamic crack stability. Phys Rev B 55:3445

Henry H (2008) Study of the branching instability using a phase field model of inplane crack propagation. EPL 83:16004

Henry H, Levine H (2004) Dynamic instabilities of fracture under biaxial strain using a phase field model. Phys Rev Lett 93:105504

Holloway DG (1968) The fracture of glass. Phys Educ 3:317

Hull D (1999) Fractography: observing, measuring and interpreting fracture surface topography. Cambridge University Press, Cambridge

Irwin GR, Kies JA (1952) Fracturing and fracture dynamics. Weld J Res Suppl 31:95

Kalthoff JF, Winkler S, Beinert J (1976) Dynamical stress inetnsity factors for arresting cracks in dbc specimens. Int J Fract 12:317

Karihaloo BL, Xiao QZ (2001) Higher order terms of the crack tip asymptotic field for a wedge-splitting specimen. Int J Fract 112:129

Kies JA, Sullivan AM, Irwin GR (1950) Interpretation of fracture markings. J Appl Phys 21:716

Kobchenko M, Panahi H, Renard F, Dysthe KD, Malthe-Sørenssen A, Mazzini A, Scheibert J, Jamtveit B, Meakin P (2011) 4D imaging of fracturing in organic-rich shales during heating. J Geophys Res 116:B12201

Livne A, Bouchbinder E, Svetlizky I, Fineberg J (2010) The near-tip fields of fast cracks. Science 327:1359

Moore DE, Lockner DA (1995) The role of microcracking in shear-fracture propagation in granite. J Struct Geol 17:95

Murali P, Guo T, Zhang Y, Narasimhan R, Li Y, Gao H (2011) Atomic scale fluctuations govern brittle fracture and cavitation behavior in metallic glasses. Phys Rev Lett 107:215501

Panahi H, Kobchenko M, Renard F et al (2012) A 4D Synchrotron X-ray-tomography study of the formation of hydrocarbon-migration pathways in heated organic-rich shale. SPE J. SPE-162939-PA (in press; posted 27 November 2012)

Persson BNJ, Brener EA (2005) Crack propagation in viscoelastic solids. Phys Rev E 71:036123

Prades S, Bonamy D, Dalmas D, Bouchaud E, Guillot C (2005) Nano-ductile crack propagation in glasses understress corrosion: spatiotemporal evolution of damage in the vicinity of the crack tip. Int J Solids Struct 42:637

Rabinovitch A, Bahat D (2008) Mirror-mist transition in brittle fracture. Phys Rev E 78:067102

Rabinovitch A, Belizovsky G, Bahat D (2000) Origin of mist and hackle patterns in brittle fracture. Phys Rev B 61:14968

Ravi-Chandar K (1998) Dynamic fracture of nominally brittle materials. Int J Fract 90:83

Ravi-Chandar K (2004) Dynamic fracture. Elsevier Ltd, Amsterdam

Ravi-Chandar K, Knauss WG (1984) An experimental investigation into dynamic fracture-ii microstructural aspects. Int J Fract 25:65

Ravi-Chandar K, Yang B (1997) On the role of microcracks in the dynamic fracture of brittle materials. J Phys Mech Solids 45:535

Regel VR (1951) O mekhanizme kheupkogo razusheniya plastmass. Zhurnal Tekhnicheskoi Fiziki 21:287

Rice JR (1968) A path independent integral and the approximate analysis of strain concentration by notches and cracks. J Appl Mech 35:379

Rosakis AJ, Duffy J, Freund LB (1984) The determination of dynamic fracture toughness of aisi 4340 steel by the shadow spot method. J Mech Phys Solids 32:443

Rountree CL, Kalia RK, Lidorikis E, Nakano A, Brutzel LV, Vashishta P (2002) Atomistic aspects of crack propagation in brittle materials: multimillion atom molecular dynamics simulations. Annu Rev Mater Res 32:377

Rountree CL, Bonamy D, Dalmas D, Prades S, Kalia RK, Guillot C, Bouchaud E (2010) Fracture in glass via molecular dynamics simulations and atomic force microscopy experiments. Phys Chem Glass Eur J Glass Sci Technol B 51:127

Scheibert J, Guerra C, Célarié F, Dalmas D, Bonamy D (2010) Brittle/quasi-brittle transition in the dynamic fracture of nominally brittle materials: an energetic signature. Phys Rev Lett 104:045501

Sharon E, Fineberg J (1999) Confirming the continuum theory of dynamic brittle fracture for fast cracks. Nature 397:333

Smekal (1953) Zum bruchvorgang bei sprodem stoffverhalten unter ein-and mehrachsigen beanspruchungen. Osterr Ing Arch 7:49

Spatschek R, Hartmann M, Brener E, Muller-Krumbhaar H, Kassner K (2006) Phase field modeling of fast crack propagation. Phys Rev Lett 96:015502

Washabaugh PD, Knauss W (1994) A reconciliation of dynamic crack velocity and rayleigh wave speed in isotropic brittle solids. Int J Fract 65:97

DOI 10.1007/978-3-319-04397-5_9
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9840-5

ORIGINAL PAPER

Tight sedimentary covers for CO_2 sequestration

D. Leguillon · E. Karnaeva · A. Baroni · C. Putot

Received: 31 October 2012 / Accepted: 19 March 2013 / Published online: 29 March 2013

Abstract CO_2 storage at depth is a promising way to reduce the spread of greenhouse gases in the atmosphere. Obviously the sedimentary cover should ensure the sealing of reservoirs. The latter are periodically fractured to allow injection and we propose a model to predict the maximum bearable gas pressure before reinitiating these fractures in the caprock or incidentally along the interface between the reservoir and the caprock. The method is based on a twofold criterion merging energy and stress conditions. Specific conditions related to the gas pressure acting on the crack faces and the swelling of the reservoir due to the pressure rise require taking into account several terms in addition to the classical singular term that describes the state of stress at the tip of the main crack.

Keywords Rock mechanics · Brittle fracture · Asymptotics

1 Introduction

Due to increased human activity, reducing the amount of greenhouse gas emitted in the atmosphere to prevent global warming is becoming a big problem for mankind (GIEC 2005). Although efforts are made to reduce fuel consumption especially in transportation and replace them with less pollutant fluids, it seems necessary to find methods for their elimination. For CO_2, one possible solution is the storage in suitable deep underground geological formations made of porous rocks and covered by layers of impermeable sediments like depleted oil reservoirs, coal seams or aquifers. However, this technique is not free of drawbacks. The risk of leakage through the sediments along existing fracture networks, or caused by the high pressure injection, is a weak point. Moreover, this phenomenon is enhanced by the corrosive properties of CO_2 that can be interpreted, from the mechanical viewpoint, as a degradation of the properties of the surrounding rocks.

The reservoir itself is highly fractured (otherwise injection would be impossible) and our aim is to establish the conditions for which the elastic contrast between the reservoir and the sediment layers allows inhibiting the intrusion of pre-existing dyke-type fractures (i.e. perpendicular to the bedding) into the caprock or deflecting them into sill-type cracks (parallel to the bedding, i.e. along the interface between reservoir and caprock).

A first criterion for crack deflection by an interface, based on energy arguments, was proposed by He and Hutchinson (HH) in 1989. It has been discussed in several papers and we revisited it a first time in 2000 (Leguillon et al. 2000) considering the mechanism proposed by Cook and Gordon (1964). More

D. Leguillon (✉) · E. Karnaeva
Institut JLRA, CNRS UMR 7190, Université P. et M. Curie,
4 Place Jussieu, 75252 Paris Cedex 05, France
e-mail: dominique.leguillon@upmc.fr

E. Karnaeva · A. Baroni · C. Putot
IFP-Energies Nouvelles, DMA, 1-4 avenue de Bois-Préau,
92852 Rueil-Malmaison Cedex, France

recently, a general purpose criterion to predict crack initiation in a brittle material has been developed (Leguillon 2002). It is based on two conditions in energy and stress which must be simultaneously fulfilled and has proven effective in many situations. It led to a new version of the HH criterion involving both toughness and strength of the materials and the interfaces (Leguillon and Martin 2012a).

In the present work, the additional difficulty is to take into account the variation of fluid pressure acting on the walls of the fracture before and following a crack extension either in the next sediment bed or along the interface. On the one hand, the classic Williams' series at the tip of a crack impinging on an interface is enriched to include terms accounting for the fluid pressure and the poroelastic behaviour of the reservoir. It provides a description of the near stress field before the onset of a crack extension. On the other hand, a matched asymptotic expansions procedure yields an expression for the incremental energy release rate due to a virtual crack extension. Emphasis is put: (i) on the contrast between the materials analyzing different situations; and (ii) on the role of the fluid pressure.

Provided that mechanical parameters and boundary conditions are known for a given coverage area, this approach should provide an upper bound for admissible fluid pressure to avoid growth of a potential fracture and thus to have a safe storage.

It should be emphasized that it is by no means the study of crack propagation under the effect of a fluid pressure as in hydraulic fracturing (e.g. Detournay 1999; Loret and Radi 2001; Mishuris et al. 2012). There is no dynamic effect, the pre-existing cracks in the reservoir are fixed, the loading due to CO_2 injection is quasi-static and we are interested only in the possibility of reinitiating, either by penetrating the caprock or by deflecting along the interface between the reservoir and the caprock. Clearly, for safety reasons, the cracks should not penetrate the caprock, thus the initiation step is crucial. The subsequent (dynamic) growth of these cracks is outside the scope of this work.

2 The model

The application to CO_2 storage is presented on a "representative" volume (we do not discuss here the concept of representativeness) (Fig. 1). It takes into account the pressure inside the reservoir p_{CO_2} (MPa), the overburden vertical compression p_V (MPa) and the horizontal tectonic confinement (Karnaeva 2012; Karnaeva et al. 2012). To take full advantage of the supercritical properties of CO_2, the injection depth is at least 800 m. This determines the overburden load p_V due to the sediments weight which increases with depth with a gradient around 2 MPa per 100 m. Throughout this paper $p_V = 20$ MPa is selected corresponding to an injection depth of approximately 1,000 m. The tectonic confinement is modelled by vanishing horizontal displacements on the vertical boundaries. Such a condition can be also interpreted as a symmetry condition, the same situation reproducing periodically. This assumption is known to correspond to a low confinement, other models can be used, either a constant tectonic compression directly proportional to the sediments weight, or better a stress state depending both on the overburden load and the fluid pressure in the reservoir.

There are some difficulties in setting equations and boundary conditions. Following Terzaghi's law (1936) the total stress is considered. The rock forming the reservoir is assumed permeable and poro-elastic (Young's modulus $E^{(R)} = 2$ GPa, Poisson's ratio $\nu^{(R)} = 0.3$, Biot's coefficient $\alpha^{(R)} = 1$). The constitutive law is analogous to that of thermo-elasticity (Segall and Fitzgerald 1998), the temperature change being replaced by the pore pressure change, and the thermal expansion coefficient by a term involving Biot's coefficient $\alpha^{(R)}$

$$\underline{\underline{\sigma}} = \mathbf{C}^{(R)} : \left(\underline{\underline{\varepsilon}} - \frac{1-2\nu^{(R)}}{E^{(R)}} \alpha^{(R)} p_{CO_2} \underline{\underline{I}} \right) \tag{1}$$

where, for simplicity, $\mathbf{C}^{(R)}$ (MPa) denotes the elasticity matrix relying on Young's modulus and Poisson's ratio of the reservoir, $\underline{\underline{\varepsilon}}$ and $\underline{\underline{\sigma}}$ (MPa) the linearized strain and stress tensors and $\underline{\underline{I}}$ the identity matrix. The indices (R) and (C) are used to denote respectively the reservoir and the caprock.

The layer of covering sediments is assumed impermeable with an elastic constitutive law

$$\underline{\underline{\sigma}} = \mathbf{C}^{(C)} : \underline{\underline{\varepsilon}} \tag{2}$$

The total stress merges with the effective stress in the caprock.

Weak and strong contrasts in elasticity coefficients are analysed (reservoir parameters are keeping constant while for the caprock: $E^{(C)} = 5$ MPa, or $E^{(C)} = 40$ MPa, $\nu^{(C)} = 0.3$, $\alpha^{(C)} = 0$).

If the fluid pressure in the reservoir is low, the main crack is closed and the stress components of the solu-

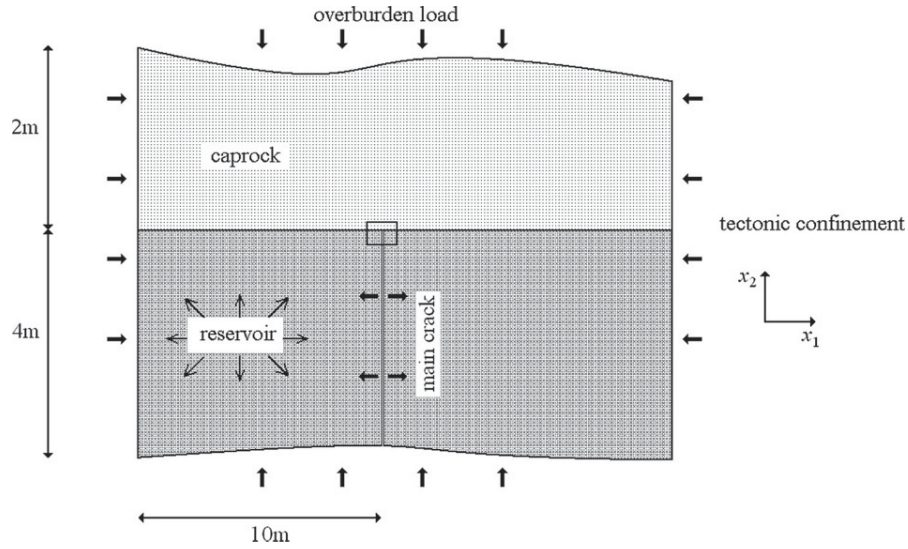

Fig. 1 The "representative" volume with a dyke in the reservoir impinging on the interface with the caprock made of sediments

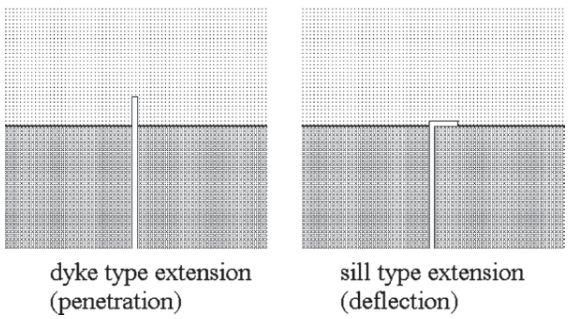

Fig. 2 The two possible mechanisms of crack growth

tion to the elastic problem are constant in the 2 domains and can be written (the horizontal displacement vanishes everywhere, $U_1 = 0$) under the assumption of plane strain elasticity

$$\begin{cases} \sigma_{11}^{(C)} = -\frac{\nu^{(C)}}{1-\nu^{(C)}} p_V \; ; \sigma_{22}^{(C)} = -p_V \; ; \sigma_{12}^{(C)} = 0 \\ \sigma_{11}^{(R)} = -\frac{\nu^{(R)}}{1-\nu^{(R)}} p_V - \alpha^{(R)} \frac{1-2\nu^{(R)}}{(1-\nu^{(R)})(1+\nu^{(R)})} p_{CO_2} \; ; \sigma_{22}^{(R)} \\ \quad = -p_V \; ; \sigma_{12}^{(R)} = 0 \end{cases} \quad (3)$$

The main crack (Fig. 1) opens as soon as the inside pressure exceeds the horizontal stress in the reservoir

$$p_{CO_2} > \frac{\nu^{(R)}\left(1+\nu^{(R)}\right)}{1-\nu^{(R)2}-\alpha^{(R)}\left(1-2\nu^{(R)}\right)} p_V \quad (4)$$

Then, the problem is to know if the main vertical crack in the reservoir is likely to extend or to arrest, and in case of extension to know if it evolves in a dyke or a sill type extension (Fig. 2).

3 Williams' expansion

For carrying out the matched asymptotic expansions (Van Dyke 1964), it is necessary, in a first step, to define the Williams expansion (1959) in the vicinity of the crack tip prior to its extension. In addition to the usual singular terms, complementary terms must be inserted to take into account both the fluid pressure acting on the faces of the main crack and the reservoir swelling as a consequence of the injection of fluid. Then since these two terms have constant stress components, it is not allowed omitting the T-stress term which enjoys the same property and is thus of the same order of magnitude. (throughout this paper the Cartesian and polar coordinates are mixed without confusion, Fig. 3)

$$\begin{aligned} \underline{U}^0(r,\theta) = \underline{C} &+ kr^{\lambda}\underline{u}(\theta) \\ &+ Tr\underline{t}(\theta) - p_{CO_2} r\underline{v}(\theta) \\ &+ p_{CO_2}\beta^{(R)} r\underline{w}(\theta) + \cdots \\ &\text{with; } \beta^{(R)} = \frac{\alpha^{(R)}}{1+\nu^{(R)}} \end{aligned} \quad (5)$$

$$\begin{aligned} \underline{\underline{\sigma}}^0(r,\theta) = kr^{\lambda-1}\underline{\underline{s}}(\theta) + T\underline{\underline{\tau}}(\theta) - p_{CO_2}\underline{\underline{v}}(\theta) \\ + \beta^{(R)} p_{CO_2}\underline{\underline{\omega}}(\theta) + \cdots \end{aligned} \quad (6)$$

The second line (6) derives from (5) through the constitutive laws (1) and (2). The explanation for the index 0 will be found further in Eq. (14). Units for $\underline{U}^0$ and $\underline{C}$ are m; for k and T they are respectively MPa.m$^{1-\lambda}$ and MPa; units for $\underline{u}$, $\underline{t}$, $\underline{v}$ and $\underline{w}$ are MPa^{-1}; $\underline{\underline{s}}$, $\underline{\underline{\tau}}$, $\underline{\underline{v}}$ and $\underline{\underline{\omega}}$ are dimensionless.

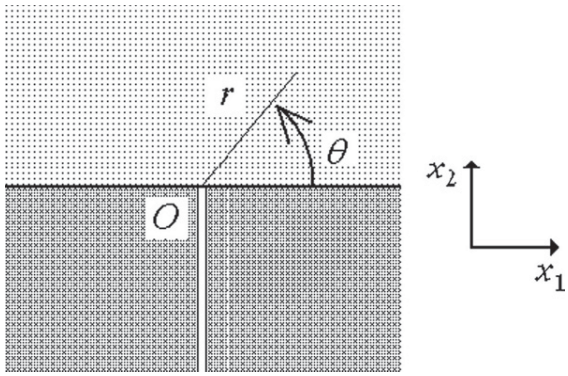

Fig. 3 The Cartesian x_1, x_2 and polar r, θ coordinates emanating from the crack tip O

Besides the constant term $\underline{C}$, the first relevant one is singular (i.e. $\lambda < 1$) but although there is a crack tip, the exponent differs from 1/2, it is larger if $E^{(C)} > E^{(R)}$ (as in the present case) and smaller in the opposite case. It is associated with a symmetric mode $\underline{u}(\theta)$ (another mode is associated with λ but is antisymmetric and thus not activated in the symmetric case illustrated in Fig. 1). The exponent and the associated mode are easily computed (Leguillon and Sanchez-Palencia 1987) solving numerically an eigenvalue problem.

This analysis is limited to the case $\lambda > 1/2$ named "weak singularity" (Leguillon and Sanchez-Palencia 1992) which is the only one to have a strengthening effect (Leguillon and Martin 2012b). The opposite situation is unfavourable to crack arrest and promotes penetration (He and Hutchinson 1989; Leguillon et al. 2000; Leguillon and Martin 2012a).

The next term is the T-stress. It is defined, under the assumption of plane strain elasticity, in a vicinity of the main crack tip (Fig. 1) and has not to satisfy remote boundary conditions. It can be calculated easily by hand knowing that it must check the first relations in (8) and that the parameter a is adjusted to fulfil the displacement continuity through the interface

$$\begin{cases} t_1^{(C)}(\theta) = \left[\frac{(1+\nu^{(C)})(1-\nu^{(C)})}{E^{(C)}}a - \frac{\nu^{(C)}(1+\nu^{(C)})}{E^{(C)}}\right]\cos\theta \\ t_2^{(C)}(\theta) = \left[\frac{(1+\nu^{(C)})(1-\nu^{(C)})}{E^{(C)}} - \frac{\nu^{(C)}(1+\nu^{(C)})}{E^{(C)}}a\right]\sin\theta \\ t_1^{(R)}(\theta) = -\frac{\nu^{(R)}(1+\nu^{(R)})}{E^{(R)}}\cos\theta \\ t_2^{(R)}(\theta) = \frac{(1+\nu^{(R)})(1-\nu^{(R)})}{E^{(R)}}\sin\theta \end{cases}$$
$$\text{with } a = \frac{\nu^{(C)}}{1-\nu^{(C)}} - \frac{E^{(C)}}{E^{(R)}}\frac{\nu^{(R)}(1+\nu^{(R)})}{(1+\nu^{(C)})(1-\nu^{(C)})} \tag{7}$$

It enjoys the following properties

$$\sigma_{22}^{(C)} = \sigma_{22}^{(R)} = 1; \quad \sigma_{12}^{(C)} = \sigma_{12}^{(R)} = 0;$$
$$\sigma_{11}^{(R)} = 0; \quad \sigma_{11}^{(C)} = a \tag{8}$$

The last property is original and due to the contrast between materials, in addition to a tension parallel to the crack the material undergoes a transverse tension ahead of the tip (parameter a vanishes if there is no contrast between materials).

The third term is associated with constant pressure acting on the main crack walls, again it can be calculated easily by hand

$$\begin{cases} v_1^{(C)}(\theta) = \frac{(1+\nu^{(R)})(1-\nu^{(R)})}{E^{(R)}}\cos\theta \\ v_2^{(C)}(\theta) = -\frac{\nu^{(C)}(1+\nu^{(R)})(1-\nu^{(R)})}{E^{(R)}(1-\nu^{(C)})}\sin\theta \\ v_1^{(R)}(\theta) = \frac{(1+\nu^{(R)})(1-\nu^{(R)})}{E^{(R)}}\cos\theta \\ v_2^{(R)}(\theta) = -\frac{\nu^{(R)}(1+\nu^{(R)})}{E^{(R)}}\sin\theta \end{cases} \tag{9}$$

$$\sigma_{22}^{(C)} = \sigma_{22}^{(R)} = 0; \quad \sigma_{12}^{(C)} = \sigma_{12}^{(R)} = 0; \quad \sigma_{11}^{(R)} = 1;$$
$$\sigma_{11}^{(C)} = \frac{E^{(C)}}{E^{(R)}}\frac{\left(1+\nu^{(R)}\right)\left(1-\nu^{(R)}\right)}{\left(1+\nu^{(C)}\right)\left(1-\nu^{(C)}\right)} \tag{10}$$

The fourth term taking into account the swelling of the reservoir is calculated by hand as well

$$\begin{cases} w_1^{(C)}(\theta) = \frac{(1+\nu^{(C)})(1-\nu^{(C)})}{E^{(C)}}b\cos\theta \\ w_2^{(C)}(\theta) = -\frac{\nu^{(C)}(1+\nu^{(C)})}{E^{(C)}}b\sin\theta \\ w_1^{(R)}(\theta) = \frac{(1-2\nu^{(R)})(1+\nu^{(R)})}{E^{(R)}}\cos\theta \\ w_2^{(R)}(\theta) = \frac{(1-2\nu^{(R)})(1+\nu^{(R)})}{E^{(R)}}\sin\theta \end{cases}$$
$$\text{with } b = \frac{E^{(C)}}{E^{(R)}}\frac{\left(1-2\nu^{(R)}\right)\left(1+\nu^{(R)}\right)}{\left(1+\nu^{(C)}\right)\left(1-\nu^{(C)}\right)} \tag{11}$$

$$\sigma_{11}^{(R)} = \sigma_{22}^{(R)} = 1; \quad \sigma_{12}^{(C)} = \sigma_{12}^{(R)} = 0; \quad \sigma_{22}^{(C)} = 0;$$
$$\sigma_{11}^{(C)} = b \tag{12}$$

In the Williams expansion (5), this term has the multiplier $\beta^{(R)}$ which accounts for the poro-elastic law (1).

The generalized stress intensity factors (GSIF) k and T in (5) and (6) are computed using a path independent integral (Appendix 1). Note that in the present analysis the GSIF's k and T are extracted from

$$\hat{\underline{U}}^0(r,\theta) = \underline{U}^0(r,\theta) + p_{CO_2}\, r\underline{v}(\theta) - p_{CO_2}\, r\beta^{(R)}$$
$$\underline{w}(\theta) = \underline{C} + kr^{\lambda}\underline{u}(\theta) + Tr\underline{t}(\theta) + \cdots \tag{13}$$

Taking advantage of the properties of the path integral (the right hand side is a classical Williams expansion

with equilibrium and boundary conditions homogeneous to 0, Appendix 1). Clearly k is mainly governed by p_{CO_2} but receives also contributions from p_V and conversely T is mainly governed by p_V but receives contributions from p_{CO_2}, through $\hat{\underline{U}}^0$ in (13).

4 Outer and inner expansions

The crack extension lengths l_d(deflected) and l_p (penetrated) are assumed to be small compared to the dimensions of the representative volume. The outer expansion can be written

$$\begin{aligned}&\underline{U}(x_1, x_2, l_{d/p}) = \underline{U}^0(x_1, x_2) + \text{corrective terms ;}\\ &\text{where } \underline{U}^0(x_1, x_2) = \underline{U}(x_1, x_2, 0)\end{aligned} \tag{14}$$

It means that, at the leading order, the outer expansion is independent of the extension mechanism, either deflection or penetration. The crack extension is small and not visible at this scale.

The inner expansion takes a special form which is consequence of the matching conditions (Van Dyke 1964; Leguillon and Sanchez-Palencia 1987) governed at the leading order by the Williams expansion (5)

$$\begin{aligned}\underline{U}(x_1, x_2, l_{d/p}) &= \underline{U}(l_{d/p}y_1, l_{d/p}y_2, l_{d/p})\\ &= \underline{C} + k\, l_{d/p}^{\lambda}\left[\rho^{\lambda}\underline{u}(\theta) + \underline{V}^1_{d/p}(y_1, y_2)\right]\\ &+T\, l_{d/p}\left[\rho\underline{t}(\theta) + \underline{V}^2_{d/p}(y_1, y_2)\right]\\ &+p_{CO_2} l_{d/p}\left[\rho\underline{v}(\theta) + \underline{V}^3_{d/p}(y_1, y_2)\right]\\ &+\beta^{(R)} p_{CO_2} l_{d/p}\left[\rho\underline{w}(\theta) + \underline{V}^4_{d/p}(y_1, y_2)\right] + \cdots\end{aligned} \tag{15}$$

The $\underline{V}^j_{d/p}$'s are solution to well posed problems on the inner unbounded domain spanned by the dimensionless space variables $y_i = x_i/l_{d/p}(\rho = r/l_{d/p})$ (Appendix 2). In this domain the crack extension length is 1 whatever its actual value (Fig. 2).

So now we have a description of the solution prior to crack extension (with (14) and (5) or better (15) where all the $\underline{V}^j$'s vanish) and following it (with (14) and (15)). Then it is possible to calculate the change in potential energy $\delta W = W(l_{d/p}) - W(0)$(Appendix 3) between these two states and finally the incremental energy release rate

$$\begin{aligned}G^{inc}_{d/p} = -\frac{\delta W}{l_{d/p}} =& k^2 l_{d/p}^{2\lambda-1} A_{d/p} + kT l_{d/p}^{\lambda} B_{d/p}\\ &+kp_{CO_2} l_{d/p}^{\lambda} C_{d/p} + T^2 l_{d/p} D_{d/p} + p_{CO_2}^2 l_{d/p}\\ &E_{d/p}+Tp_{CO_2} l_{d/p} F_{d/p} + \cdots\\ =& k^2 l_{d/p}^{2\lambda-1}\left[A_{d/p} + m_1(l_{d/p})B_{d/p} + m_2(l_{d/p})C_{d/p}\right.\\ &+ m_1(l_{d/p})^2 D_{d/p} + m_2(l_{d/p})^2 E_{d/p}\\ &\left.+ m_1(l_{d/p})m_2(l_{d/p})F_{d/p}\right] + \cdots\end{aligned} \tag{16}$$

where the coefficients the $A_{d/p} \ldots F_{d/p}$ are extracted from the $\underline{V}^j_{d/p}$'s (Appendix 2, Karnaeva 2012). The functions $m_1(l_{d/p})$ and $m_2(l_{d/p})$ are dimensionless mixity parameters

$$m_1(l_{d/p}) = \frac{T}{k} l_{d/p}^{1-\lambda}\,; \quad m_2(l_{d/p}) = \frac{p_{CO_2}}{k} l_{d/p}^{1-\lambda} \tag{17}$$

It is assumed here that the constant gas pressure p_{CO_2} acts on the crack extension faces, which means that the fluid instantaneously reacts and fills the extension. It is a reasonable assumption for a fluid in a supercritical state. Nevertheless it corresponds to the more complicated situation, assuming that no pressure acts along the extension faces is easier, some terms vanish (Appendix 2). An intermediate solution is to postulate a known pressure profile.

5 The mixed criterion

The mixed criterion (Leguillon 2002) states that two conditions must be fulfilled to predict crack initiation at stress concentrations due to "weak" singularities (at corners for example): an energy rule and a stress condition. The energy balance implies that if failure occurs then

$$\begin{aligned}&k^2 l_{d/p}^{2\lambda-1} X(p_V, p_{CO_2}, l_{d/p}) \geq G_c^{(I/C)}\\ &\quad\text{with } X(p_V, p_{CO_2}, l_{d/p})\\ &= A_{d/p} + m_1(l_{d/p})B_{d/p} + m_1(l_{d/p})C_{d/p}\\ &\qquad + m_1(l_{d/p})^2 D_{d/p} + m_2(l_{d/p})^2 E_{d/p}\\ &\qquad + m_1(l_{d/p})m_2(l_{d/p})F_{d/p}\end{aligned} \tag{18}$$

where $G_c^{(I)}$ and $G_c^{(C)}$ denote respectively the toughness of the interface between the reservoir and the caprock and the toughness of the material forming the caprock. The dependence on p_V of X defined in (18) is through k and T via m_1 and m_2 (17), the dependence on p_{CO_2} is explicit through m_1 and arises also through k and T.

Note that the failure parameters of the interface ($G_c^{(I)}$ here and $\sigma_c^{(I)}$ further in Eq. (20)) are very difficult to determine and a good compromise is to select those of the weaker of the two materials separated by the

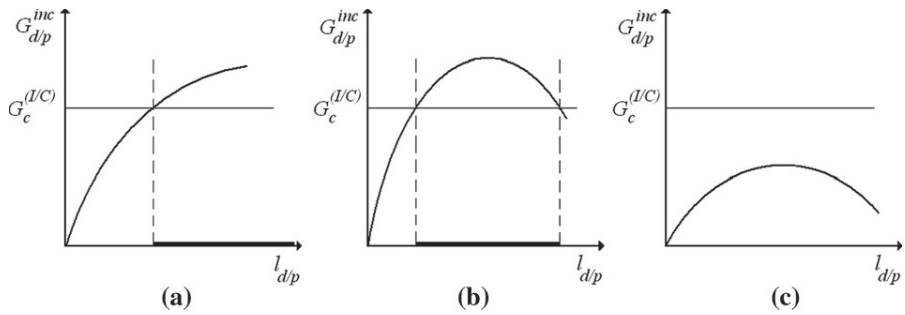

Fig. 4 Different situations occurring when solving inequality (18): **a** A wide range of admissible crack extension lengths (*thick line*), **b** A short range, **c** No solution

interface (Hutchinson et al. 1987; Karnaeva 2012), in general the most porous material.

If $\lambda > 0.5$, the function $G_{d/p}^{inc}$ is an increasing function of $l_{d/p}$ when $l_{d/p}$ is small. Then, on the one hand, this function can keep growing (within a reasonable range fulfilling the assumption of smallness on $l_{d/p}$) and thus inequality (18) has solutions and defines a lower bound of admissible crack extension lengths (thick line in Fig. 4a). On the other hand, if after the growing phase the function becomes decreasing, the range of admissible lengths is reduced to a segment (Fig. 4b) and may even disappear (Fig. 4c). In this latter case, when p_{CO_2} increases (c) turns into (b).

As observed in Fig. 4, we must draw attention to the fact that the admissible lengths cannot be indefinitely small. According to the energy balance equation, at failure there is a crack jump.

The stress condition can be stated as follows: prior to crack nucleation, the tension must exceed the tensile strength all along the presupposed crack path. This can be written using the Williams expansion (6) and the mixity parameters (17)

$$\sigma(r,\theta_{d/p}) = k\, r^{\lambda-1}\left[s(\theta_{d/p}) + m_1(r)\tau(\theta_{d/p}) - m_2(r)\nu(\theta_{d/p}) + \beta m_2(r)\omega(\theta_{d/p})\right] + \cdots \tag{19}$$

The angle $\theta_{d/p}$ is the angle made by the crack extension: $\theta_d = \pi/2$(deflection) and $\theta_p = \pi$(penetration). The functions σ, s, τ, ν and ω denote the normal stress component of the tensors $\underline{\underline{\sigma}}$, $\underline{\underline{s}}$, $\underline{\underline{\tau}}$, $\underline{\underline{\nu}}$ and $\underline{\underline{\omega}}$ in the crack extension direction (see (6)), i.e. the component 22 for the crack deflection and the component 11 for the crack penetration (Fig. 3).

It is observed that, within the range defined by the energy condition, the function $\sigma(r,\theta)$ is a decreasing function of r, then the stress condition can be written

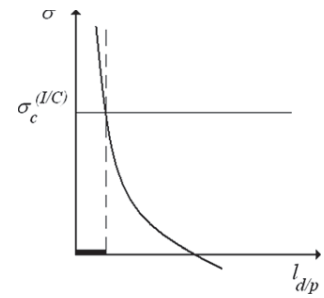

Fig. 5 Admissible crack extension lengths (*thick line*) when solving inequality (21)

$$\sigma(r,\theta_{d/p}) \geq \sigma_c^{(I/C)} \quad \text{for} \quad 0 < r \leq l_{d/p} \Rightarrow \sigma(l_{d/p},\theta_{d/p}) \geq \sigma_c^{(I/C)} \tag{20}$$

Where $\sigma_c^{(I/C)}$ denotes respectively the tensile strength of the interface (index I) and of the material forming the caprock (index C). Then

$$k\, l_{d/p}^{\lambda-1} Y(p_V, p_{CO_2}, l_{d/p}) \geq \sigma_c^{(I/C)} \quad \text{with } Y(p_V, p_{CO_2}, l_{d/p}) = s(\theta_{d/p}) + m_1(r)\tau(\theta_{d/p}) - m_2(r)\nu(\theta_{d/p}) + \beta m_2(r)\omega(\theta_{d/p}) \tag{21}$$

Solving this inequality leads to define an upper bound of admissible crack extension lengths (Fig. 5).

6 Numerical results—deflection—strong contrast

Compatibility between (18) and (21) must be considered separately in the cases illustrated in Fig. 4a–c. Obviously, the fracture parameter values $G_c^{(I/C)}$ and $\sigma_c^{(I/C)}$ are crucial to discriminate between the different cases. Figure 4a is the most favorable case. Indeed, (18) provides a lower bound for admissible crack extension lengths whereas (21) gives an upper bound. The compatibility leads to an implicit equation for $l_{d/p}$

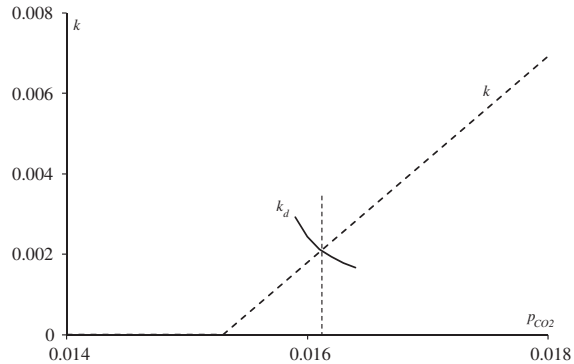

Fig. 6 The GSIF k compared to the critical deflection value $k_d(\text{GPam}^{1-\lambda})$ versus CO_2 injection pressure (GPa) for $p_V = 20$ MPa

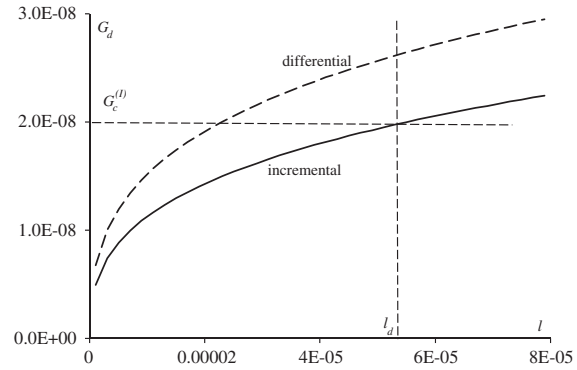

Fig. 7 Incremental (*solid line*) and differential (*dashed line*) energy release rates (GPam) at deflection near the onset length ($l = l_d$) versus the crack extension length l(m) for $p_{CO_2} = 16.1$ MPa

$$l_{d/p} = \frac{G_c^{(I/C)}}{X(p_V, p_{CO_2}, l_{d/p})} \frac{Y(p_V, p_{CO_2}, l_{d/p})^2}{\sigma_c^{(I/C)2}} \quad (22)$$

Once this equation is solved, either (18) or (21) gives the condition for failure

$$k \geq k_{d/p} = \left(\frac{G_c^{(I/C)}}{X(p_V, p_{CO_2}, l_{d/p})} \right)^{1-\lambda} \times \left(\frac{\sigma_c^{(I/C)}}{Y(p_V, p_{CO_2}, l_{d/p})} \right)^{2\lambda-1} \quad (23)$$

where k_d and k_p denote respectively the critical values of the GSIF k for deflection and penetration. If k_d is smaller than k_p then deflection is promoted and vice versa.

Such a situation is encountered when studying the deflection in case of a strong contrast between reservoir and caprock with $G_c^{(I)} = 20\ \text{Jm}^{-2}$ and $\sigma_c^{(I)} = 2$ MPa (quite small but realistic for a porous rock like the reservoir). Strong contrast means $E^{(C)}/E^{(R)} = 20$ (see Sect. 2) and $\lambda = 0.687$ in the Williams expansion (5). The injection depth is supposed around 1,000 m leading to a vertical compression $p_V = 20$ MPa.

As expected, according to (4), the main crack opening occurs for $p_{CO_2} = 15.3$ MPa. For lower p_{CO_2} values, the crack remains closed and $k = 0$, above k increases linearly with p_{CO_2}. The critical value k_d can be computed also as a function of p_{CO_2}, and deflection occurs when k exceeds k_d, i.e. for $p_{CO_2} \geq 16.1$ MPa (Fig. 6). The corresponding length is rather small $l_d = 55\ \mu\text{m}$, mainly because the compression prevails when one moves away from the crack tip and from the influence of the singular term. But at this step, i.e. when the short crack exists, the classical fracture mechanics takes precedence and a simple Griffith analysis can be carried out. Figure 7 shows the (differential) energy release rate derived from (16) for $p_{CO_2} = 16.1$ MPa

$$G_{d/p} = -\frac{\partial W_{d/p}}{\partial l} = k^2 l_{d/p}^{2\lambda-1} \left[2\lambda A_{d/p} + (\lambda+1) m_1(l_{d/p}) B_{d/p} + (\lambda+1) m_1(l_{d/p}) C_{d/p} + 2m_1(l_{d/p})^2 D_{d/p} + 2m_2(l_{d/p})^2 E_{d/p} + 2m_1(l_{d/p}) m_2(l_{d/p}) F_{d/p} \right] + \cdots \quad (24)$$

Clearly $G_d > G_c^{(I)}$ at deflection onset and the gap increases within Fig. 7 range. Passing at a larger scale we see in Fig. 8 that the curve bends and G_d passes below the value $G_c^{(I)}$ and go on decreasing. This point (roughly $l_{da} = 1$ cm) is a good candidate for crack

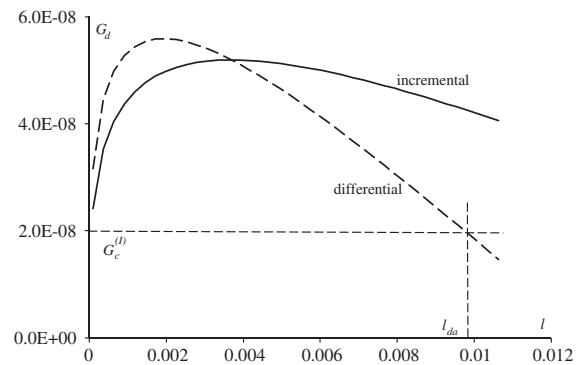

Fig. 8 Incremental (*solid line*) and differential (*dashed line*) energy release rates (GPam) at deflection near the crack arrest point ($l = l_{da}$) versus the crack extension length l(m) for $p_{CO_2} = 16.1$ MPa

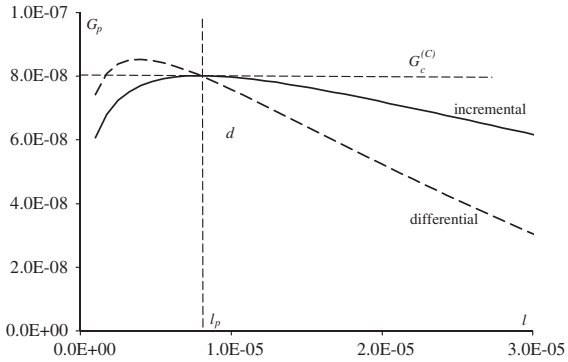

Fig. 9 Incremental (*solid line*) and differential (*dashed line*) energy release rates (GPa.m) at penetration ($l = l_p$) versus the crack extension length (m) for $p_{CO_2} =$ 20.2MPa

arrest, note that it remains small compared to the representative volume (Fig. 1) and well within the asymptotic framework.

Nevertheless, this arrest length l_{da} is very sensitive to a small increase in p_{CO_2}; for $p_{CO_2} = 16.5$ MPa, $l_{da} = 30\ cm$ and finally for $p_{CO_2} = 17$ MPa no arrest can be found (i.e. no small arrest length within the asymptotic framework). As a conclusion $p_{CO_2} \approx$ 16 MPa is the critical value not to be exceeded to avoid crack deflection.

7 Numerical results: penetration—strong contrast

This case refers to the situation observed in Fig. 4b, c for $G_c^{(C)} = 80$ Jm^{-2}, $\sigma_c^{(C)} = 10$ MPa. Mother crack opening still occurs for $p_{CO_2} \geq 15.3$ MPa, but G_p^{inc}(see (16)) passes through a maximum which remains smaller than $G_c^{(C)}$up to $p_{CO_2} = 20.2$ MPa (Fig. 9), preventing any penetration mechanism. At this point the tensile stress acting on the presupposed extension path (penetration) $\sigma = 289$ MPa, is far above the tensile strength because l_p is very small, around 8 μm, and the Young modulus of the material is high (strong contrast). Thus if penetration occurs, it is mainly governed by the energy criterion. But this defines also an arrest point since the (differential) energy release rate G is decreasing and takes exactly the value $G_c^{(C)}$(Fig. 9) for $l = l_p = l_{pa}$.

At this point, it is not necessary to go much further. Penetration appears over a meaningless length at a CO_2 pressure well above that which causes deflection, thus it is the latter mechanism that is naturally selected.

8 Numerical results—weak contrast

In case of a more realistic weak contrast: $E^{(C)}/E^{(R)} = 2.5$ (see Sect. 2) then $\lambda = 0.585$. The situation differs slightly, the exponent λ is closer to 0.5 and, as can be seen for instance in (23), the role of the stress condition regresses: $2\lambda - 1 = 0.17$ and then $1 - \lambda = 0.417$ (whereas it was respectively 0.374 and 0.313 in the strong contrast case). Both mechanisms, deflection and penetration, are mainly governed by the energy condition.

According to (4), the mother crack opening occurs for p_{CO_2} =15.3 MPa as before since Poisson's ratio $\nu^{(R)}$ and Biot's coefficient $\alpha^{(R)}$ of the reservoir are unchanged. There is a kind of hinge defined by $G_c^{(I)} =$ 20 Jm^{-2}, $\sigma_c^{(I)} = 2$ MPa, $G_c^{(C)} = 40$ Jm^{-2} and $\sigma_c^{(C)} =$ 4 MPa where the situation is that of Fig. 4b, c in both cases. One or the other mechanism is triggered when p_{CO_2} reaches 15.8 MPa, it is not possible to decide. In particular, the two initiation (and arrest) lengths l_d and l_p are almost equal, around 30 μm. Nevertheless, the mechanism seems more stable than in Sect. 6, for $p_{CO_2} = 20$ MPa, it comes $l_{da} \approx$9 cm and $l_{pa} \approx$8 cm.

On the one hand, if $G_c^{(I)}$ changes from 10 to 100 J m^{-2} without modifying the other parameters the results evolve as follows: deflection occurs for p_{CO_2} varying from 15.7 to 16.3 MPa and l_d increases from 15 to 146 μm. The initiation length changes much, but the arrest length for $p_{CO_2} = 20$ MPa changes little from 9.1 to 8.8 cm. On the other hand, changes in $\sigma_c^{(I)}$ within a reasonable range (from 1 to 10 MPa for instance) do not produce significant changes although when $\sigma_c^{(I)} \geq 3$ MPa, the solution is calculated using (23) because the stress condition (20) is no longer fulfilled when the curve G_d^{inc} is tangent to the horizontal line $G_c^{(I)}$. Thus the load must be (slightly) increased and the solution shifts to the left in the increasing part of G_d^{inc}.

Similarly, if $G_c^{(C)}$ changes from 10 to 100 J m^{-2} without modifying the other parameters: penetration occurs for p_{CO_2} varying from 15.6 to 16.0 MPa and l_p increases from 9 to 83 μm. The arrest length for $p_{CO_2} = 20$ MPa changes little from 7.8 to 7.6 cm. But now, any variation of the tensile strength $\sigma_c^{(C)}$ (within a wide range 1–30 MPa) has no effect at all because at initiation the stress condition (20) is by far true.

9 Conclusion

With almost realistic elastic and fracture parameters close to that of porous rocks (weak contrast and low fracture parameters), it is clear that if the dyke type crack in the reservoir starts to open (p_{CO_2} =15.3 MPa) as a result of the gas pressure rise due to injection, the conditions are nearly met (p_{CO_2} =15.8 MPa) for crack propagation either along the interface (sill) or in the caprock (dyke). Of course deflection is safer than penetration but both mechanisms are likely to trigger at CO_2 pressure values too close to each other to be meaningful.

The rock mass formed by the reservoir and the caprock retains its stability due to strong compressions prevailing at depth (tectonic confinement and overburden load). But as soon as the conditions are met for pre-existing cracks to open under tension, this equilibrium is near to being broken.

Appendix 1

Let us consider $\underline{f}$ and $\underline{g}$ solutions to an elastic problem locally homogeneous to 0 in the vicinity of the main crack tip (Fig. 1), i.e. such that the equilibrium equation has a vanishing right hand side member and such that the boundary conditions are vanishing as well in the vicinity under consideration. Then the following integral is path independent

$$\Psi\left(\underline{f},\underline{g}\right)=\frac{1}{2}\int_{\Gamma}\left[\underline{\underline{\sigma}}\left(\underline{f}\right).\underline{N}.\underline{g}-\underline{\underline{\sigma}}\left(\underline{g}\right).\underline{N}.\underline{f}\right]\mathrm{d}s$$

Here Γ is any contour starting on one face and finishing on the other face of the main crack (Fig. 1) and $\underline{N}$ the normal to this contour pointing toward the crack tip.

Then if $\underline{f}$ can be expanded in a Williams series

$$\underline{f}(x_1,x_2)=\underline{C}+kr^{\lambda}\underline{u}(\theta)+\cdots$$

The Generalized Stress Intensity Factor (GSIF) k is given by (Leguillon and Sanchez-Palencia 1987; Labossiere and Dunn 1999)

$$k=\frac{\Psi\left(\underline{f}(x_1x_2),r^{-\lambda}\underline{u}^{-}(\theta)\right)}{\Psi\left(r^{\lambda}\underline{u}(\theta),r^{-\lambda}\underline{u}^{-}(\theta)\right)}$$

where $r^{-\lambda}\underline{u}^{-}(\theta)$ is the so-called dual mode to $r^{\lambda}\underline{u}(\theta)$, it plays the role of an extraction function. Indeed, in 2D if λ is solution to the eigenvalue problem with and $\underline{u}(\theta)$ as eigenmode, then $-\lambda$ is also a solution with its own eigenvector $\underline{u}^{-}(\theta)$ (Leguillon and Sanchez-Palencia 1987). The primal mode has a finite energy in the vicinity of the crack tip while the dual mode has not but it plays no role since the contour Γ encompasses the crack tip at a finite distance.

Then, according to (13)

$$k=\frac{\Psi\left(\hat{\underline{U}}^{0}(x_1x_2),r^{-\lambda}\underline{u}^{-}(\theta)\right)}{\Psi\left(r^{\lambda}\underline{u}(\theta),r^{-\lambda}\underline{u}^{-}(\theta)\right)};$$

$$T=\frac{\Psi\left(\hat{\underline{U}}^{0}(x_1x_2),r^{-1}\underline{t}^{-}(\theta)\right)}{\Psi\left(r\underline{t}(\theta),r^{-1}\underline{t}^{-}(\theta)\right)}$$

Appendix 2

Let us consider now two solutions $\underline{f}^{0}$ and $\underline{f}^{l_{d/p}}$ to an elastic problem, again locally homogeneous to 0 in the vicinity of the main crack tip, corresponding respectively to the initial state illustrated in Fig. 1 and to a final state embedding a crack extension (deflection or penetration) with length $l_{d/p}$ (Fig. 2). Then the change in potential energy between these two states is given by

$$\delta W=\Psi\left(\underline{f}^{l_{d/p}}(x_1,x_2),\underline{f}^{0}(x_1,x_2)\right)$$

The term $\underline{f}^{0}$ admits a Williams expansion

$$\underline{f}^{0}(x_1,x_2)=\underline{C}+kr^{\lambda}\underline{u}(\theta)+\cdots$$

As a consequence of matching conditions the inner expansion of $\underline{f}^{l_{d/p}}$ is

$$\underline{f}^{l_{d/p}}(x_1,x_2)=\underline{C}+kl_{d/p}^{\lambda}\left[\rho^{\lambda}\underline{u}(\theta)+\underline{V}^{1}_{d/p}(y_1,y_2)\right]+\cdots\quad\text{with } y_i=x_i/l \text{ and } \rho=r/l$$

where $\underline{V}^{1}_{d/p}$ is solution to an elastic problem posed on the unbounded domain spanned by the y_i's, with an homogeneous equilibrium equation, vanishing forces along the mother crack faces, a decay to 0 at infinity and prescribed forces on the two faces $\gamma_{d/p}$(with normal $\underline{n}_{d/p}$) of the crack extension

$$\underline{\underline{\sigma}}\left(\underline{V}^{1}_{d/p}(y_1,y_2)\right).\underline{n}_{d/p}=-\underline{\underline{\sigma}}\left(\rho^{\lambda}\underline{u}(\theta)\right).\underline{n}_{d/p}$$

Equations for $\underline{V}^{2}_{d/p}$, $\underline{V}^{3}_{d/p}$ and $\underline{V}^{4}_{d/p}$ (see (14)) can be derived in the same way.

And then

$$\delta W=A_{d/p}k^2l^{2\lambda}+\cdots\quad\text{with } A_{d/p}=\Psi\left(\underline{V}^{1}_{d/p}(y_1,y_2),\rho^{\lambda}\underline{u}(\theta)\right)$$

The difficulty in the present analysis comes from the non homogeneous conditions due to the pressure p_{CO_2} acting on the main crack walls and on the crack extension and the poro-elastic behaviour in the reservoir triggering a swelling. Nevertheless these non homogeneous conditions are taken into account by the special terms (9) and (11), finally all calculations give

$$
\begin{aligned}
A_{d/p} &= \Psi\left(\underline{\hat{V}}^1_{d/p}(y_1, y_2), \rho^\lambda \underline{u}(\theta)\right)\\
B_{d/p} &= \Psi\left(\underline{\hat{V}}^2_{d/p}(y_1, y_2), \rho^\lambda \underline{u}(\theta)\right)\\
&\quad +\Psi\left(\underline{\hat{V}}^1_{d/p}(y_1, y_2), \rho \underline{t}(\theta)\right)\\
C_{d/p} &= \Psi\left(\underline{\hat{V}}^3_{d/p}(y_1, y_2), \rho^\lambda \underline{u}(\theta)\right)\\
&\quad +\Psi\left(\underline{\hat{V}}^1_{d/p}(y_1, y_2), \rho \underline{v}(\theta)\right)\\
&\quad +\frac{1}{2}\int_{\gamma_{d/p}} \underline{n}_{d/p}.\underline{\hat{V}}^1_{d/p}(y_1, y_2)\mathrm{d}S\\
D_{d/p} &= \Psi\left(\underline{\hat{V}}^2_{d/p}(y_1, y_2), \rho \underline{t}(\theta)\right)\\
E_{d/p} &= \Psi\left(\underline{\hat{V}}^3_{d/p}(y_1, y_2), \rho \underline{v}(\theta)\right)\\
&\quad +\frac{1}{2}\int_{\gamma_{d/p}} \underline{n}_{d/p}.\underline{\hat{V}}^3_{d/p}(y_1, y_2)\mathrm{d}S\\
F_{d/p} &= \Psi\left(\underline{\hat{V}}^2_{d/p}(y_1, y_2), \rho \underline{v}(\theta)\right)\\
&\quad +\Psi\left(\underline{\hat{V}}^3_{d/p}(y_1, y_2), \rho \underline{t}(\theta)\right)\\
&\quad +\frac{1}{2}\int_{\gamma_{d/p}} \underline{n}_{d/p}.\underline{\hat{V}}^2_{d/p}(y_1, y_2)\mathrm{d}S
\end{aligned}
$$

Of course, under the simplifying assumption that p_{CO_2} does not act on the crack extension faces, the integrals on $\gamma_{d/p}$ disappear.

References

Cook J, Gordon JE (1964) A mechanism for the control of crack propagation in all brittle systems. Proc R Soc 282A:508–520

Detournay E (1999) Fluid and solid singularities at the tip of a fluid-driven fracture. In: Durban D, Pearson JRA (eds) Proceedings of the IUTAM symposium on Non-linear singularities in deformation and flow. Kluwer, Dordrecht

GIEC (2005) Special report on carbon dioxide capture and storage: technical summary

He MY, Hutchinson JW (1989) Crack deflection at an interface between dissimilar elastic materials. Int J Solids Struct 25:1053–1067

Hutchinson J, Mear ME, Rice JR (1987) Crack paralleling an interface between dissimilar materials. J Appl Mech 54:828–832

Karnaeva E (2012) Amorçage et propagation des réseaux de fracture dans le context du stockage de CO_2: étanchéité des couvertures. PhD thesis, University Pierre and Marie Curie, Paris, France

Karnaeva E, Leguillon D, Baroni A, Putot C (2012) Fracture propagation in a multilayered caprock, associated with CO_2 injection. Submitted to Int. J. Rock Mech. and Min, Sci

Labossiere PEW, Dunn ML (1999) Stress intensities at interface corners in anisotropic bimaterials. Eng Fract Mech 62:555–575

Leguillon D, Sanchez-Palencia E (1987) Computation of singular solutions in elliptic problems and elasticity. Wiley, New York

Leguillon D, Lacroix C, Martin E (2000) Interface debonding ahead of a primary crack. J Mech Phys Solids 48:2137–2161

Leguillon D (2002) Strength or toughness ? A criterion for crack onset at a notch. Eur J Mech A/Solids 21:61–72

Leguillon D, Martin E (2012b) The strengthening effect caused by an elastic contrast—part I: the bimaterial case. Int J Fract 179:157–167

Leguillon D, Martin E (2012a) Crack nucleation at stress concentration points in composite materials—application to the crack deflection by an interface. To appear in Mathematical Methods and Models in Composites, Computational and Experimental Methods in Structures, V. Mantic Ed., Imperial College Press, London

Leguillon D, Sanchez-Palencia E (1992) Fracture in Heterogeneous materials, weak and strong singularities. In: Ladeveze P, Zienkiewicz OC (eds) New advances in computational structural mechanics. Studies in applied mechanics, 32 Elsevier, Amsterdam, pp 423–434

Loret B, Radi E (2001) The effect of inertia on crack growth in poroelastic fluid saturated media. J Mech Phys Solids 49:995–1020

Mishuris G, Wrobel M, Linkov A (2012) On modeling hydraulic fracture in proper variables: stiffness, accuracy, sensitivity. Int J Eng Sci 61:10–23

Segall P, Fitzgerald SD (1998) A note on induced stress changes in hydrocarbon and geothermal reservoirs. Techtonophysics 289:117–128

Terzaghi K (1936) The shearing resistance of saturated soils. In: Proceedings of the first international conference of soil mechanics and foundation engineering, Cambridge

Van Dyke M (1964) Perturbation methods in fluid mechanics. Academic Press, New York

Williams ML (1959) The stress around a fault or crack in dissimilar media. Bull Seismol Soc Am 49:199–204

DOI 10.1007/978-3-319-04397-5_10
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9841-4

ORIGINAL PAPER

Calibration of brittle fracture models by sharp indenters and inverse analysis

V. Buljak · G. Cocchetti · G. Maier

Received: 8 November 2012 / Accepted: 19 March 2013 / Published online: 3 April 2013

Abstract In several engineering areas structural analyses concern also fracture processes of brittle materials and employ cohesive crack models. Calibrations of such models, i.e. identification of their parameters by tests, computer simulations of the tests and inverse analyses, have been investigated in the literature particularly with reference to non-destructive indentation tests at various scales.To this timely research, the following contributions are presented in this paper: a simple piecewise-linear cohesive crack model is considered for brittle materials (here glass, for example); for its calibration by "non-destructive" indentation tests novel shapes are attributed to instrumented indenters, in order to make fracture the dominant feature of the specimen response to the test; such shapes are comparatively examined and optimized by sensitivity analyses; a procedure for inverse analysis is developed and computationally tested, based on penetration versus increasing force only (no imprint measurements by profilometers) and is made "economical" (i.e. computationally fast, "in situ" by small computers) by model reduction through proper orthogonal decomposition in view of repeated industrial applications.

V. Buljak
Department of Strength of Materials, Mechanical Engineering Faculty, University of Belgrade, Belgrade, Serbia

G. Cocchetti · G. Maier (✉)
Department of Structural Engineering, Technical University (Politecnico) of Milan, Milano, Italy
e-mail:maier@stru.polimi.it

Keywords Instrumented indentation · Inverse analysis · Calibration of brittle fracture models · Model reduction

1 Introduction

Indentation experiments, originally rooted in classical hardness tests, see e.g. Suresh and Giannakopoulos (1998), at present provide the experimental basis of a popular and still growing methodology for material characterizations and model calibrations centered on computer simulations of tests and inverse analyses. The parameters to identify are included in the constitutive model or/and in the residual stress tensor to be employed for direct analyses of structures or plant components. The main advantages arise from the "non-destructive" features of the test and from possible fast and economical, even often "in situ", estimations of parameters. Typical and representative contributions to developments of indentation testing methodology can be found in a vast literature, e.g. Abdul-Baqi and Van der Giessen (2002), Oliver and Pharr (2004), Cao and Lu (2004).

Several relatively recent research contributions in this area are centered on the following innovative features of the diagnostic methodology:

(a) substantial increase of experimental data with recourse to imprint geometry measurements (by laser profile-meter, atomic force microscope or other instruments, depending on the testing scale),

data which are additional to those provided by the instrumented indenter, see e.g. Bolzon et al. (2004), Bocciarelli et al. (2005), Maier et al. (2011);

(b) acceleration and economization of the inverse analysis procedure in practice by means of "a priori" once-for-all computational efforts for "model reduction", see e.g. Ostrowski et al. (2008), Buljak and Maier (2011), Buljak (2012).

In some practical engineering situations improvement (a) turns out to be almost prohibitive (e.g. in offshore engineering for applications under water) or clearly expensive for routine applications (e.g. repeated along an extensive network of pipelines). However, tensorial entities such as anisotropic elastic-plastic material models and residual stresses clearly cannot be identified on the basis of indentation curves generated by the popular axisymmetric (conical or spherical) or pyramidal indenters. Therefore, for the above purposes, indenters with "elliptical" conical shapes and test repetitions after rotations have been proposed and designed in Buljak and Maier (2012).

In this paper, a new procedure is envisaged and investigated for the calibration of isotropic fracture models on the basis only of a penetration-versus-force curve generated by an instrumented indenter endowed with a novel shape with the following features: a typical cross-section orthogonal to the indenter axis consists of two circular arches, so that in one of the symmetry axes two sharp edges ("cutting edges") are generated; the shape-defining lengths (called here "diameters") of the indenter cross-section grow from the bottom generating two diverse shapes, both with a sharp "cutting edge" in one of the two symmetry planes which contain the indenter axis. As an alternative, "double indenters" are devised and examined, with a single plane containing the two parallel axes and the "cutting edges".

The new geometric configurations of the indenter are suggested by the clear usefulness of conferring, in the present context, to the fracture propagation a prevailing role in the response of the specimen to the indentation test. Such prevalence, however, can be expected in brittle and quasi-brittle behaviors, to which the present study is confined, with further limitations to the simplest cohesive models. As for ductile fracture characterizations, the exploitation of imprint geometry data (provision (a) mentioned above) is expected to be practically necessary, see e.g. Maier et al. (2006) and (2011).

The developments investigated and proposed herein concerning brittle materials (glasses, ceramics "et alia") can be outlined as follows.

The novel geometry of the indenter pair is designed in such a way that the penetration in the tested specimen or structure generates first stress concentrations and then two cracks starting from the sharp internal edges and propagating along the "fracture plane" in the specimen or in the real structure, with "non-destructive" consequences in view of the small scale.

Digitalized indentation curves represent sources of experimental data to be selected by sensitivity analyses in preliminary design of the experiment after a suitable choice of the inelastic material models apt to describe brittle fracture.

For the inverse analyses the recently developed computational procedure, above mentioned as provision (b), is adopted herein, namely "proper orthogonal decomposition" followed by "radial basis functions" interpolations and by mathematical programming (or elsewhere "soft computing" algorithm) for the minimization of the "discrepancy function", here with a deterministic approach.

The simple cohesive crack models considered for the present preliminary study are described in Sect. 2. Finite element modeling of the experiment is discussed in Sect. 3. Section 4 and 5 are devoted to sensitivity analyses and to the design of indenters with novel shapes characterized by slenderness and by sharp "cutting edges". In Sects. 6 and 7 computational procedures are outlined apt to make the inverse analyses proposed herein more economical in practical applications. Some numerical exercises are comparatively presented in order to check the validity of the methodological novelties proposed herein and to support the conclusions gathered in Sect. 8.

2 Cohesive crack piecewise-linear models and relevant parameters

Overall analyses of structures consisting primarily of brittle materials (such as concrete, rock, ceramics, cement, glass) are usually carried out to practical engineering purposes with the following reasonable assumptions: fracture processes concentrated as growing displacement discontinuities, surrounded by undamaged, linear elastic, frequently isotropic media.

For computer simulations of structural behaviors with nonlinearities confined to the surface of fracture propagation "cohesive crack models" are frequently adopted in engineering practice.

These models, which include contact friction as a special case, describe a relationship between relative displacements and tractions in "mode I" or "mixed mode", relationship formulated with various features in the literature, see e.g. Xu and Needleman (1994), Jirasek and Bažant (2001), Broek (1989), Stavroulakis (2000), Zhi-He et al. (2003).

In the present study the piecewise-linear formulation of cohesive crack models, proposed in Maier (1970), Cocchetti et al. (2002), Tin-Loi and Que (2001) and (2002), is adopted because of the following reasons: (a) the mathematical description is a "linear complementarity problem" (LCP) not only in terms of rates, but also for finite steps and under holonomy (i.e. reversibility) hypothesis, which turns out to be acceptable and useful in the present context; (b) inverse analysis by discrepancy minimization may be tackled as "mathematical programming under equilibrium constraints" (MPEC), subject of vast literature; (c) if applied to non-hardening ("perfect") plasticity, piecewise linearization reduces limit analysis and shakedown analyses to classical linear programming, see e.g. Jirasek and Bažant (2001), Kaliszky (1989).

The most general piece-wise linear (PWL) cohesive crack model (CCM) reads (matrix notation, bold symbols for vectors and matrices; details in Cocchetti et al. 2002; Bolzon et al. 1994 and 2002):

$$\boldsymbol{\varphi} = \mathbf{N}^{\mathrm{T}}\mathbf{t} - (\boldsymbol{\varphi}_0 - \mathbf{H}\,\boldsymbol{\lambda})\,, \quad \mathbf{w} = \bar{\mathbf{N}}\,\boldsymbol{\lambda} \tag{1}$$

$$\boldsymbol{\varphi} \le \mathbf{0}, \qquad \dot{\boldsymbol{\lambda}} \ge \mathbf{0}, \qquad \boldsymbol{\varphi}^{T}\dot{\boldsymbol{\lambda}} = 0 \tag{2}$$

Reference axes 1, 2 and 3 are related to a surface as normal ("mode one") and tangential to it ("shear modes"), surface which is assumed "smooth" and on which tractions **t** are related through the model to possible relative displacements **w**.

The other quantities here involved have the following meanings rooted in classical plasticity theory: "yield functions" $\{\varphi_1 \ldots \varphi_m\}^{\mathrm{T}} = \boldsymbol{\varphi}$ defining, when zero, current yield planes, subsequent to the initial ones defined in **t** space by $\boldsymbol{\varphi}_0$ ("yield limits"); "flow coefficients" $\{\lambda_1 \ldots \lambda_m\}^{\mathrm{T}} = \boldsymbol{\lambda}$ governing irreversible "plastic flow" by their rates (derivatives, marked by dots, with respect to ordinative, non-physical, time); matrices **N** and $\bar{\mathbf{N}}$ contain the unit vectors which define, respectively, orientations of the yield planes and orientations of the relevant contributions by yielding to displacements **w**; $\bar{\mathbf{N}} = \mathbf{N}$ implies "associativity"; matrix **H** governs hardening which here is merely softening (therefore in Eq. (1) a minus precedes **H**).

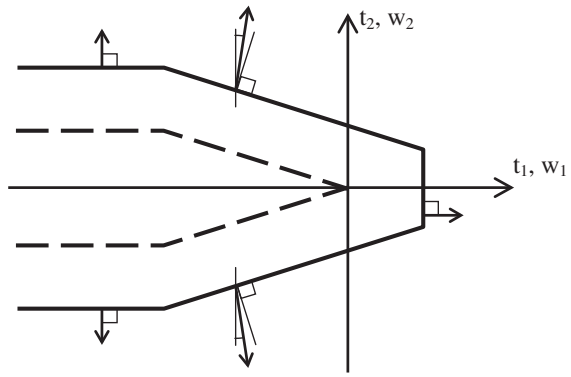

Fig. 1 Mixed mode cohesive crack model with piecewise linearization (*dashed lines* define the final yield domain after softening processes)

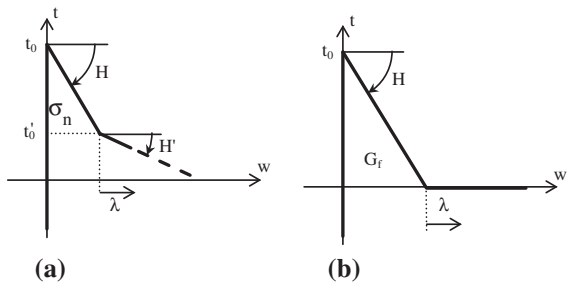

Fig. 2 Piecewise linearized models of mode I cohesive crack fracture

Details on the above general PWL-CCM and on its potentialities and limitations with respect to other fracture models can be found in Cocchetti et al. (2002), Tin-Loi and Que (2002). An illustrative special case for mixed-mode brittle fracture is shown in Fig. 1, where: m = 5 yielding modes due to symmetry assumed with respect to normal axis 1; two non-zero hardening coefficients, one concerning the yield plane normal to axis 1, the second governing softening on the other 4 external yield planes. It is worth noting that the "final" perfectly plastic behavior described by dashed lines can be interpreted as Coulomb friction.

To the present purposes of a preliminary methodological study concerning novel indentation tests of brittle fracture and subsequent inverse analyses, the following further particularizations are assumed in the above PWL-CCM context, with reference to Fig. 2.

mode-I fracture (no sliding, hence "associativity" in plasticity jargon); softening branch approximated by a sequence of linear steps, Fig. 2a, but in what follows by a single straight segment as shown in Fig. 2b (namely with H'=0 and $t_0' = 0$); holonomy (i.e. reversibility) assumed in view of the present use of the opening phase alone. Therefore, the simple fracturing model of Fig. 2b to consider in the sequel, as particularization of Eqs. (1) and (2), reads (primed symbols concern the "detachment stage"):

$$\varphi = t + H\,(w - \lambda) - t_0 \le 0, \quad w \ge 0, \quad \varphi\, w = 0 \tag{3}$$

$$\varphi' = w - \lambda - \frac{t_0}{H} \le 0, \quad \lambda \ge 0, \quad \varphi'\,\lambda = 0 \tag{4}$$

The parameters to identify for the calibration of the above simple model of brittle fracture are merely two: tensile strength t_0 and softening coefficient H; alternatively, t_0 and fracture energy G_f shown in Fig. 2b. Correlation is expected between these main parameters and the others in a general PWL model of brittle fracture.

The present holonomy assumption (reversibility of deformation path) is justified by the following circumstance: experimental data to exploit are extracted from penetration curves only, since the descending branch provided by instrumented indenters may differ from the loading branch only because of possible fragments in the crack and/or possible inelastic strains; both events are not considered here in the computer simulations since marginal as for mechanical consequences in the present context of brittle materials.

In what follows the material to characterize as for fracture behavior is assumed isotropic linear elastic, with Young modulus E and Poisson ratio ν "a priori" known, identified by usual tension tests. The contact between the indenter and the indented material will be described as Coulomb friction without dilatancy, so that another parameter to identify in practical applications may be the friction coefficient.

3 Test modeling

The constitutive models for brittle materials considered in what precedes contain (at least two) parameters in cohesive crack model to be estimated by inverse analysis resting on non-destructive economical indentation tests. Clearly, crack generation and propagation should be central and dominant events in such testing procedure to design. Therefore, advantageous prospects arise for a sharp indenter, a sort of blade with edges, quite different from the traditional and popular indenters.

Assessments of residual stresses by indentation alone have already led to propose indenter shapes without axial symmetry with respect to the tip axis ("elliptical indenters", Buljak and Maier 2012). The shape optimization to the present purposes will be performed in the next Section. Basic features of the proposed experiments and of their simulations are selected here below.

According to an usual praxis in present experimental mechanics, the "instrumented" indenter provides in digital form the "indentation curve", namely the force F imposed on the tip and the consequent tip penetration u into the specimen (with surface suitably polished according to pertinent codes). The indenter tip material considered herein is usual diamond, isotropic elastic with moduli $E^I = 1170\,\text{GPa}$ and $\nu^I = 0.1$.

However, rigidity might be attributed to the indenter, not only here for the comparative numerical exercises performed in order to optimize its shape, but also in practical applications to materials like glass.

As for the simplest cohesive crack model (mode I, PWL with 2 parameters) specified in Sect. 2, Eqs. (3), (4), glass is considered here with the following "search intervals" conjectured by an "expert" (see e.g. Le Bourhis 2007; McGee 2006):

$$\begin{aligned} &20\,\text{MPa} \le t_0 \le 100\,\text{MPa}, \\ &20\,\text{MPa}/\mu\text{m} \le H \le 80\,\text{MPa}/\mu\text{m} \end{aligned} \tag{5}$$

The averages of the above lower and upper bounds are considered in what follows as "reference values" of the parameters to estimate, namely:

$$\bar{t}_0 = 50\ \text{MPa}, \quad \bar{H} = 50\,\text{MPa}/\mu\text{m} \tag{6}$$

The cohesive model, Eqs. (3), (4), with the above parameters is attributed at the beginning of each simulation to the whole symmetry plane area within which the crack propagation is expected to extend during the test loading phase. The Young modulus and Poisson ratio, are assumed here as "a priori" known, and read: $E = 70\,\text{GPa}$, $\nu = 0.25$, respectively.

On the growing interface between indenter and specimen, to the coefficient f of Coulomb friction without dilatancy the following "a priori" conjectured lower and upper bound and their mean value are attributed here:

$$0.01 \le f_0 \le 0.35, \qquad \bar{f}_0 = 0.15 \tag{7}$$

For computer simulations of tests, the following provisions are adopted herein: finite element (FE) modeling

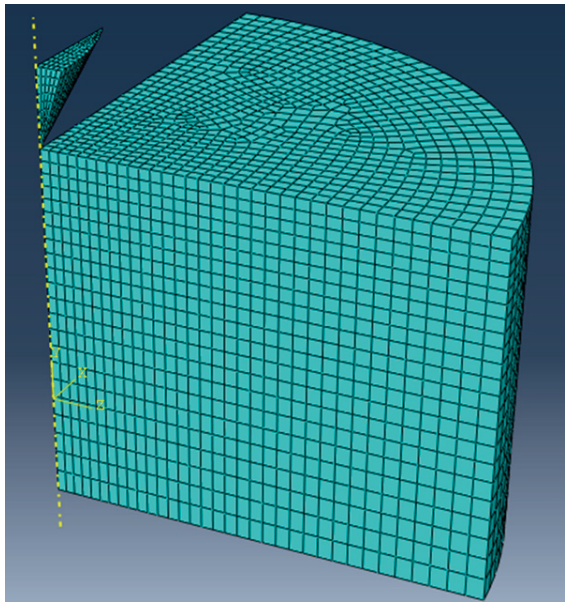

Fig. 3 Adopted mesh for test modeling with single indenter

by a commercial code (Abaqus 2010); boundary conditions with zero displacements (alternatively, "infinite elements") in an area considered unperturbed by the test in the specimen; exploitation of two orthogonal planes of symmetry consistent with the expected symmetries of the imprint (parallel to indenter central longitudinal axis); large strain context.

The peculiar idealization of the test (a crack localized in a pre-selected plane normal to the specimen surface) would make possible, in a small-strain context, the once-for-all "condensation" of the elastic domain surrounding the crack: a provision, not adopted here, which may alleviate the computational burden of repeated simulations.

Other data which quantify the main features of the numerical exercises in what follows are listed here: 8-node hexahedron elements for continuum in specimen and indenter, 8 nodes cohesive elements to model fracturing surface; about 60,000 degrees of freedom (relatively small changes with diverse geometries considered for indenter optimization). A typical adopted FE mesh is visualized in Fig. 3 by exploiting the two symmetries preserved in all indenter shapes.

4 New indenter shapes for brittle fracture tests

The novel geometrical features of indenters considered here for the calibration of brittle fracture models are

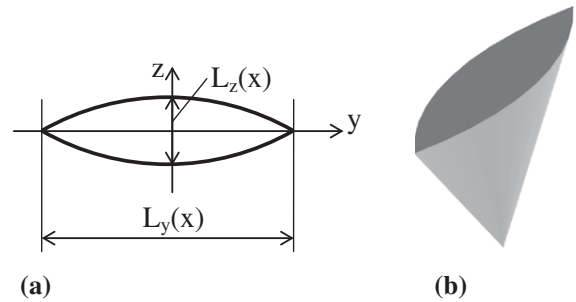

Fig. 4 "Bi-circular conical" (BCC) indenter. **a** Cross section; **b** overall shape

presented below. The common purpose pursued is the presence of sharp edges as a provision contributing to confer dominant role to crack generation and propagation.

(A) The simple basic geometry of popular conical indenters (like those for Rockwell hardness tests, see e.g. Oliver and Pharr 2004; Chen et al. 2007) can be described by:

$$D(x) = 2x \tan\alpha \tag{8}$$

where: x is an abscissa along the cone axis, with origin on the tip vertex V (smoothing not to be considered here); α is half of the opening angle; D is the diameter of the cross-section circle. The new geometry considered here is visualized in Fig. 4: along the two orthogonal axes y and z, by preserving symmetry with respect to xy and xz planes, two lengths L_Y and L_Z are selected as maximal and minimal "diameters" of the new section defined by circular arches ending at sharp edges, as shown in Fig. 4a.

A second parameter, either β or "thickness" T, is considered for the governance of the new geometry, namely:

$$L_Y(x) = \beta \cdot D(x),\ L_Z(x) = \frac{1}{\beta} D(x),\ T = \frac{L_Z}{L_Y} = \frac{1}{\beta^2} \tag{9}$$

The indenter defined by Eqs. (8) and (9) are called henceforth "bi-circular conical" (BCC); clearly, it preserves cross-section areas and the vertex tip ($x = 0$) of the original cone (which is recovered for $T = 1$).

(B) An indenter geometry alternative to the preceding one is considered now.

The denomination proposed is "bi-circular blade" (BCB) indenter. Its generation can be described as follows, with reference to Figs. 5a and 6. A fictitious circular cone described by Eq. (8) is considered as for its

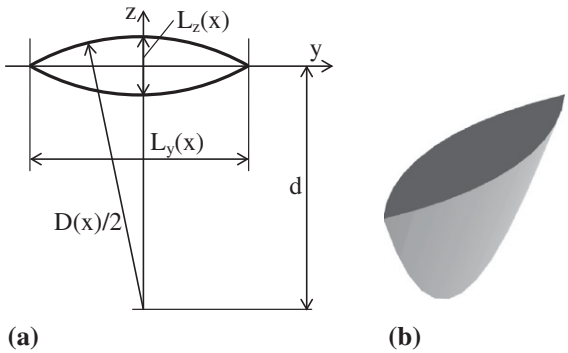

Fig. 5 Cross section (**a**) and overall shape (**b**) of a sharp "bi-circular blade" (BCB) indenter

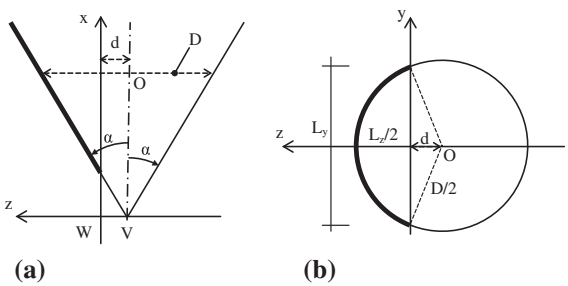

Fig. 6 Geometry of indenter BCB. **a** Section in the plane of symmetry for both the original fictitious cone and new indenter; **b** cross section

intersection with a plane parallel to its axis at a distance d from it, with the role of symmetry plane of the new indenter.

The cross section is therefore defined by circular arches and exhibits the following "diameters":

$$L_Z(x) = D(x) - 2d,$$
$$L_Y(x) = \left[2D(x) \cdot L_Z(x) - L_Z^2(x)\right]^{1/2} \tag{10}$$

The above geometric configuration is governed by two parameters: α and d. Clearly, the origin from the fictitious cone, Eq. (8), with vertex V (at x = 0), implies a surfaces consisting of straight lines through V.

Now the "thickness" T of the cross-section changes with the longitudinal coordinate x:

$$T(x) = \frac{L_Z(x)}{L_Y(x)} = \sqrt{\frac{D(x) - 2\mathrm{d}}{D(x) + 2\mathrm{d}}} \tag{11}$$

Figure 6 clarifies the generation of this indenter geometry which exhibits no singular point as vertex and a sharp edge represented for z = 0 by the equation:

$$(z + \mathrm{d})^2 + y^2 = \left[\frac{D(x)}{2}\right]^2, \quad z = 0 \tag{12}$$

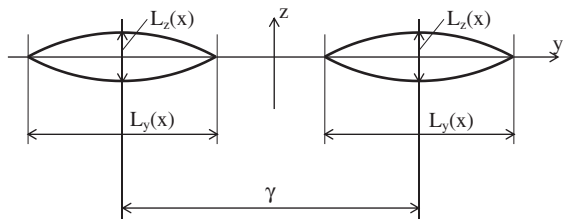

Fig. 7 Cross section either of a "double bi-circular conical" (DBCC) indenter or of a "bi-circular blade indenter" (DBCB)

The sharp edge line in the symmetry plane containing the axis and the maximal diameter L_y is smooth (no longer bi-linearly straight like in BCC indenters and now without tip as vertex), like in a sort of blade or knife, see Fig. 5b.

(C) Configurations of more complex and costly indenters might be suggested by the advantages for fracture parameters estimation expected from increased role or dominance of crack-propagation promoted in the indentation experiment. Therefore, it may be of practical interest the association in a single indentation instrument of either two indenters BCC or two BCB described in what precedes. The denominations adopted include "double", namely DBCC and DBCB indenters, respectively. Figure 7 evidences the new geometry requiring as third governing parameter the distance γ between the axes. The two symmetries with respect to planes *xy* and *xz* are preserved.

The two novel indenter shapes specified in Figs. 4 and 5 are further illustrated in Fig. 8, which visualizes the two planes of symmetry and the FE meshes adopted in the sequel (with features specified in Sect. 3).

5 Sensitivity analyses for indenter shape design

Sensitivity analysis in general means assessment of the influence which the sought parameters have on quantities measurable in the experiment, see e.g. Kleiber et al. (1997).

Clearly, such influence must be quantified through simulations of the test in order to select the quantities to measure and to design the experimental procedure. Crucial is here the latter purpose, pursued in what follows by two approaches, both using the FE discretizations specified in Sect. 3 and visualized in Fig. 4.

(i) For the present innovation, in order to select new indenter geometries, sensitivity analyses may provide checks of compliance with the following

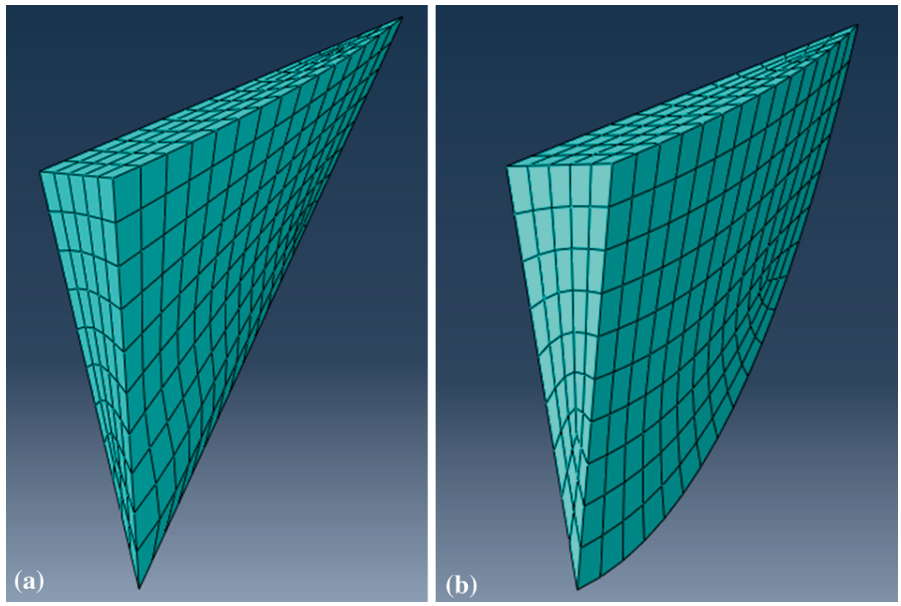

Fig. 8 Novel shape of indenters (one fourth in view of the two symmetries): **a** "bi-circular conical"; **b** "bi-circular blade" and relevant FE mesh for test simulations

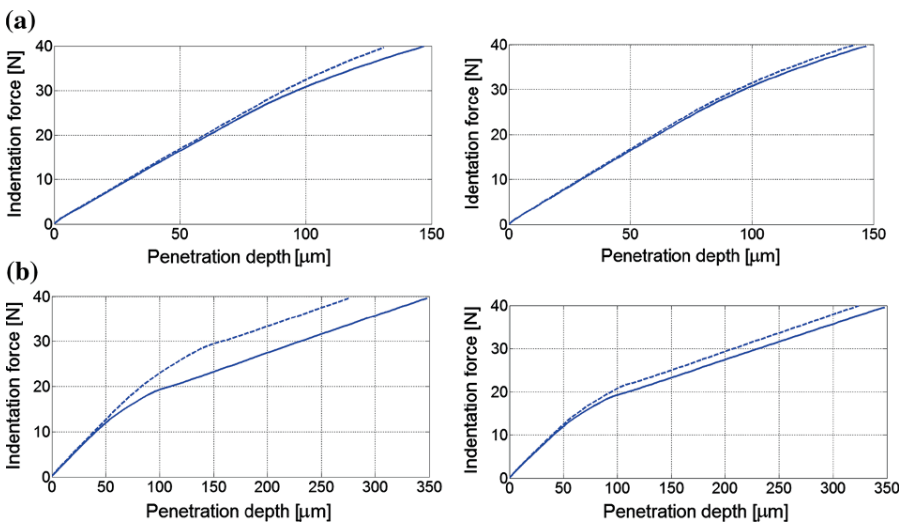

Fig. 9 Loading indentation curves generated by BCC indenter with $\alpha = 16°$ and **a** with thickness $T = 0.4$ and **b** with $T = 0.2$: on the *left* for H at reference value and tensile strength t_0 at conjectured *upper bound* (*dashed line*) and *lower bound* (*solid line*); on the *right*, analogously, for t_0 at reference value and H at *upper bound* (*dashed line*) and *lower bound* (*solid line*)

requirement: the differences between measurable quantity values achieved when the parameters are at the extremes of the intervals conjectured preliminarily "by an expert" to localize the search, should be larger (say by two orders of magnitude) than expected experimental errors. While a conjectured extreme value is attributed to one parameter, reference values are assigned to the other parameters. In the present preliminary shape design, preferable turns out to be the indenter with higher sensitivity in the above sense. Clearly, at this stage of investigation experimental errors cannot be realistically quantified, but only assumed to comparison purposes.

Figures 9, 10, 11 and 12 visualize some of the computational exercises carried out in the present orientative study. Such comparative exercises led to select as preliminarily "optimal" indenter a DBCC with the geometry parameters $\alpha = 16°, T = 0.2$ (namely $\beta = 2.236$) and $\gamma = 0.5$mm.

The above outlined computations lead to the following conclusions: the main parameters in simple cohesive crack models significantly influence the inden-

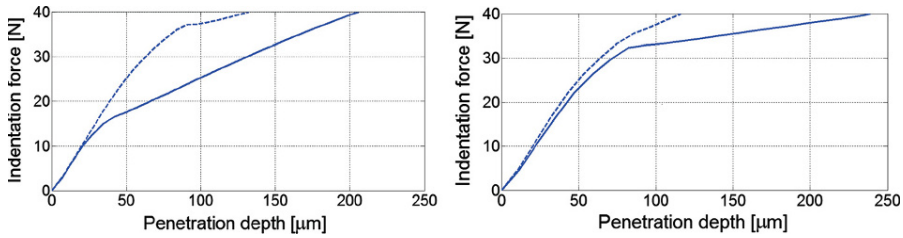

Fig. 10 Diagrams with same meaning as those in Fig. 9, but here generated through simulations of tests by doubling with $\gamma = 0.5$mm the same indenter (i.e. by a DBCC), and by applying the same final load level (namely 40 N) on the double indenter

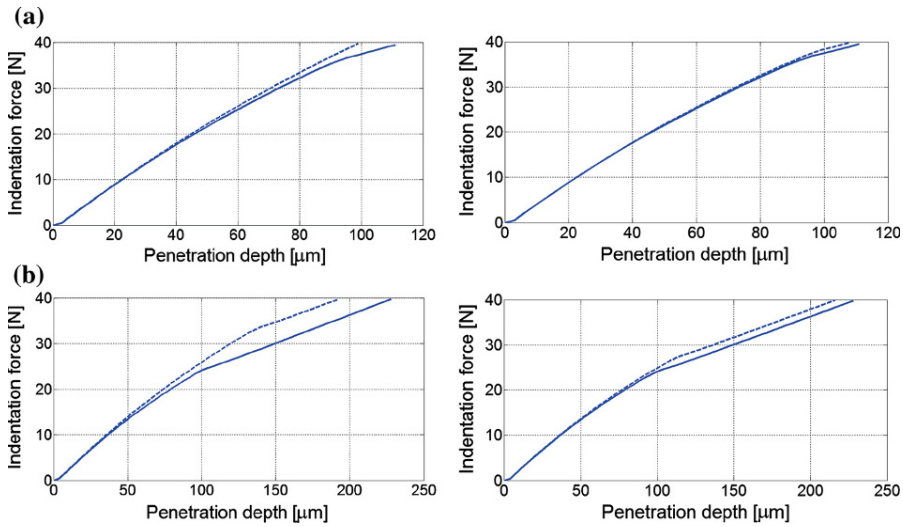

Fig. 11 Pseudo–experimental results comparable to those in Fig. 9 (concerning a BCC), here achieved by simulation of indentation with same loading but by a BCB indenter with geometry governed again by $\alpha = 16°$ and by d = 0.26 mm and d = 0.6 mm

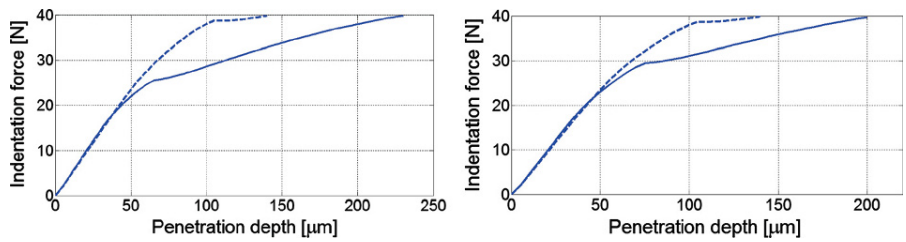

Fig. 12 Pseudo–experimental results by a DBCB: comparable to those in Fig. 10 (concerning DBCC indenter), but here generated by doubling with $\gamma = 0.5$ mm the same indenter and by doubling the final load level

tation curve, generated by bi-circular conical (BCC) indenter, more with lower thickness T and more with double pins. However, such conclusions obviously should be combined with technological and stability requirements of the indenter production and routine employments. Moreover, the above numerical results have no general quantitative meaning, since they depend on a particular choice of materials (here a kind of glass).

(ii) The traditional sensitivity analysis in inverse problems, see e.g. Kleiber et al. (1997), has a meaning different from that above adopted at (i). Specifically, this meaning is provided through the derivative of a typical measurable quantity with respect to a sought parameter, by attributing conjectured reference values to all the other parameters to estimate. Such usual sensitivity analysis quantifies the influence of sought parameters on measurable quantities independently

Table 1 Sensitivities of penetration measures under various indentation loads F

F [N]	5	10	15	20	25
$t_0/u_{\max} du/dt_0$	0.004	0.016	0.043	0.232	0.329
$H/u_{\max} du/dH$	0.000	0.005	0.031	0.095	0.149

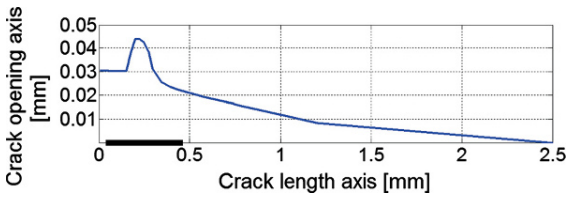

Fig. 13 Geometric configuration (account taken of symmetries) of the crack on the specimen surface at the end of the penetration by the DBCC indenter with thickness $T = 0.2$; diameter $L_y = 0.46$mm is visualized by a thick segment on the abscissae axis

from accuracies of expert's conjectures and of experimental data.

By considering the above selected DBCC indenter, sensitivities in the above sense have been computed (by forward finite-difference approximation with increments of 5 % with respect to reference values of the two parameters) for 5 penetration depths corresponding to 5 equal increase stages of indenter load. The normalized results are collected in Table 1, which evidences the growth of sensitivity with tip penetration, as obviously expected. The normalization is performed by the maximum of the penetration ($u_{\max}$) and by the reference values of the two sought parameters (tensile strength t_0 and negative hardening or softening H).

The FE simulation of the fracture test, above considered as an example of DBCC indenter employment, has led to the crack shown in Fig. 13. The geometry at the end of penetration (corresponding to the loading branch top of the indentation curves) is characterized by the following numerical results: depth of the indenter pin amounts to 0.22 mm; depth of the crack to ~ 0.5 mm; crack length and central opening on the surface: $L \cong 5$ mm and $l \cong 0.06$ mm, respectively.

6 Inverse analysis procedures

The approach adopted in what follows for the parameter identification is traditional and popular in practice, see e.g. Bui (1994), Buljak (2012). It is centered on the minimization with respect to the sought parameters (here vector $\mathbf{p} = \{t_0, H\}^{\mathrm{T}}$) of the "discrepancy function" ω, discrepancy between experimental data (vector $\bar{\mathbf{u}}$, here of "pseudo-experimental data") and their counterparts $\mathbf{u}$ computed by test simulations as functions of the variable parameter vector $\mathbf{p}$:

$$\omega(\mathbf{p}) = \left[\mathbf{u}(\mathbf{p}) - \bar{\mathbf{u}}\right]^T \mathbf{C}^{-1} \left[\mathbf{u}(\mathbf{p}) - \bar{\mathbf{u}}\right],$$
$$\min_{\mathbf{p}} \{\omega(\mathbf{p})\} = \omega(\hat{\mathbf{p}}) \tag{13}$$

The covariance matrix $\mathbf{C}$, which quantifies the measurement inaccuracy, is assumed unit in this preliminary deterministic investigation. For numerical illustrative example, a double bi-circular conical (DBCC) with thickness $T = 0.2$ is considered; with maximum $F_{\max} = 20\text{N} + 20\text{N} = 40\text{N}$ of the applicable force. Vector $\bar{\mathbf{u}}$ contains here 50 "pseudo-experimental data", namely, 50 penetration depths corresponding to 50 levels of force at equal intervals over the loading part of the indentation curve, starting from zero up to the above maximum force.

The numerical quantities (concerning materials and FE mesh) assumed in what precedes, and the averages of the bounds conjectured for the model parameters to identify, are now considered as input for test simulations (direct analysis) apt to provide vector $\bar{\mathbf{u}}$ of the measurable response quantities, here confined to upward indentation curve (penetration phase).

The values, say $\bar{\mathbf{p}}$, attributed to the sought parameters for the direct analysis leading to $\bar{\mathbf{u}}$, if compared to the results $\hat{\mathbf{p}}$ of inverse analyses based on $\bar{\mathbf{u}}$, obviously lead to checks of validity and accuracy of the estimation procedure.

The relationship $\mathbf{u}(\mathbf{p})$ has been formulated in what precedes as a nonlinear complementarity problem (NLCP). Actually, in each FE simulation step, the linear complementarity problems (LCP) of the cohesive crack model, Eqs. (3), (4), and of Coulomb friction model, are associated with linear elasticity, and with equilibrium and compatibility equations. These equations are nonlinear in view of large strains considered herein according to their implementation in the adopted commercial code Abaqus (2010). Therefore, problem (13) turns out to be a "mathematical programming under equilibrium constraints" (MPEC) in the jargon of classical mathematics originally oriented to economy, i.e. an optimization with non-convex and non-smooth constraints and with possible lack of convexity of the objective function $\omega(\mathbf{p})$, see e.g. Cottle et al. (1992), Luo et al. (1996), Facchinei and Pang (2003).

For the minimization of the discrepancy functions, a ("first order" iterative) Trust Region Algorithm (TRA) is employed to the present purposes by using its implementation in Matlab (2009). Descriptions of TRA are widely available in the literature, e.g. Conn et al. (2000), and are omitted here. The attainment of the absolute minimum is here checked by initialization multiplicity.

The double sharp bi-circular conical indenter (DBCC) described in Sects. 4 and 5, and in Figs. 10 and 13, is considered here again with geometry defined by the same parameters, namely $\alpha = 16°, T = 0.2$ (hence $\beta = 2.236$) and $\gamma = 0.5$ mm. With reference to the present example, Fig. 14a visualizes by a map the discrepancy function $\omega(\mathbf{p})$ over the search domain and the step sequences starting from two different initializations; the convergence in one of them is shown in Fig. 14b. The initial and final values of the objective function ω turn out to be 1.1022×10^6 and 0.069, respectively, in one minimization; 4.2359×10^6 and 0.0325 in the other minimization process.

The following convergence criterion was adopted for the end of the iterative procedure: the optimization is terminated either when the change of the objective function from one iteration to another is less than 0.01 or when the norm of the step length (i.e. vector of parameter changes from one iteration to another) is smaller than 0.001 (with parameters normalized about reference values in this case).

Similar numerical exercises have corroborated the prospect, arising in various practical applications, to consider Coulomb friction coefficient as the third parameter to estimate. One of several results encouraging the present method is visualized in Fig. 15 with TRA convergence plots for estimation of strength t_0, softening H and friction f on the basis of the same indenter shapes BCC and DBCC and relevant tests considered earlier. As a verification of the above inverse analysis accuracy, the following numerical results related to Fig. 15 are cited here as example: after 11 TRA steps for one case and 8 steps for the other, estimates of t_0, H and f exhibit the following errors in percentage: 0.27, 0.9 and 0.31 %, respectively, from BCC test; 0.001, 0.02 and 0.001 %, respectively, from DBCC test.

The comparison is limited to the identification of the three parameters t_0, H and f by the same algorithm TRA with same initializations which are visualized in Fig. 15.

The comparative computational checks (and others not presented here for brevity) of the novel indentation-based inverse analysis proposed herein lead to the following remarks. On the basis of the loading vs. penetration curve alone provided by instrumented indenters, simple cohesive crack models for brittle fracture can be calibrated by the novel sharp shapes envisaged here, with priority to be attributed to double indenters.

7 Model reduction

Computer simulations of indentation tests are usually expensive, and rather high numbers of simulations are required in order to solve numerically the problem of discrepancy function minimization, Eq. (13). Such circumstance is crucial for repeated frequent applications of inverse analyses in industrial contexts, and concerns practically all approaches and computational methods for parameter identification (mathematical programming like TRA adopted here, genetic algorithms, artificial neural networks).

A remedy frequently adopted at present consists of "model reduction" by "proper orthogonal decomposition" (POD). Details on this methodology can be found in, e.g., Buljak (2012), Buljak and Maier (2011) and will not be presented here.

As anticipated in the Introduction, the sequence of operative stages proposed in the present context can be outlined as follows, with reference to the test by the double indenter considered in what precedes DBCC with $\alpha = 16°, T = 0.2$ (hence $\beta = 2.236$), $\gamma = 0.5$ mm (see Sect. 4) and to the estimation by it of two parameters alone (t_0 and H) for clearer graphical descriptions of some operative stages.

(a) In the "search domain" suggested by an expert and specified in Sect. 3, a grid of N "nodes" is generated, here with $N = 9 \times 9 = 81$. The grid generation is here performed by equi-intervals in both axis directions. For higher number of parameters to estimate it would be advantageous in practical applications to a priori select the number of nodes and then to optimize their collocations to form the grid over the search domain.

(b) Assuming the parameters of each node $\mathbf{p}_i$, with $i = 1 \ldots N$, the $M = 50$ penetrations depths as measurable quantities (corresponding to equi-intervalled indenter loads) are computed by FE test

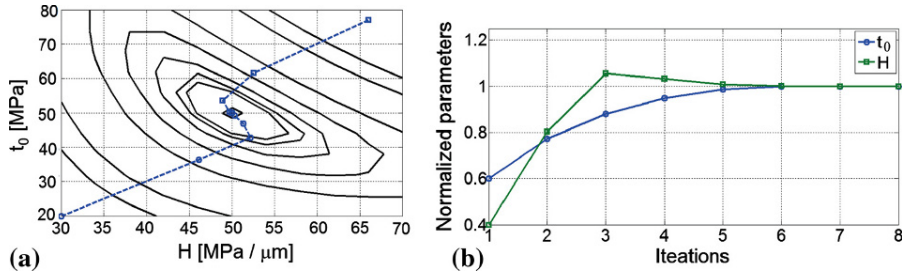

Fig. 14 With reference to the same DBCC indenter as in Figs. 10 and 13. **a** Parameters search domain with a map of the discrepancy function and two sequences of TRA steps. **b** Convergence of the TRA procedure from one of the two initializations

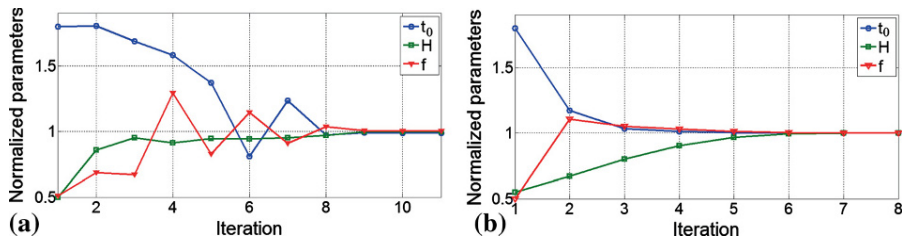

Fig. 15 Convergence of TRA, from the same initializations, for the identification of friction coefficient f besides brittle fracture parameters t_0 and H; indentation curve provided **a** by a single pin indenter BCC, and **b** by a double indenter (DBCC)

simulations and are gathered in vector $\mathbf{u}_i$. All these "snapshots" are collected in $M \times N$ (here 50×81) matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]$.

(c) The eigenvalues and eigenvectors are computed for the (positive semi-definite) matrix $\mathbf{D} = \mathbf{U}^T\mathbf{U}$, here of order 81. According to the POD theory (see e.g. Buljak (2012)) such computations are central to the generation of new "basis" $\boldsymbol{\phi}$ by exploiting the "correlation" physically motivated by the common origin of all the considered "snapshots" of the pseudo-experimental data. The eigenvalues of $\mathbf{D}$ provide guidance to "truncation", namely to removal of axes with negligible components and thus to the generation of a new basis $\bar{\boldsymbol{\phi}}$, with $\bar{N} << N$ dimensions, $\bar{N} = 5$ in the present context.

Thus, a controllable approximation of any test response snapshot $\mathbf{u}_i (i = 1 \ldots N)$ can be computed by a linear transformation of an "amplitude" vector $\mathbf{a}_i$ with much lesser dimensionality, here reduced to 5.

(d) Let $\mathbf{p}$ be one parameter vector among many within the sequence of steps implied by TRA for the solution of the optimization, Eq. (13). For any $\mathbf{p}$ of such variable vectors, the corresponding vector $\mathbf{u}$ of measurable quantities is needed to assess the consequent value of the objective function ω. If the results of the POD are available, interpolation among the amplitudes $\mathbf{a}_i (i = 1 \ldots N)$ leads to $\mathbf{a}(\mathbf{p})$, namely to the amplitude $\bar{N}$-vector which defines, with controllable approximation, in the original basis the corresponding $\bar{N}$-vector $\mathbf{u}$, "snapshot" of the response to the test if the parameter were $\mathbf{p}$. The interpolations can be efficiently performed by "radial basis functions" (RBF).

The above outlined POD+RBF+TRA procedure has been applied frequently in recent times (e.g. Ostrowski et al. 2008; Buljak and Maier 2011; Buljak 2012) because of the earlier mentioned computational advantages, which are remarkable in practical applications, if compared to TRA (or other optimization algorithms) involving test simulations by FE in a long sequence. Such advantages are evidenced in this study by the following typical assumptions and results of comparative exercises.

For the indentation, same DBCC indenter is employed as in the preceding Sections; same test modeling; average computing time for a single FE simulation is about 45 minutes by a computer with processor Pentium i5, and the amount of RAM memory is 4GB; total computing time up to TRA convergence amounts to about 24 hours.

The adopted RBFs read, with the value 1 attributed to the factor r according to advice in Buhmann (2003):

$$g_i(\mathbf{p}) = \frac{1}{\sqrt{\|\mathbf{p} - \mathbf{p}_i\|^2 + r^2}} \tag{14}$$

Other numerical features adopted in the present examples are specified in what follows. Equi-interval grid of

$N = 81$ nodes already specified; truncation in eigenvalues of matrix $\mathbf{D} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$: once the eigenvalues have been ordered in terms of decreasing values, the neglected eigenvalues are those for which their sum, starting from the smallest one, is less than 10^{-6} of the sum of the whole set of eigenvalues; the average computing time to assess snapshot $\mathbf{u}$ corresponding to parameters $\mathbf{p}$ amounts to: 0.003 sec.; computing time for the solution of the inverse problem by POD + RBF + TRA: 0.1sec.; average errors of the estimates resulting from TRA: 0.01 % on t_0, 0.04 % for H; errors of the estimates achieved by POD + RBF + TRA: 0.001 % on t_0, 0.06 % for H.

The robustness of the present inverse analysis procedure has been checked with positive results by stochastic numerical exercises like the following one. A random perturbation of $\pm 2\,\%$ with uniform probability density has been induced into the inputs $\bar{\mathbf{u}}$ of pseudo-experimental data; as a consequence, mean values and standard deviation of the assessed parameters corresponding to 10 different inverse analysis with randomly extracted noise of the above mention level turned out to be $t_0 = 49.97 \pm 0.13$ MPa and $H = 50.04 \pm 0.23$ MPa/μm instead of reference values specified in Eq. (6). The error of the assessed parameters by the POD+RBF+TRA cohesive crack model calibration turned out to be less than 1 % for the noise level equal to $\pm 2\,\%$.

An alternative assessment of consequences due to measurement inaccuracy may be based on attribution of same measurement error (not in percentage) to all experimental data concerning penetration depths during the loading process. The errors attributed as an example to such data are specified in the first row of Table 2. Such errors have led to the percentage changes indicated there (second and third row) as for the estimates of strength t_0 and hardening H. The parameter H turns out to be affected by larger errors if compared to strength. Clearly, such percentages might be reduced by larger depth of indentation.

Furthermore, it is worth noting that the experimental errors assumed in this numerical tests are more than the resolution of displacement measurements by the most advanced modern instrumented indenters operating on the force scale used in present numerical examples.

The "a priori" evaluation of the Young modulus may be affected by some errors which, quantified in percentage as examples in Table 3 first row, have caused the percentage errors pointed out there.

Table 2 Influence of instrumental errors on measurable quantities u_i on the fracture estimates

Δu_i[μm]	+2 (%)	+5 (%)	−2 (%)	−5 (%)
Δt_0 [%]	1.9	5.8	1.5	2.8
ΔH [%]	3.7	9.1	3.3	6.3

Table 3 Consequences on the parameter estimates due to possible errors in the preceding assessment of the Young modulus

ΔE[%]	+5 (%)	+10 (%)	−5 (%)	−10 (%)
Δt_0[%]	1.61	2.95	1.73	3.61
ΔH[%]	5.11	9.02	4.51	9.95

The above numerical checks contribute to the conclusion that the inverse problem is stable and well posed and that the influence of elastic properties uncertainty turns out to be quite limited. The latter circumstance is due to the specific design of the indenter. Unlike classical, still widely employed indenters, the novel sharp shape presented here reduces the elastic energy involved in the indentation process, with respect to the one related to the fracture.

Similar perturbations in percentage of the "a priori" Poisson ratio have affected about an order-of-magnitude less the two-fracture parameters resulting from the inverse analyses, as exemplified in Table 3.

In the present investigation of possible fracture parameter estimation by novel kinds of indenters and inverse analysis, the dimensionality of the unknown parameters space has been kept rather small, also in order to easily visualize comparative results. With possibly larger number of parameters to identify, the trivial approach to subdivide each conjectured parameter interval in equal sub-intervals may become a practical difficulty, since the number of nodes over the search domain obviously grows exponentially with the number of parameters to estimate. An alternative approach to grid generation might be based on Latin hypercube sampling, McKay et al. (1979). It provides the advantage that the nodes number N is chosen "a priori", as useful in practice.

8 Closing remarks

The results of the study outlined in what precedes leads to the following conclusions.

(a) Brittle materials can be mechanically characterized as for the main parameters governing simple fracture models by means of indentation tests on the basis of experimental data extracted merely from the loading phase of the indentation curves.
(b) The shape of the indenter should be apt to make brittle fracture as dominant feature of the specimen response to indentation tests. Indenters with sharp edges, particularly if double in the same instrument, turn out to satisfy such requirement provided that their geometry is optimized with respect to some parameters governing it.
(c) Production of indenters with shapes not available on the market might impose some constraints not considered in this preliminary investigation. Another limitation of the present results is their confinement to the simplest "mode-I" piecewise-linearized cohesive crack model, with only two parameters to identify (or three if friction coefficient is considered as unknown) and with isotropic elasticity assumed as "a priori" known. Further investigations are desirable by "pseudo-experimental" comparative approaches in order to overcome the above limitations.
(d) Model reduction by proper orthogonal decomposition as proposed herein turns out to be crucial in order to make computing efforts consistent with economical constraints in routine industrial applications either in laboratory or "in situ".
(e) Dimensions of indentation imprints (and consequently sizes of indenters and entities of maximum loads) has to satisfy a compromise between two main requirements: as small as possible to avoid damages by "in situ" testing; much larger than typical lengths in the material nanoscale structure. The almost microscale dimensions selected in the present study clearly might turn out to be inadequate in various applications, but the main features of the method are likely to represent an improvement of the current practices concerning brittle structures diagnoses.
(f) If anisotropy is expected in the brittle fracture properties of the material, the non-destructive diagnostic method presented herein based on sharp instrumented indenters only (i.e. without profilometers) might be generalized, according to research in progress, by recourse to the following main provisions: some mixed-mode piecewise-linear cohesive crack model like the one visualized in Fig. 1, in order to preserve the mathematical formulation and procedure adopted herein; repetitions of the indentation tests on the vertices of a small triangle containing the superficial place under examination, by turning 45° the major diameters of the indenter cross-sections. Such procedure has been adopted in Buljak and Maier (2011) for identification of residual stress tensors (as an economical alternative to the procedure proposed in Bocciarelli and Maier 2007); this is the subject of current research on calibrations of fracture models in anisotropic media.
(g) Clearly, stochastic approaches, e.g. such that adopted in Bolzon et al. (2002) for diverse experiments, will represent a further step ahead in the present methodological context of inverse analyses.

Acknowledgments The results presented in this paper are related to research projects on structural diagnoses by non-destructive tests in collaboration with VeTec Co., Marghera (Venezia, Italy), and in the framework of the European KMM-VIN coordinated by the Polish Academy of Sciences. This study is also related to the research project no 35006 supported by Serbian Ministry of Science and Technological Development. The authors acknowledge with thanks the supports of the above institutions to the research which led to the results here summarized, and the useful information provided by Professor Luigi Biolzi on glass materials.

References

ABAQUS/Standard (2010) Theory and user's manuals, release 6.10, Dassault Systèmes Simulia Corp., Providence, RI, USA

Abdul-Baqi A, Van der Giessen E (2002) Numerical analysis of indentation-induced cracking of brittle coatings on ductile substrates. Int J Solids Struct 39:1427–1442. doi:10.1016/S0020-7683(01)00280-3

Bocciarelli M, Bolzon G, Maier G (2005) Parameter identification in anisotropic elastoplasticity by indentation and imprint mapping. Mech Mater 37:855–868. doi:10.1016/j.mechmat.2004.09.001

Bocciarelli M, Maier G (2007) Indentation and imprint mapping method for identification of residual stresses. Comput Mater Sci 39:381–392. doi:10.1016/j.commatsci.2006.07.001

Bolzon G, Fedele R, Maier G (2002) Parameter identification of a cohesive crack model by Kalman filter. Comput Methods Appl Mech Eng 191:2847–2871. doi:10.1016/S0045-7825(02)00223-2

Bolzon G, Maier G, Panico M (2004) Material model calibration by indentation, imprint mapping and inverse analysis. Int J Solids Struct 41:2957–2975. doi:10.1016/j.ijsolstr.2004.01.025

Bolzon G, Maier G, Novati G (1994) Some aspects of quasi-brittle fracture analysis as a linear complementarity problem. In: Bažant ZP, Bittnar Z, Jirásek M, Mazars J (eds) Fracture

and damage in quasi-brittle structures. E and FN Spon, London, pp 159–174
Broek D (1989) The practical use of fracture mechanics. Kluwer, London
Buhmann MD (2003) Radial basis functions. Cambridge University Press, Cambridge
Bui HD (1994) Inverse problems in the mechanics of materials: an introduction. CRC Press, Boca Raton
Buljak V (2012) Inverse analysis with model reduction— Proper Orthogonal Decomposition in structural mechanics. Springer, Berlin. doi:10.1007/978-3-642-22703-5
Buljak V, Maier G (2011) Proper orthogonal decomposition and radial basis functions in material characterization based on instrumented indentation. Eng Struct 33:492–501. doi:10.1016/j.engstruct.2010.11.006
Buljak V, Maier G (2012) Identification of residual stresses by instrumented elliptical indentation and inverse analysis. Mech Res Commun 41:21–29. doi:10.1016/j.mechrescom.2012.02.002. (Corrigendum in Mechanics Research Communications 46,p. 90, 2012, doi:10.1016/j.mechrescom.2012.07.005
Cao YP, Lu J (2004) A new method to extract the plastic properties of metal materials from an instrumented spherical indentation loading curve. Acta Materialia 52:4023–4032. doi:10.1016/j.actamat.2004.05.018
Chen X, Ogasawara N, Zhao M, Chiba N (2007) On the uniqueness of measuring elastoplastic properties from indentation: the indistinguishable mystical materials. J Mech Phys Solids 55:1618–1660. doi:10.1016/j.jmps.2007.01.010
Cocchetti G, Maier G, Shen X (2002) Piecewise-linear models for interfaces and mixed mode cohesive cracks. Comput Methods Eng Sci 3:279–298. doi:10.3970/cmes.2002.003.279
Conn AR, Gould NIM, Toint PL (2000) Trust-region methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
Cottle RW, Pang JS, Stone RE (1992) The linear complementarity problem. Academic Press, Boston
Facchinei F, Pang JS (2003) Finite-dimensional variational inequalities and complementarity problems, vol II. Springer Series in Operations Research and Financial Engineering, Berlin
Jirasek M, Bažant ZP (2001) Inelastic analysis of structures. Wiley, New York
Kaliszky S (1989) Plasticity—theory and engineering applications. Elsevier, Amsterdam
Kleiber M, Antunez H, Hien TD, Kowalczyk P (1997) Parameter sensitivity in nonlinear mechanics. Theory and finite element computations. Wiley, New York
Le Bourhis E (2007) Glass structure. In: Mechanics and technology, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi:10.1002/9783527617029.ch5
Luo ZQ, Pang JS, Ralph D (1996) Mathematical programs with equilibrium constraints. Cambridge University Press, Cambridge
Maier G (1970) A matrix structural theory of piecewise-linear plasticity with interacting yield planes. Meccanica 5:55–66. doi:10.1007/BF02133524
Maier G, Bocciarelli M, Bolzon G, Fedele R (2006) Inverse analyses in fracture mechanics. Int J Fract 138:47–73. doi:10.1007/s10704-006-7153-7
Maier G, Bolzon G, Buljak V, Miller B (2011) Comparative assessment of three different approaches in material parameter characterization based on instrumented indentation, using only the indentation curve, only the residual imprint, and both experimental data. Inverse Probl Sci Eng 19:815–837. doi:10.1080/17415977.2011.551931
Matlab (2009). Matlab user's guide, The Math Works Inc., Natick, MA. www.mathworks.it
McGee T (2006) Glass manufacturing engineering. Wiley, New York
McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables from a computer code. Technometrics 21:239–245. doi:10.1080/00401706.1979.10489755
Oliver WC, Pharr GM (2004) Measurement of hardness and elastic modulus by instrumental indentation: advances in understanding and refinements to methodology. J Mater Res 19:3–20. doi:10.1557/jmr.2004.19.1.3
Ostrowski Z, Bialecki RA, Kassab AJ (2008) Solving inverse heat conduction problems using trained POD-RBF network. Inverse Probl Sci Eng 16:39–54. doi:10.1080/17415970701198290
Stavroulakis GE (2000) Inverse and crack identification problems in engineering mechanics. Kluwer, Dordrecht
Suresh S, Giannakopoulos AE (1998) A new method for estimating residual stresses by instrumented sharp indentation. Acta Materialia 46:5755–5767. doi:10.1016/S1359-6454(98)00226-2
Tin-Loi F, Que NS (2001) Parameter identification of quasi-brittle materials as a mathematical program with equilibrium constraints. Comput Methods Appl Mech Eng 190:5819–5836. doi:10.1016/S0045-7825(01)00199-2
Tin-Loi F, Que NS (2002) Nonlinear programming approaches for an inverse problem in quasi-brittle fracture. Int J Mech Sci 44:843–858. doi:10.1016/S0020-7403(02)00035-8
Xu XP, Needleman A (1994) Numerical simulation of fast crack growth in brittle solids. J Mech Phys Solids 42:1397–1434. doi:10.1016/0022-5096(94)90003-5
Zhi-He J, Glaucio H, Paulino RH, Dodds J (2003) Cohesive fracture modeling of elastic-plastic crack growth in functionally graded materials. Eng Fract Mech 70:1885–1912. doi:10.1016/S0013-7944(03)00130-9

DOI 10.1007/978-3-319-04397-5_11
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9846-z

Statistics of ductile fracture surfaces: the effect of material parameters

Laurent Ponson · Yuanyuan Cao · Elisabeth Bouchaud · Viggo Tvergaard · Alan Needleman

Received: 21 January 2013 / Accepted: 26 March 2013 / Published online: 9 April 2013

Abstract The effect of material parameters on the statistics of fracture surfaces is analyzed under small scale yielding conditions. Three dimensional calculations of ductile crack growth under mode I plane strain, small scale yielding conditions are carried out using an elastic-viscoplastic constitutive relation for a progressively cavitating plastic solid with two populations of void nucleating second phase particles represented. Large particles that result in void nucleation at an early stage are modeled discretely while small particles that require large strains to nucleate are homogeneously distributed. The three dimensional analysis permits modeling of a three dimensional material microstructure and of the resulting three dimensional stress and deformation states that develop in the fracture process region. Material parameters characterizing void nucleation are varied and the statistics of the resulting fracture surfaces is investigated. All the fracture surfaces are found to be self-affine over a size range of about two orders of magnitude with a very similar roughness exponent of 0.56 ± 0.03. In contrast, the full statistics of the fracture surfaces is found to be more sensitive to the material microscopic fracture properties: height fluctuations are shown to crossover from a Student's distribution with power law tails at small scales to a Gaussian behavior at large scales, but this transition occurs at a material dependent length scale. Using the family of Student's distributions, this transition can be described introducing an additional exponent $\mu = 0.15 \pm 0.02$, the value of which compares well with recent experimental findings. The description of the roughness distribution used here gives a more complete quantitative characterization of the fracture surface morphology which allows a better comparison with experimental data and an easier interpretation of the roughness properties in terms of microscopic failure mechanisms.

Keywords Fracture surfaces · Roughness statistics · Ductile fracture · Crack growth · Scaling behavior · Finite elements

L. Ponson · Y. Cao
Institut Jean le Rond d'Alembert (UMR 7190), CNRS, Université Pierre et Marie Curie, Paris, France

E. Bouchaud
CEA-Saclay and ESPCI, Paris Tech, Paris, France

V. Tvergaard
Department of Mechanical Engineering, The Technical University of Denmark, Lyngby, Denmark

A. Needleman (✉)
Department of Materials Science and Engineering, University of North Texas, Denton, TX, USA
e-mail:needle@unt.edu

1 Introduction

Mandelbrot et al. (1984) first characterized the self-affine scaling properties of fracture surfaces. A hope in this study was to provide a quantitative correlation between fracture surface roughness and a material's crack growth resistance. This hope has yet to be

realized. Nevertheless, Mandelbrot et al. (1984) stimulated a large body of experimental and theoretical work aimed at characterizing the scaling properties of fracture surfaces. Experimentally, the self-affine nature of the roughness of fracture surfaces has been observed in a wide variety of materials (metals, ceramics, glasses, rocks…) under a wide variety of loading conditions (quasi-static, dynamic, fatigue) and is typically characterized by two exponents: one measured parallel to the crack front, denoted by ζ, and one measured along the crack propagation direction, denoted by β, see Ponson (2007), Alava et al. (2006), Bonamy and Bouchaud (2011) for overviews. In some studies, e.g. Bouchaud et al. (1990), Måløy et al. (1992), it was suggested that the values of these exponents are universal.

Others have maintained that the characterization of the roughness of fracture surfaces is more complex. For example, in Bonamy and Bouchaud (2011), Bonamy et al. (2006) it is argued that there are two roughness regimes: one pertaining to length scales smaller than the fracture process zone and the other to length scales larger than the fracture process zone, with each regime characterized by different values of the fracture surface roughness exponents. Alternatively, a multifractal characterization of fracture surface roughness has been suggested as discussed in Cherepanov et al. (1995). The extent to which the appropriate characterization of fracture surface roughness depends on the material microstructure and/or the fracture mechanism remains an open question.

A variety of theoretical analyses of fracture surface roughness have been carried out using a linear elastic fracture mechanics framework, e.g. Ramanathan et al. 1997, or a molecular dynamics framework, e.g. Nakano et al. (1995). For ductile fracture of structural metals at room temperature the governing mechanism is the nucleation, growth and coalescence of micron scale voids, which involves large plastic deformations, and occurs over size and time scales much larger than currently accessible by molecular dynamics. This process governs the evolution of ductile fracture surface roughness. The importance of accounting for porosity evolution and the accompanying plastic deformation in modeling fracture surface roughness is emphasized in Bouchaud (2003), Bouchbinder et al. (2004).

Calculations of fracture surface roughness using a finite deformation continuum mechanics formulation for a progressively cavitating solid were carried out in Needleman et al. (2012). The analyses in Needleman et al. (2012) were based on a modified Gurson (1975) constitutive relation for a porous viscoplastic solid with two types of void nucleating particles modeled: large particles that nucleate voids at an early stage and smaller particles that nucleate voids at a later stage. The larger particles were represented as uniform sized discrete "islands" of void nucleation, thus introducing a material length scale, while the smaller particles were uniformly distributed. This framework naturally accounts for the effects of damage evolution on the stress and deformation state in the fracture process zone.

In Needleman et al. (2012), the calculations were carried out for fixed material properties but for four random distributions of the larger particles. Here, we employ the same theoretical framework as in Needleman et al. (2012) and solve the same small scale yielding problem. We focus attention on one of the four spatial distributions of the larger particles in Needleman et al. (2012) and vary two material parameters that characterize void nucleation. We also use a more robust procedure to define the fracture surface geometry from the output of our simulations and investigate the roughness statistical properties for the cases calculated in Needleman et al. (2012) as well as for the simulations presented here. We go beyond the characterization of fracture surfaces by their correlation function and the value of the roughness exponent, and investigate the scaling of the full distribution of roughness. This scaling can be described by a second exponent characterizing the transition from power law tail at small scales to Gaussian roughness statistics at large scales. This quantitative characterization of fracture surface geometry allows a better comparison with experimental data and an easier interpretation of the roughness properties in terms of microscopic failure mechanisms.

2 Problem formulation

The boundary value problem analyzed is identical to that in Needleman et al. (2012). Only parameters characterizing void nucleation differ. For completeness, the formulation and constitutive relation are briefly stated here.

The focus is on quasi-static crack growth but dynamic analyses are carried out for numerical reasons. The calculations are carried out using a Lagrangian, convected coordinate formulation and the dynamic

principle of virtual work written as

$$\int_V \tau^{ij}\delta E_{ij}\,dV = \int_S T^i \delta u_i\,dS - \int_V \rho\frac{\partial^2 u^i}{\partial t^2}\delta u_i\,dV \quad (1)$$

All field quantities are taken to be functions of the convected coordinates, y^i, and time, t; E_{ij} are the covariant components of the Lagrangian strain tensor, T^i are the contravariant components of the traction vector, τ^{ij} are the contravariant components of Kirchhoff stress on the deformed convected coordinate net ($\boldsymbol{\tau} = J\boldsymbol{\sigma}$, with $\boldsymbol{\sigma}$ being the Cauchy or true stress and J the ratio of current to reference volume), ν_j and u_j are the covariant components of the reference surface normal and displacement vectors, respectively, ρ is the mass density, V and S are the volume and surface of the body in the reference configuration, and $(\,)_{,i}$ denotes covariant differentiation in the reference (y^1, y^2, y^3) Cartesian frame. In presenting the results we will use the notations x, y and z for y^1, y^2 and y^3, respectively.

A slice of material orthogonal to the initial crack plane is analyzed and the quasi-static mode I isotropic elastic singular displacement field is imposed on the remote boundaries. Also imposed are symmetry conditions corresponding to an overall plane strain constraint.

As in Needleman et al. (2012), the in-plane block dimensions are $h_x = h_y = 0.4\,\text{m}$ with an initial crack tip with an opening of $b_0 = 1.875 \times 10^{-4}\,\text{m}$. The finite element mesh consists of 428, 256 twenty node brick elements giving 1, 868, 230 nodes and 5, 604, 690 degrees of freedom. Ten uniformly spaced elements are used through the thickness h_z of 0.005 m, with 10 elements through the thickness, and a uniform 208×64 in-plane mesh is used in a $0.02\,\text{m} \times 0.006\,\text{m}$ region immediately in front of the initial crack tip giving an in-plane element size of $9.62 \times 10^{-5}\,\text{m}$ by $9.38 \times 10^{-5}\,\text{m}$.

2.1 Constitutive relation

The constitutive framework is the modified Gurson constitutive relation with

$$\mathbf{d} = \mathbf{d}^e + \mathbf{d}^\Theta + \mathbf{d}^p \quad (2)$$

and

$$\mathbf{d}^e = \mathbf{L}^{-1} : \hat{\boldsymbol{\sigma}} \quad (3)$$

$$\mathbf{d}^\Theta = \alpha\dot{\Theta}\mathbf{I} \quad (4)$$

$$\mathbf{d}^p = \left[\frac{(1-f)\bar{\sigma}\dot{\bar{\epsilon}}}{\boldsymbol{\sigma} : \frac{\partial\Phi}{\partial\boldsymbol{\sigma}}}\right]\frac{\partial\Phi}{\partial\boldsymbol{\sigma}} \quad (5)$$

Small elastic strains are assumed, $\mathbf{L}$ is the tensor of isotropic elastic moduli, $\hat{\boldsymbol{\sigma}}$ is the Jaumann rate of Cauchy stress and Θ is the temperature.

Adiabatic conditions are assumed so that

$$\rho c_p\frac{\partial\Theta}{\partial t} = \chi\boldsymbol{\tau} : \mathbf{d}^p \quad (6)$$

with $\rho = 7,600\,\text{kg/m}^3 = 7.6 \times 10^{-3}\,\text{MPa/(m/sec)}^2$, $c_p = 465\,\text{J/(kg°K)}$, $\chi = 0.9$ and α in Eq. (4) is 1×10^{-5}/K.

The flow potential is (Gurson 1975),

$$\Phi = \frac{\sigma_e^2}{\bar{\sigma}^2} + 2q_1 f^* \cosh\left(\frac{3q_2\sigma_h}{2\bar{\sigma}}\right) - 1 - \left(q_1 f^*\right)^2 = 0 \quad (7)$$

where $q_1 = 1.25$, $q_2 = 1.0$ are parameters introduced in Tvergaard (1981), Tvergaard (1982a), f is the void volume fraction, $\bar{\sigma}$ is the matrix flow strength, and

$$\sigma_e^2 = \frac{3}{2}\boldsymbol{\sigma}' : \boldsymbol{\sigma}', \quad \sigma_h = \frac{1}{3}\boldsymbol{\sigma} : \mathbf{I}, \quad \boldsymbol{\sigma}' = \boldsymbol{\sigma} - \sigma_h\mathbf{I} \quad (8)$$

The function f^*, introduced in Tvergaard and Needleman (1984), is given by

$$f^* = \begin{cases} f & f < f_c \\ f_c + (1/q_1 - f_c)(f - f_c)/(f_f - f_c) & f \geq f \end{cases} \quad (9)$$

where the values $f_c = 0.12$ and $f_f = 0.25$ are used.

The matrix plastic strain rate, $\dot{\bar{\epsilon}}$, is taken as

$$\dot{\bar{\epsilon}} = \dot{\epsilon}_0\left[\frac{\bar{\sigma}}{g(\bar{\epsilon}, \Theta)}\right]^{1/m},$$
$$g(\bar{\epsilon}, \Theta) = \sigma_0 G(\Theta)\,[1 + \bar{\epsilon}/\epsilon_0]^N \quad (10)$$

with $\bar{\epsilon} = \int\dot{\bar{\epsilon}}dt$, $E = 70\,\text{GPa}$, $\nu = 0.3$, $\sigma_0 = 300\,\text{MPa}$ ($\epsilon_0 = \sigma_0/E = 0.00429$), $N = 0.1$, $m = 0.01$ and $\dot{\epsilon}_0 = 10^3$/s.

The function defining the temperature-dependence of the flow strength is

$$G(\Theta) = 1 + b_G\exp(-c[\Theta_0 - 273]) \times \left[\exp(-c[\Theta - \Theta_0]) - 1\right] \quad (11)$$

with $b_G = 0.1406$ and $c = 0.00793$/K. In (11), Θ and Θ_0 are in K and $\Theta_0 = 293$K. Also, the initial temperature is taken to be uniform and 293K.

The initial void volume fraction is taken to be zero and the evolution of the void volume fraction is governed by

$$\dot{f} = (1 - f)\mathbf{d}^p : \mathbf{I} + \dot{f}_{nucl} \tag{12}$$

where the first term on the right hand side of Eq. (12) accounts for void growth and the second term for void nucleation.

Eight point Gaussian integration is used in each twenty node element for integrating the internal force contributions. Lumped masses are used and twenty-seven point Gaussian integration is used for the element mass matrix. The discretized equations are integrated using the explicit Newmark β-method ($\beta = 0$), Belytschko et al. (1976). The constitutive updating is based on the rate tangent modulus method of Peirce et al. (1984) while material failure is implemented via the element vanish technique in Tvergaard (1982b).

2.2 Inclusions and fracture properties

The calculations model a material with two populations of void nucleating second phase particles: (1) uniformly distributed small particles that are modeled by plastic strain controlled nucleation and (2) large, low strength inclusions that are modeled as "islands" of stress controlled nucleation. In each case, void nucleation is assumed to be described by a normal distribution as in Chu and Needleman (1980).

For plastic strain nucleation,

$$\dot{f}_{nucl} = D\dot{\bar{\epsilon}}, \quad D = \frac{f_N}{s_N\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{\bar{\epsilon} - \epsilon_N}{s_N}\Big)^2\Big] \tag{13}$$

The parameters $f_N = 0.04$ and $s_N = 0.1$ are fixed.

For stress controlled nucleation

$$\dot{f}_{nucl} = A\left[\dot{\bar{\sigma}} + \dot{\sigma}_h\right] \tag{14}$$

with

$$A = \frac{f_N}{s_N\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{\bar{\sigma} + \sigma_h - \sigma_N}{s_N}\Big)^2\Big] \tag{15}$$

if $(\bar{\sigma} + \sigma_h)$ is at its maximum over the deformation history. Otherwise $A = 0$.

We confine attention to a single inclusion distribution that is the distribution labeled 411 in Needleman et al. (2012). There are 2016 possible inclusions (mean spacing about 6.7×10^{-4} m) in the uniform mesh region. Each inclusion radius is $r_0 = 1.5 \times 10^{-4}$ m. For an inclusion governed by stress nucleation centered at (y_0^1, y_0^2, y_0^3), the value of f_N in Eq. (15) at the point (y^1, y^2, y^3) is

$$f_N = \begin{cases} \bar{f}_N & \text{for } \sqrt{(y^1 - y_0^1)^2 + (y^2 - y_0^2)^2 + (y^3 - y_0^3)^2} \le r_0; \\ 0 & \text{for } \sqrt{(y^1 - y_0^1)^2 + (y^2 - y_0^2)^2 + (y^3 - y_0^3)^2} > r_0 \end{cases} \tag{16}$$

The values $\bar{f}_N = 0.04$ and $s_N/\sigma_0 = 0.2$ are fixed.

In the following, materials having various values of parameters characterizing void nucleation are investigated (Fig. 1). For all materials analyzed the inclusion distribution is that termed case411 in Needleman et al. (2012):

- Material # 1 has the parameter values analyzed in Needleman et al. (2012) which are $\epsilon_N = 0.3$ and $\sigma_N/\sigma_0 = 1.5$. The fracture surfaces for this case are calculated using the procedure introduced here for defining the roughness geometry and the statistical geometry of this case is investigated in more detail than in Needleman et al. (2012).
- Material # 2 has the nucleation strain of the small particles changed from the reference value to $\epsilon_N = 0.4$ with the nucleation stress of the inclusions fixed at $\sigma_N/\sigma_0 = 1.5$.
- Material #2x has the nucleation strain of the small particles changed from the reference value to $\epsilon_N = 0.2$ with the nucleation stress of the inclusions fixed at $\sigma_N/\sigma_0 = 1.5$. This calculation terminated after an amount of crack growth that was too small to provide a fracture surface that could be characterized statistically in the same manner as for the other cases.
- Material # 3 has the nucleation stress of the inclusions changed from the reference value to $\sigma_N/\sigma_0 = 2.0$ with the nucleation strain of the small particles fixed at $\epsilon_N = 0.3$.

The results of the calculations are regarded as modeling quasi-static response via dynamic relaxation. The results are reported as for a quasi-static solution; for example, the absolute magnitude of geometric dimensions do not matter; only geometric ratios matter. In the following, the results are presented as for a quasi-static solution.

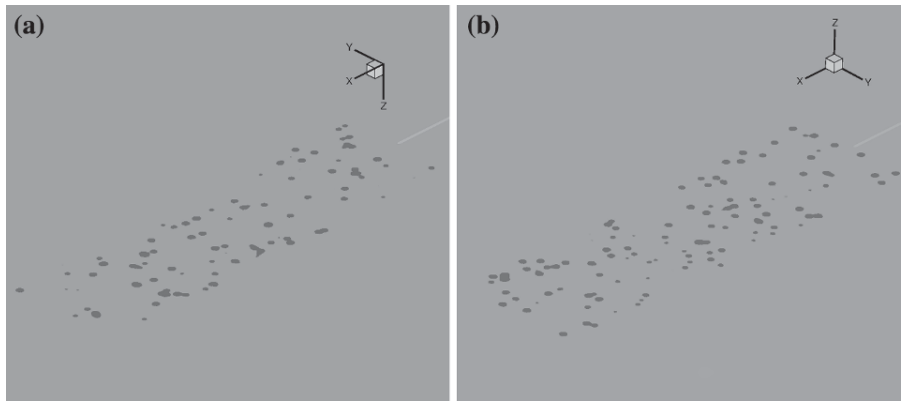

Fig. 1 Initial inclusion distribution. **a** On $z = 0$. **b** On $z = h_z$. Note that the positive $y-$axis is in opposite directions in **a** and **b**

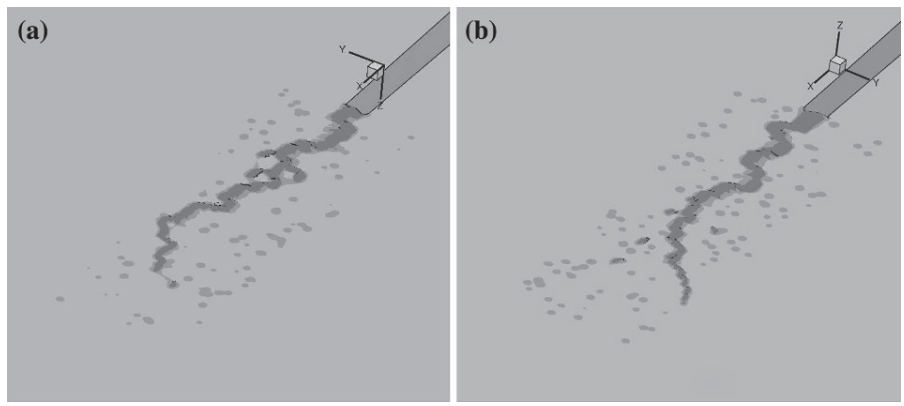

Fig. 2 Void volume fraction distribution showing the mode of crack growth for Material#2 ($\epsilon_N = 0.4$, $\sigma_N/\sigma_0 = 1.5$). **a** On $z = 0$. **b** On $z = h_z$. Note that the positive $y-$axis is in opposite directions in **a** and **b**

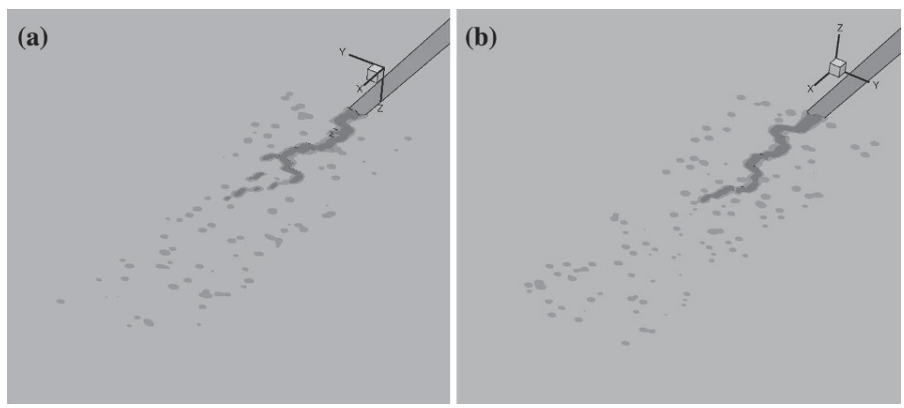

Fig. 3 Void volume fraction distribution showing the mode of crack growth for Material #2x ($\epsilon_N = 0.2$, $\sigma_N/\sigma_0 = 1.5$). **a** On $z = 0$. **b** On $z = h_z$. Note that the positive $y-$axis is in opposite directions in **a** and **b**

3 Results

3.1 Crack growth

Figures 2, 3 and 4 show distributions of void volume fraction f that indicate the crack growth path on the planes $z = 0$ and $z = h_z$ for Material #2, Material #2x and Material #3, respectively. In each case, the stage of loading corresponds to the last stage of loading computed for that case. Note that in (a) for all three figures the $y-$axis is positive to the left whereas in (b) for all three figures the $y-$axis is positive to the right. The dark gray regions show where $f \geq 0.1$ so that the material within those regions has essentially lost all stress

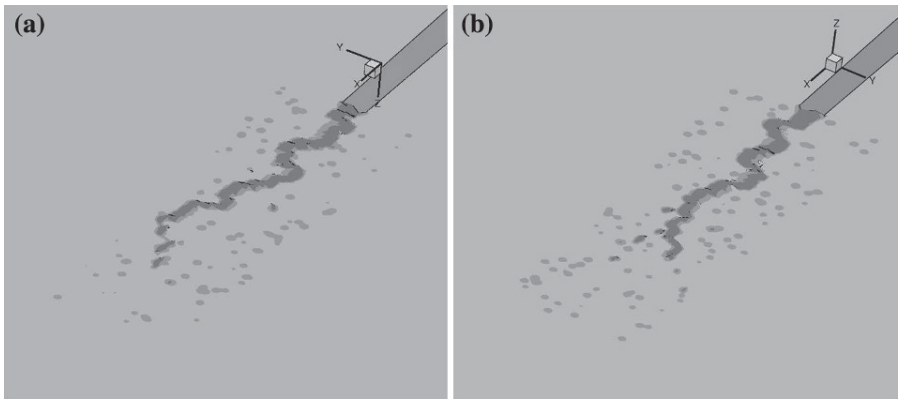

Fig. 4 Void volume fraction distribution showing the mode of crack growth for Material #3 ($\epsilon_N = 0.3$, $\sigma_N/\sigma_0 = 2.0$). **a** On $z = 0$. **b** On $z = h_z$. Note that the positive $y-$axis is in opposite directions in **a** and **b**

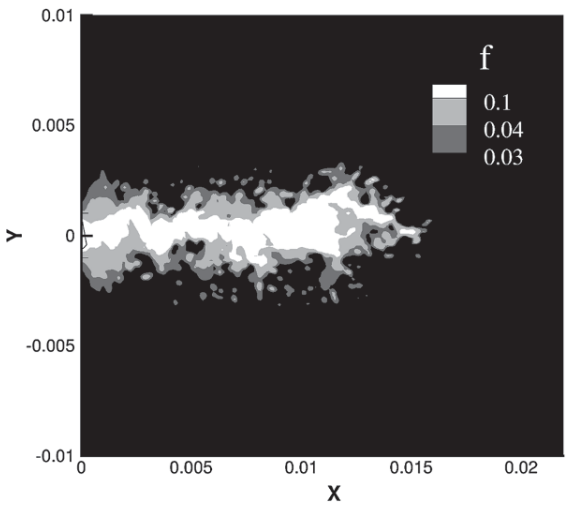

Fig. 5 Distributions of void volume fraction f in a plane $z =$ constant for material # 3 illustrating the definition of fracture surfaces based on void volume fraction threshold values of $f_t = 0.03,\ 0.04$ and 0.10

carrying capacity. The lighter gray regions which correspond to $0.1 \geq f \geq 0.01$ show the inclusions on the two planes that have nucleated voids. Shear localization can play a significant role in linking voids to the main crack. For Material #2x with $\epsilon_N = 0.2$, the calculation was terminated after a relatively small amount of crack growth because intense localized deformation led to the stable time step becoming very small.

3.2 Fracture surface definition

The first step in defining the fracture surface is to plot the void volume fraction f distribution on a cross section $z =$ constant. Next a threshold value f_t is chosen so that $f \geq f_t$ corresponds to a connected crack with f_t being the largest value that gives a connected crack over a predefined region. Figure 5 shows a typical cross section with contour values of $f = 0.03,\ 0.04$ and 0.10. The fracture surfaces for these three contour values are rather close. Changing the size of the cracked region modifies the value of f_t needed for a connected crack. Interestingly, the scaling behavior, and especially the value of the roughness exponent, was found to be largely independent of this value. On the contrary, the roughness amplitude was observed to slightly increase while the value of f_t was decreased.

For the fracture surfaces analyzed in the following, f_t is taken close to 0.10 but sufficiently small to avoid any uncracked ligaments behind the main crack. Regions with $f \geq f_t$, but that are not connected to the main crack, were found—these are not shown in Fig. 5. They do not contribute to the roughness of the fracture surfaces.

The roughness is computed for the two fracture surfaces produced by each calculation. One fracture surface, termed as the lower fracture surface, is obtained from the smallest value of z, for a given position (x, y) in the mean fracture plane, for which $f \geq f_t$ on the connected crack while the other fracture surface, termed the upper fracture surface, corresponds to the largest value of z, for a given position (x, y) in the mean fracture plane, for which $f \geq f_t$ on the connected crack. The crack profiles obtained from the lower and upper fracture surfaces are denoted by $h^{\mathrm{bot}}(x)$ and $h^{\mathrm{top}}(x)$, respectively. These are shown on Fig. 6 for the void volume fraction distribution of Fig. 5. The jumps in the profiles reflect the overhangs in the connected crack

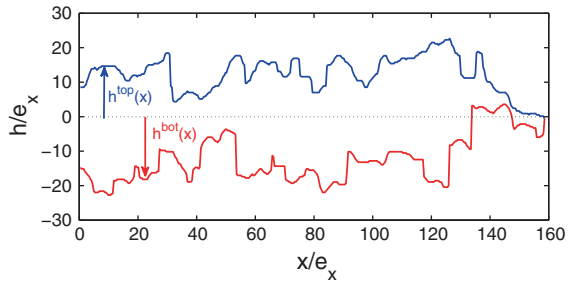

Fig. 6 Fracture profiles extracted along the crack propagation direction x obtained from the void volume fraction distribution of Fig. 5 using $f_t = 0.04$ for material # 3

as seen in the porosity distribution. Our definition of the crack profiles mimics a profilometer that measures only the highest point on the fracture surface and so prevents the presence of overhang on the crack surfaces analyzed subsequently.

4 Statistical analysis of fracture surfaces

4.1 Correlation functions of fracture surfaces and roughness exponent

We first characterize the fluctuations of heights of the fracture surface using the correlation function, Δh, defined as

$$\Delta h(\delta x) = \sqrt{\langle [h(x+\delta x,\ y) - h(x,\ y)]^2 \rangle_{x,\,y}} \qquad (17)$$

Here, $\langle\ \rangle_x$ denotes the average over x and y. The quantity $\Delta h(\delta x)$ can be interpreted as the typical difference of height between two points separated by a distance δx along the mean fracture plane. We focus on the correlation of heights in the direction of propagation only (x-axis), the width of the specimen in the perpendicular direction z being too small to allow a statistical analysis of the roughness in that direction.

The correlation function is computed for both the upper and lower fracture surfaces. The final correlation function is then obtained by averaging over these statistically equivalent surfaces. Indeed, from symmetry arguments, we expect the upper and lower fracture surfaces in each calculation to share similar statistical properties. The correlation function $\Delta h(\delta x)$ is shown in Fig. 7 on a logarithmic scale, after normalization of the axis by the element spacing e_x along the x-axis in the fine mesh region. Figure 7a shows plots for the four inclusion distributions analyzed in Needleman et al. (2012) corresponding to material #1 ($\epsilon_N = 0.3$ and $\sigma_N/\sigma_0 = 1.5$) using the same notation as in Needleman et al. (2012) to identify the inclusion distributions. The correlation functions exhibit a power law behavior

$$\Delta h(\delta x) \propto \delta x^{\beta} \qquad (18)$$

over more than two decades, characterized by the roughness exponent β. In this logarithmic representation, β is the slope of the solid straight lines, power law fits of the correlation functions computed from the fracture surfaces. Although the roughness amplitude—the position of the lines along the ordinates—may vary from one sample to another, the slope remains rather independent of the inclusion spatial distribution, as illustrated by the values of β given in the legend of the figure and obtained from a power law fit of these curves. These results give an average roughness exponent $\beta = 0.55 \pm 0.02$, independent of the inclusion distribution.

Our calculations capture the self-affine nature of ductile fracture surfaces, as observed experimentally in Bonamy and Bouchaud (2011), and as also found numerically in Needleman et al. (2012), but here over a larger range of length scales than in Needleman et al. (2012). Since quadratic elements are used the mesh spacing is $e_x/2$ but the scaling in Fig. 7 holds for even smaller δx values. This indicates that the interpolation procedure between mesh points when defining the fracture surface geometry still preserves the roughness scaling. In Needleman et al. (2012) values $0.4 < \beta < 0.6$ were reported for the simulations in Fig. 7a. The procedure used in this study to define the fracture surfaces from the porosity field gives a more precise characterization of the crack roughness, a more uniform value of the roughness exponent and the value of the roughness exponent β exhibits a smaller variation with the inclusion distribution than in Needleman et al. (2012). There are variations in the amplitude of the roughness that reflect variations in the threshold value f_t of the porosity used to define the fracture surfaces. Nevertheless, despite these fluctuations, the value of the roughness exponent is very robust, and changes very weakly from one inclusion distribution to another. We have also checked that the value of the roughness exponent does not depend much on the value of the threshold f_t used to define the fracture surfaces using both $f_t = 0.04$ and $f_t = 0.10$, supporting the value $\beta \simeq 0.55$ reported in this study.

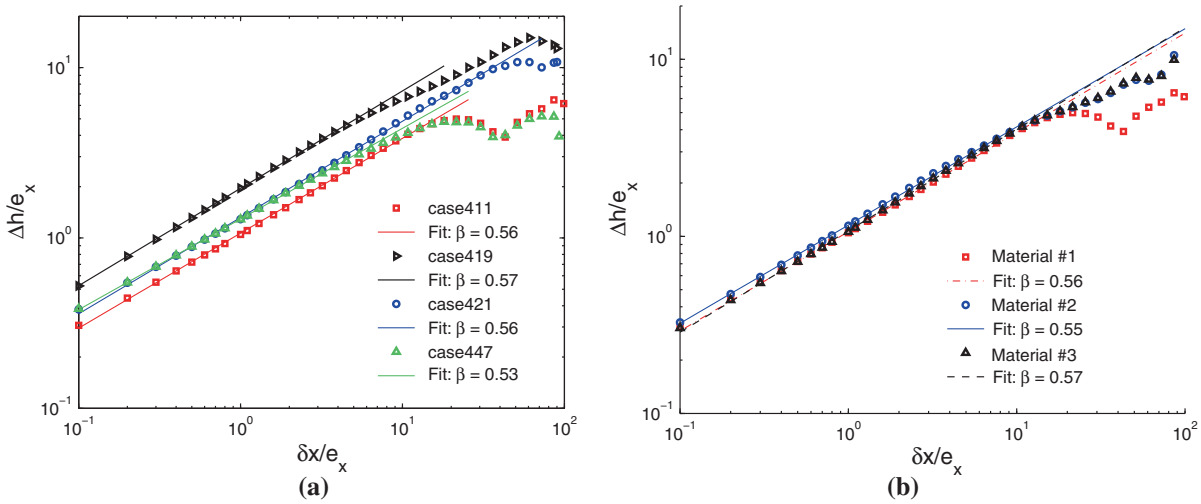

Fig. 7 Correlation function of the roughness of simulated fracture surfaces. **a** Material # 1: scaling behavior of the crack roughness observed for four different spatial distributions of the inclusions (see Needleman et al. 2012) characterized by a similar roughness exponents $\beta = 0.55 \pm 0.02$. **b** Comparisons of materials #1, #2 and #3: Scaling behavior of the crack roughness for the same spatial distribution of inclusions, but with different material properties. regardless of the value of the nucleation threshold, the roughness follows power law behavior with a roughness exponent $\beta = 0.56 \pm 0.01$

Figure 7b shows the effect of varying a material parameter for a fixed spatial distribution of inclusions. The inclusion distribution chosen here is case411 (Needleman et al. 2012). In one calculation, the void nucleation strain is taken to be $\epsilon_N = 0.4$ (material #2) while in the other calculation $\sigma_N/\sigma_0 = 2.0$ (material #3). All other material and void nucleation parameters remain fixed. Also, for comparison purposes, the data for case411 with the reference void nucleation parameters (material # 1) is also plotted. The simulated fracture surfaces here show a power law behavior, with a roughness exponent, $\beta \simeq 0.56$, rather independent of the material properties in the range investigated. This result is consistent with a large body of experimental work that reports a universal value of the roughness exponent, independent to a large extent of the microstructure of the material studied, its mechanical properties or the loading conditions used during the fracture test as seen in Bonamy and Bouchaud (2011), Bouchaud et al. (1990), Ponson et al. (2006).

4.2 Non-Gaussian statistics of height fluctuations

As seen in Fig. 7, the Hurst exponents calculated for the various inclusion distributions and material parameter variations considered here and in Needleman et al. (2012) are nearly the same. In Vernède et al. (2013) the full statistics of fracture surface height fluctuations were obtained for cracks in a variety of materials. It was found that the height fluctuations could be described by a distribution that differed from a Gaussian by having power law tails. The deviation from Gaussian statistics was found to be material dependent. Therefore, in order to explore possible effects of material characteristics on the predicted fracture surface morphology, we investigate the full statistics of height variations $\delta h(x, y)$ defined by

$$\delta h(x, y) = h(x + \delta x, y) - h(x, y) \tag{19}$$

Following Ponson (2007), Santucci et al. (2007), the procedure used to compute the histogram $P(\delta h)$ at a given scale δx is the following.

1. We fix first a value of δx.
2. For each location (x, y) on both the upper and lower fracture surface, the corresponding height variations δh are computed. This procedure results in a large set $\{\delta h\}_{\delta x}$ of height variations at a given scale δx.
3. The histogram of this set of values is computed. The histogram of δh is calculated by placing into 'boxes' $[b_{\min} b_2]$, $[b_2 b_3]$,..., $[b_{n-1} b_{\max}]$ the values of δh where the side b_1, b_2 of the boxes are distrib-

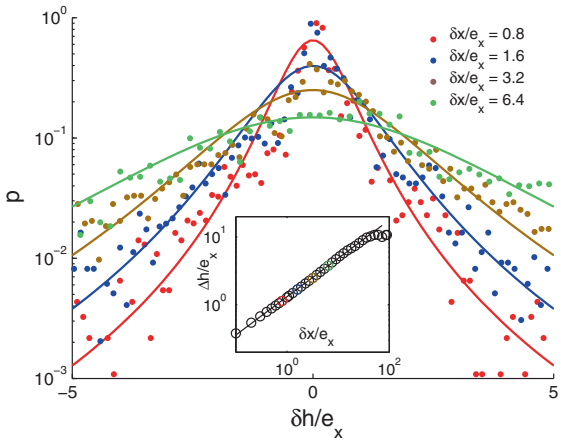

Fig. 8 Histograms $P(\delta h|\delta x)$ of height variations δh (see Eq. (19)) for various values of δx for material #1 and the spatial distribution of particles labeled case421 in Needleman et al. (2012). The *solid lines* are fits based on the Student's distribution of Eq. (20) using various values of the parameters k and δh_c.The *inset* represents the correlation function of the surface. The *points* (or second moment) corresponding to the distributions represented in the main panel are plotted using the same color code

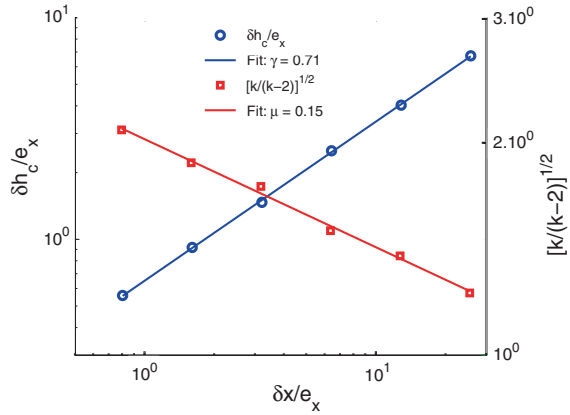

Fig. 9 Variations of the parameters $\sqrt{k/(k-2)}$ and $\delta h_c/e_x$ obtained from the fit of the roughness distributions $P(\delta h)$ shown in Fig. 8 for case 421 in Needleman et al. (2012) using Student's distribution (see Eq. (20)). The transition from power law tails ($\sqrt{k/(k-2)} \gg 1$) to Gaussian ($\sqrt{k/(k-2)} \simeq 1$) statistics as the length scale δx increases is characterized by the exponent $\mu = 0.15$. The typical value of δh_c of the roughness extracted using fits by Student's distribution also evolves as a power law of the length scale δx with exponent $\gamma = 0.71$

uted homogeneously between $b_{\min} = \min[\delta h]$ and $b_{\max} = \max[\delta h]$.

4. The histogram or probability density

$$P_i\left(\delta h_i = \frac{b_i + b_{i+1}}{2}|\delta x\right)$$

is calculated as the fraction of values of δh contained in the i^{th} box.

In order to study the scaling behavior of the fracture surface roughness, this procedure is repeated for various scales δx, leading to a family of histograms $P(\delta h|\delta x)$. An important property is that the standard deviation of the distributions $P(\delta h|\delta x)$ corresponds to the correlation function $\Delta h(\delta x)$ of the fracture surfaces, as can be observed from its definition in Eq. (17).

The distribution $P(\delta h|\delta x)$ is shown in Fig. 8 for four values of δx for the simulation labeled case421 in Needleman et al. (2012). The larger the value of δx, the broader the distribution, as expected from the scaling of the correlation function $\Delta h(\delta x) \propto \delta x^{\beta}$. The standard deviation of these histograms is plotted in the inset as a function of δx using the same color code as in the main panel. As expected, the standard deviation evolves as a power law with exponent $\beta = 0.56$, in agreement with the plot of $\Delta h(\delta x)$ shown in Fig. 7.

The distribution of height variations computed on numerical fracture surfaces are not Gaussian, but exhibit fat tails with power law behavior $P(\delta h) \propto \delta h^{(k+1)/2}$ for large values of δh. This means that large height fluctuations are not exponentially rare on ductile fracture surfaces as is the case for brittle fracture surfaces (Ponson et al. 2007). Indeed, the crack profiles in the calculations here display a non-negligible number of abnormally large fluctuations δh, as qualitatively seen in Fig. 6.

To describe this effect more quantitatively, the distributions $P(\delta h)$ are described using a family of probability distributions referred to as Student's t distributions

$$t_{k,\delta h_c}(\delta h) \propto \frac{1}{\delta h_c}\left(1 + \frac{1}{k}\left(\frac{\delta h}{\delta h_c}\right)^2\right)^{-(k+1)/2} \tag{20}$$

with parameters k and δh_c, and represented by solid lines in Fig. 8. For reasons discussed in the next section, it is more appropriate to consider the parameter $\sqrt{k/(k-2)}$ instead of k.

These parameters, obtained from the fit of the distributions $P(\delta h|\delta x)$, are represented in Fig. 9 as a function of the scale δx. The first parameter that characterizes the shape of the distribution decreases as a power law with the scale δx

$$\sqrt{\frac{k}{k-2}} \propto \delta x^{-\mu} \tag{21}$$

with $\mu = 0.15$ as illustrated by the straight line variations in the logarithmic representation of Fig. 9.

The second fitting parameter used to describe the roughness distributions $P(\delta h|\delta x)$ is also represented in Fig. 9 as a function of the scale δx. It here increases as a power law

$$\delta h_c \propto \delta x^{\gamma} \tag{22}$$

with exponent $\gamma = 0.71$.

The ductile crack profiles studied here exhibit a more complex behavior than brittle fracture surfaces. Indeed, in brittle materials, distributions of height variations follow a Gaussian behavior at all scales (Ponson 2007; Ponson et al. 2007). In our description based on Student's t distribution, this corresponds to $k \to \infty$, or equivalently, to $\sqrt{k/(k-2)} = 1$, for any value of δx. In other words, only one exponent is needed to describe the scaling of the roughness distribution, since $\mu = 0$. For ductile fracture surfaces, not one, but three scaling exponents β, μ and γ are required. However, this can be reduced to two independent exponents based on a simple relationship. Using the definition (20) of Student's t distribution, one can show that its second moment, and so the correlation function, varies as

$$\Delta h(\delta x) \propto \delta h_c \sqrt{k/(k-2)} \tag{23}$$

As a result, from the scaling relations in Eqs. (18), (21) and (22), one obtains

$$\beta = \gamma - \mu. \tag{24}$$

This relation is satisfied by the exponents extracted here for which $\beta = 0.56$ (case421) was measured (see Fig. 7). This supports our description of the roughness statistics based on Student's t distribution.

The relation between exponents suggests an interpretation of the roughness exponent β that characterizes the scaling behavior of the correlation function with the scale δx as the combined effect of the variations of the typical roughness δh_c with the scale δx and the variations in the actual shape of the distribution of height fluctuations. It also suggests that the universal value of the roughness exponent β relies on the value of μ that describes the change in the shape of roughness distribution. Experimental results in Vernède et al. (2013) are consistent with these findings: the statistical analysis of both aluminum alloy and mortar fracture surfaces indicates a similar behavior, with a measured value $\mu^{\exp} \simeq 0.15$. However, the experimental value of the second exponent, $\gamma^{\exp} \simeq 0.9$, is larger than the value obtained for the calculated fracture surfaces here.

In the following, we use the above analysis of the distribution of height variations to compare the fracture surfaces obtained for materials #1, #2 and #3.

4.3 Roughness statistics: comparison between the materials

The analysis in Sect. 4.2 revealed that the scaling of the roughness distribution of ductile fracture surfaces could be described using Student's $t_{k,\delta h_c}$ distribution introduced in Eq. (20), with the adjustable parameters $\sqrt{k/(k-2)}$ and δh_c that follow power laws with the scale δx. We focus in this section on the parameter $\sqrt{k(k-2)}$ that characterizes the distribution shape. For a finite value of the parameter k, Student's t distribution displays fat tails with

$$t_{k,\delta h_c}(\delta h) \propto \delta h^{(k+1)/2} \quad \text{for } \delta h \gg \delta h_c$$

while $t_{k,\delta h_c}(\delta h)$ approaches a Gaussian distribution when k tends to infinity. As a result, this family of distributions is suited to describe a transition from power law tails to Gaussian statistics. In particular $\sqrt{k/(k-2)} = 1$ for a Gaussian distribution and $\sqrt{k/(k-2)} > 1$ otherwise. We expect the value of $\sqrt{k/(k-2)}$ to indicate the distance from Gaussian behavior.

The analysis of the fracture surface of material #1 has revealed that $\sqrt{k/(k-2)}$ remains larger than one, indicating fat tails statistics and deviation from a Gaussian distribution at the length scales investigated (see Fig. 9). However, extrapolating the power law behavior $\sqrt{k/(k-2)} \sim \delta x^{-\mu}$ towards larger values of δx, predicts that it will reach $\sqrt{k/(k-2)}(\delta x = \xi) = 1$ at the crossover length $\delta x = \xi$. Interestingly, it means that for $\delta x > \xi$, the fracture surface might recover Gaussian statistics. Such an extrapolation leads to $\xi = 100 e_x$ for material #1.

The transition from fat tail statistics for $\delta x < \xi$ to Gaussian behavior for $\delta x > \xi$ calls for an interpretation in terms of fracture mechanisms. In the scenario proposed in Vernède et al. (2013), Gaussian statistics of the roughness would be reminiscent of brittle fracture surfaces while distributions with fat tails would be signature of damage mechanisms. This behavior is fairly consistent with the scaling behavior of the correlation function that gives roughness exponents of the order of $\beta = 0.5$ for brittle failure and larger values around $\beta = 0.6$ at a smaller scale when failure is accompanied by damage mechanisms (Bonamy

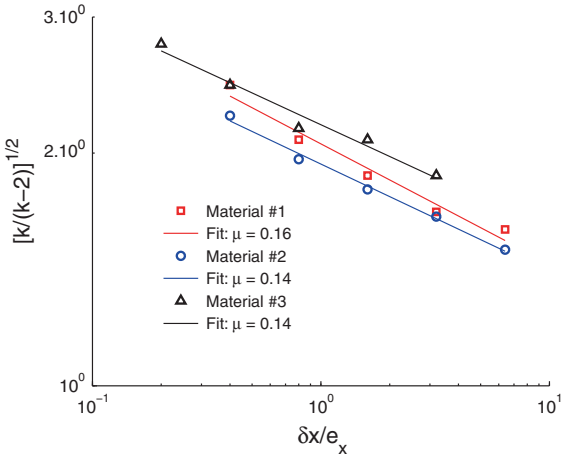

Fig. 10 Variations of the parameter $\sqrt{k/(k-2)}$ used in the fit by Student's distribution, Eq. (20), of the histograms $P(\delta h)$ of height fluctuations as a function of δx

et al. 2006; Morel et al. 2008). The crossover length between the two regimes - large exponents and fat tail statistics at small scale to small exponents with Gaussian statistics at large scale - may be a reflection of the size scale associated with the micro mechanism of fracture. If so, the value of ξ extracted from the analysis of the roughness statistics could serve as a length scale associated with the underlying damage process.

Figure 10 compares the variations of the parameter $\sqrt{k/(k-2)}$ for the materials analyzed. The power law decay differs from one material to another. The exponent $\mu \simeq 0.15$ —the slope of the straight line variations in this logarithmic representation—remains rather constant. But the variations in the position along the ordinates indicate different values of ξ, and more generally, different level of deviations from Gaussian (and so brittle) behavior.

The deviation from Gaussian behavior is more pronounced for material #3 than for material #2 for which the values of $\sqrt{k/(k-2)}$ is smaller at all length scales δx. This may indicate that material #3 is more ductile than material #2 which is expected due to the larger value of the nucleation stress for the large inclusions. Extrapolation of $\sqrt{k/(k-2)}$ to large values of δx indicates that the crossover value $\xi = 300e_x$ for material #3 which is much larger than the value estimated for the materials #1 and #2.

The comparison of material #1 with material #2 is more complex. The relative position of the curves $\sqrt{k/(k-2)}$ indicates stronger deviations from Gaussian behavior for material #1. However, the extrapolated value of $\xi \simeq 100e_x$ gives a similar value for both materials.

5 Discussion

Calculations have been carried out for a fixed size, density and distribution of discretely modeled void nucleating inclusions. The statistics of the fracture surfaces was investigated. In addition, the statistics of the fracture surfaces obtained in Needleman et al. (2012) were recalculated. In Needleman et al. (2012), the fracture surface was defined by a threshold value $f_t = 0.1$, somewhat smaller than the material parameter f_t used in the constitutive model to define final failure. In the present studies, in some cases, an even smaller value, $f_t = 0.03$, is chosen to define a well connected crack surface. Fig. 5 shows that the crack surfaces obtained by these definitions are rather close, as the value of f increases steeply in the material near the fully open crack. In Needleman et al. (2012), a thickness average of the fracture surface height was calculated and this average profile was used to obtain the correlation function $\Delta h(\delta x)$. Here, instead, $\Delta h(\delta x)$ is calculated for each cross section, and is subsequently averaged, and it turns out that this procedure gives power law scaling over a larger range, with less difference between the various inclusion distributions in Needleman et al. (2012). Also in Needleman et al. (2012) only the roughness exponent was calculated whereas here we calculate more complete fracture surface statistics.

As noted in Needleman et al. (2012) the calculations contain several length scales: (1) the mean spacing between inclusion centers; (2) the inclusion radius; (3) the slice thickness; and (4) the finite element mesh spacing. A significant difference between these length scales is that the first three are physically relevant length scales whereas any dependence of the results on the finite element mesh length scale is a numerical artifact. Whether or not there is a significant effect of the finite element mesh spacing on the predicted statistics of fracture roughness remains to be determined.

Our results give rise to a value of $\beta = 0.56 \pm 0.03$ that is essentially independent of the inclusion distributions and fracture properties considered and that is also close to the value $\beta^{3D} \simeq 0.6$ reported for 3D fracture surfaces of a wide range of ductile and quasi-brittle

materials in the crack propagation direction Ponson et al. (2006), Ponson et al. (2006). While the probability distribution of heights is a Gaussian for brittle fracture, it is more complex for ductile fracture surfaces. We showed that Student's t distribution can fit all the histograms in Fig. 8. This distribution is equivalent to a Gaussian at sufficiently small values of $\delta h/e_x$ but crosses over to a power-law at larger values. Also, as seen in Fig. 8 for small values of $\delta x/e_x$ the computed distributions display fat power law tails whereas for sufficiently large values of $\delta x/e_x$ the distribution is well-fit by a Gaussian. The transition from fat tail to Gaussian statistics is described by introducing a second exponent $\mu \simeq 0.15$. In the experiments in Vernède et al. (2013) where the roughness statistics were calculated as an average over all directions in the fracture surface, the statistics were also observed to deviate from a Gaussian and the deviation was well-described by two exponents. The value of μ was close to the one obtained here but the value of γ was larger than in our calculations. Hence, while for brittle fracture surfaces, knowledge of the roughness index β is sufficient to describe completely the probability distribution, for ductile fracture surfaces independent exponents are needed.

Our results indicate that for ductile fracture surfaces there is a length characterizing the crossover from power law to Gaussian statistics. The results also suggest that this length may correlate with the size of the fracture process zone. What remains to be determined is the extent to which this crossover length as well as other aspects of the non-Gaussian statistics correlate with measures of fracture toughness such as K_{Ic} and crack growth resistance curves. For the type of materials considered here, this requires more extensive calculations with increased crack growth, larger variations of inclusion density and distribution, and greater variation in material properties. Such calculations are being undertaken and the results here suggest the possibility of a connection being made between post-mortem fracture surface statistics and crack growth resistance.

6 Conclusions

1. A central result of this paper is that, for the material parameter variations analyzed, the computed ductile fracture surfaces exhibit self-affine properties with similar values of the fracture surface roughness exponent β but their full statistical properties differ.
2. For the ductile fracture surfaces analyzed:

- The computed fracture surfaces are self-affine over a range of length scales of about two orders of magnitude.
- The computed values of the fracture surface roughness exponents along the crack growth direction are not sensitive to the larger particle distributions nor to the fracture parameter variations investigated.
- The computed fracture surface roughness distributions are not Gaussian but they are all well fitted by Student's distribution. Both the roughness exponent $\beta = 0.56 \pm 0.03$ and the exponent $\mu \simeq 0.15$ characterizing the transition with $\delta x/e_x$ from fat tail to Gaussian statistics are found to be rather independent of the spatial distributions of the larger inclusions and of material parameters in the range investigated.
- The computed full fracture surface roughness statistics vary with the fracture parameters investigated.

Acknowledgments The financial support provided by the U.S. National Science Foundation through Grant CMMI-1200203 and by the European Union through the ToughBridge Marie Curie Grant (LP) is gratefully acknowledged. We also thank Jean-Philippe Bouchaud and Stéphane Vernède for fruitful discussions

References

Alava MJ, Nukala PKVV, Zapperi S (2006) Statistical models of fracture. Adv Phys 55:349–476

Belytschko T, Chiapetta RL, Bartel HD (1976) Efficient large scale non-linear transient analysis by finite elements. Int J Numer Methods Eng 10:579–596

Bonamy D, Ponson L, Prades S, Bouchaud E, Guillot C (2006) Scaling exponents for fracture surfaces in homogeneous glass and glassy ceramics. Phys Rev Lett 97:135504

Bonamy D, Bouchaud E (2011) Failure of heterogeneous materials: a dynamic phase transition? Phys Rep 498:1–44

Bouchaud E, Lapasset G, Planès J (1990) Fractal dimension of fractured surfaces: a universal value? Europhys Lett 13:73–79

Bouchaud E (2003) The morphology of fracture surfaces: a tool for understanding crack propagation in complex materials. Surf Rev Lett 10:797–814

Bouchbinder E, Mathiesen J, Procaccia I (2004) Roughening of fracture surfaces: the role of plastic deformation. Phys Rev Lett 92:245505

Cherepanov GP, Balankin AS, Ivanova VS (1995) Fractal fracture mechanics–a review. Eng Fract Mech 51:997–1033

Chu CC, Needleman A (1980) Void nucleation effects in biaxially stretched sheets. J Eng Mater Technol 102:249–256

Gurson AL (1975) Plastic flow and fracture behavior of ductile materials incorporating void nucleation, growth and interaction. Ph.D. Thesis, Brown University.

Måløy KJ, Hansen A, Hinrichsen EL, Roux S (1992) Experimental measurements of the roughness of brittle cracks. Phys Rev Lett 68:213–215

Mandelbrot BB, Passoja DE, Paullay AJ (1984) Fractal character of fracture surfaces of metals. Nature 308:721–722

Morel S, Bonamy D, Ponson L, Bouchaud E (2008) Transient damage spreading and anomalous scaling in mortar crack surfaces. Phys Rev E 78:016112

Nakano A, Kalia RK, Vashista P (1995) Dynamics and morphology of brittle cracks: a molecular-dynamics study of silicon nitride. Phys Rev Lett 75:3138–3141

Needleman A, Tvergaard V, Bouchaud E (2012) Prediction of ductile fracture surface roughness scaling. J Appl Mech 79:031015

Peirce D, Shih CF, Needleman A (1984) A tangent modulus method for rate dependent solids. Comput Struct 18:875–887

Ponson L, Auradou H, Vié P, Hulin JP (2006) Low self-affine exponents of fractured glass ceramics surfaces. Phys Rev Lett 97:125501

Ponson L, Bonamy D, Bouchaud E (2006) Two dimensional scaling properties of experimental fracture surfaces. Phys Rev Lett 96:035506

Ponson L, Bonamy D, Auradou H, Mourot G, Morel S, Bouchaud E, Guillot C, Hulin J-P (2006) Anisotropic self-affine properties of experimental fracture surfaces. Int Fract 140:27–36

Ponson L (2007) Crack propagation in disordered materials: how to decipher fracture surfaces. Ann Phys 32:1–120

Ponson L, Auradou H, Pessel M, Lazarus V, Hulin J-P (2007) Failure mechanisms and surface roughness statistics of fractured Fontainebleau sandstone. Phys Rev E 76:036108

Ramanathan S, Ertas D, Fisher DS (1997) Quasistatic crack propagation in heterogeneous media. Phys Rev Lett 79:873–876

Santucci S, Måløy KJ, Delaplace A, Mathiesen J, Hansen A, Bakke J, Schmittbuhl J, Vanel L, Ray P (2007) Statistics of fracture surfaces. Phys Rev E 75:016104

Tvergaard V (1981) Influence of voids on shear band instabilities under plane strain conditions. Int J Fract 17:389–407

Tvergaard V (1982a) On localization in ductile materials containing spherical voids. Int J Fract 18:237–252

Tvergaard V (1982b) Influence of void nucleation on ductile shear fracture at a free surface. J Mech Phys Solids 30:399–425

Tvergaard V, Needleman A (1984) Analysis of the cup-cone fracture in a round tensile bar. Acta Metall 32:157–169

Vernède S, Cao Y, Bouchaud JP, Ponson L (2013) Extreme events and non-Gaussian statistics of experimental fracture surfaces (submitted for publication)

DOI 10.1007/978-3-319-04397-5_12
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9847-y

ORIGINAL PAPER

Efficient pseudo-spectral solvers for the PKN model of hydrofracturing

Michal Wrobel · Gennady Mishuris

Received: 5 February 2013 / Accepted: 26 March 2013 / Published online: 11 April 2013

Abstract In the paper, a novel algorithm employing pseudo-spectral approach is developed for the PKN model of hydrofracturing. The respective solvers compute both the solution and its temporal derivative. In comparison with conventional solvers, they demonstrate significant cost effectiveness in terms of balance between the accuracy of computations and densities of the temporal and spatial meshes. Various fluid flow regimes are considered.

Keywords Hydrofracturing · PKN model · Numerical modelling

1 Introduction and preliminary results

Hydraulic fracturing is a widely used method for stimulation of hydrocarbons reservoirs. This technology has been known and successfully applied for a few decades (Khristianovic and Zheltov 1955; Hubbert and Willis 1957; Crittendon 1959). Recently it has been revived, due to economical reasons, as a basic technique for exploitation of unconventional deposits of oil and gas. The phenomenon of a fluid driven fracture propagating in a brittle medium is also present in many natural processes (e.g. magma driven dykes—Rubin 1995, subglacial drainage of water—Tsai and Rice 2010).

Throughout the years, starting from the pioneering works of Sneddon and Elliot (1946), Khristianovic and Zheltov (1955), Perkins and Kern (1961), Geertsma and de Klerk (1969), Nordgren (1972), various models of hydrofracturing have been formulated and used in applications. A broad review of the topic can be found in Kovalyshen and Detournay (2009), Kovalyshen (2010), Linkov (2011c), Kusmierczyk et al. (2012). Together with increasing complexity of the models describing this multiphysics process, the computational techniques have been continuously enhanced. A comprehensive survey on the algorithms and numerical methods used in hydrofracturing simulation can be found in Adachi et al. (2007), Economides (2000).

Responding to the recent demand, an increasing stream of publications have appeared concerning additional information on seismic events, shear stresses in the rock formation, multifracturing and others and their implementation into the solvers (Zhang et al. 2004; Moos 2012; Olson 2008; Dobroskok and Linkov 2011). Also, a considerable effort has been made to

This work has been done in the framework of the EU FP7 PEOPLE project under contract number PIAP-GA-2009-251475-HYDROFRAC.

M. Wrobel · G. Mishuris (✉)
Institute of Mathematical and Physical Sciences, Aberystwyth University, Ceredigion, Wales SY23 3BZ, UK
e-mail: ggm@aber.ac.uk

M. Wrobel · G. Mishuris
Eurotech Sp. z o.o., ul. Wojska Polskiego 3, 39-300 Mielec, Poland

improve the existing algorithms by incorporating new efficient numerical techniques (Lecampion and Detournay 2007; Linkov 2011c; Peirce and Detournay 2008, 2009).

The main computational challenges associated with the modelling of hydraulic fractures are: (a) strong nonlinearity resulting from the coupling between the solid and fluid phases, (b) singularity of the gradients of the physical fields near the crack tip, (c) moving boundaries, (d) degeneration of the governing equations at the crack tip, multiscaling and others.

To achieve the maximal possible efficiency of numerical simulations, the computational algorithms should be formulated in *proper* variables accounting for all the problem peculiarities (Linkov 2011c). As a result, they allow one to reduce the volume of processed data, which is especially important when dealing with complex geometries and/or multifracturing.

The analysis presented in this paper is devoted to the PKN model of hydrofracturing. This model contains all the peculiarities mentioned above, except for the nonlocal relation for the fluid-solid coupling. Although we restrict our interest only to a single fracture, the developed algorithms, thanks to their robustness, can be successfully applied to model a system of cracks.

The numerical analysis of the problem should be backdated to Nordgren (1972) who extended the Perkins and Kern model (Perkins and Kern 1961) to account for the fluid loss effect and fracture volume change. As a result, the crack length was determined as part of the solution. The author proposed a finite difference scheme to solve the problem, which is in fact equivalent to the finite volume (FV) method.

Further development of the PKN formulation was done by Kemp (1989), who (a) implemented the specific boundary condition at the moving crack tip into the FV scheme, (b) incorporated asymptotic behaviour of the solution near the crack tip in a special tip element, (c) indirectly used the fourth power of the crack opening (w^4) as a new dependent variable, instead of the crack opening itself. For the early-time asymptotic model Kemp proposed a power series solution, presenting its first four terms.

The recent paper by Kovalyshen and Detournay (2009) has extended most of Kemp's results, incorporating all information on the PKN model available to date. They present various asymptotics, complete analytical solution for an impermeable rock (directly extending the results from Kemp (1989) from four leading terms to an infinite series representation), FV algorithm with a special tip element and a numerical benchmark for the Carter leak-off, linking the results to the scaling approach developed in Adachi and Peirce (2007), Detournay (2004), Mitchell et al. (2007a,b).

In Linkov (2011a,b,c), the PKN model was reformulated to improve the efficiency and stability of computations by (1) introducing *proper* dependent variables (cubed fracture opening, w^3), (2) utilizing the speed equation and (3) by imposing a modified boundary condition at a small distance behind the crack tip (ε-regularisation). Additionally, the analytical solution for an impermeable rock was evaluated for the new dependent variable in a form of rapidly converging series in Linkov (2011c). Moreover, the author highlighted (Linkov 2011a) that numerical schemes exploiting a fixed position of the crack tip during the iterations may become ill-posed.

In Mishuris et al. (2012) and Kusmierczyk et al. (2012) the ε-regularisation technique was further enhanced by (1) appropriate adaptation of the speed equation to the chosen numerical scheme and (2) improved way of imposing of the regularized boundary condition. A detailed discussion on various aspects of application of implicit and explicit numerical schemes was provided.

In this paper we are presenting a novel algorithm based on the pseudo-spectral approach. Namely, we propose an efficient numerical algorithm to solve a specific self-similar problem and extend the results to the general (transient) formulation. Since, the integration schemes used in the algorithm incorporate the exact boundary conditions at the crack tip, no regularization technique is necessary. The most accurate two points representation of the temporal derivative is used to guarantee an optimal algorithm performance. Finally, two solvers are developed which show their robustness and stability. They both demonstrate high cost effectiveness in terms of the relationship between the volume of the processed data and the accuracy of computations. Moreover, additionally to the crack opening and length, the temporal derivative of the former and the crack tip velocity are automatically returned as components of the problem solution.

1.1 Problem formulation

Let us consider a symmetrical crack of length $2l$ situated in the plane $x \in [-l, l]$. The crack is fully filled by

a Newtonian liquid injected at the middle point ($x = 0$) with a known rate $q_0(t)$. Note here, that the crack length evolution, $l = l(t)$, is the result of fluid flow inside the fracture. Due to the symmetry of the problem, one can restrict the analysis to the half of the crack $x \in [0, l(t)]$.

The classic mathematical formulation of the PKN model of hydrofracturing was given in Nordgren (1972). Below we present a system of equations constituting the model. The mass conservation principle is expressed by the continuity equation:

$$\frac{\partial w}{\partial t} + \frac{\partial q}{\partial x} + q_l = 0, \quad t \geq t_0, \quad 0 \leq x \leq l(t), \tag{1}$$

while the Poiseuille equation describes the flow in a narrow channel. In the case of a Newtonian fluid, it is written in the following form:

$$q = -\frac{1}{M} w^3 \frac{\partial p}{\partial x}. \tag{2}$$

Here $w = w(t, x)$ stands for the crack opening, $q = q(t, x)$ is the fluid flow rate, $p = p(t, x)$ ($p = p_f - \sigma_0$, σ_0-confining stress) refers to the net fluid pressure. The constant M, involved in the Poiseuille equation, is computed as $M = 12\mu$, where μ denotes the dynamic viscosity (see for example Adachi and Detournay 2002). The function $q_l = q_l(t, x)$ from (1) is the volumetric rate of fluid loss to formation in the direction perpendicular to the crack surfaces per unit length of the fracture. This function is usually assumed to be given, but it may depend on the solution itself as well. To account for various leak-off regimes, we accept the following behaviour of q_l:

$$q_l(t, x) = Q_l(t)(l(t) - x)^{\eta}, \quad \text{for} \quad x \to l(t), \tag{3}$$

for some constant $\eta \geq -1/2$. Note that the case $\eta = -1/2$ corresponds to the Carter law (Carter 1957), while $\eta \geq 1/3$ guarantees that the leak-off vanishes near the crack tip as fast as the crack opening at least (see for details Kusmierczyk et al. 2012).

The group of fluid equations is to be supplemented by the relation describing deformation of the rock under applied hydraulic pressure. In the case of the PKN model, a linear relationship between the net fluid pressure and crack opening is in use:

$$p = kw, \tag{4}$$

where a known proportionality coefficient $k = \frac{2}{\pi h} \frac{E}{1-\nu^2}$ is found from the solution of a plane strain elasticity problem (Nordgren 1972) for an elliptical crack of height h. E and ν are the elasticity modulus and Poisson's ratio, respectively.

The above equations are equipped with the boundary condition at a crack mouth ($x = 0$) determining the injection flux rate:

$$-\frac{k}{M}\left[w^3 \frac{\partial w}{\partial x}\right]_{x=0} = q_0(t), \tag{5}$$

and two boundary conditions at a crack tip:

$$w(t, l(t)) = 0, \quad q(t, l(t)) = 0. \tag{6}$$

In order to define the crack length, $l(t)$, the global fluid balance equation is usually utilized (see for example Adachi et al. 2007)

$$\int_0^{l(t)} [w(t, x) - w(0, x)]dx - \int_0^t q_0(t)dt + \int_0^{l(t)} \int_0^t q_l(t, x)dtdx = 0. \tag{7}$$

Finally, the initial conditions are assumed in the following way:

$$w(0, x) = 0, \quad l(0) = 0. \tag{8}$$

System (1)–(8) constitutes the classic formulation of the PKN problem. It was shown in Kemp (1989) and Garagash and Detournay (2000) that the asymptotic behaviours of w and q near the crack tip are interrelated, and the first term of the expansion for the crack opening may be written as:

$$w(t, x) \sim w_0(t)\,(l(t) - x)^{\alpha}, \quad \text{as} \quad x \to l(t). \tag{9}$$

For the classic PKN model the exponent $\alpha = 1/3$ was found in Kemp (1989). Thus condition $(6)_2$ is always satisfied as it follows from (2) and (4). As a result, the model does not account for the standard stress singularity of fracture mechanics at the crack tip, and thus is relevant for the so-called zero toughness regime (see e.g. Adachi and Detournay 2002).

Remark 1 Despite that zero crack opening and length are considered as the initial conditions, all authors begin their studies from the asymptotic model for the small time. With the assumption of zero leak-off term in the continuity equation and constant q_0, the problem is reduced to a self-similar formulation. The full numerical analysis is then continued by taking the similarity solution as the initial state. This effectively means that the initial condition (8) can be replaced by the non-zero crack opening

$$l(0) = l_{\diamond}, \quad w(0, x) = w_{\diamond}(x), \quad x \in (0, l_{\diamond}). \tag{10}$$

In this paper, the modified formulation of the PKN model is considered, following the recent advance in the area of numerical modelling (Linkov 2011a,b,c; Mishuris et al. 2012). Thus, to trace the fracture front we use the so-called speed equation, instead of the fluid balance relationship (7):

$$\frac{dl}{dt} = v_0(t) = \frac{q}{w}\Big|_{x=l(t)}. \tag{11}$$

The speed equation assumes that the fracture tip coincides with the fluid front, which excludes the presence of a lag or an invasive zone ahead of the fracture tip. Originally it was introduced by Kemp (1989) and has been recently revisited by Linkov (2011a,b,c).

Note that, on substitution of Eqs. (2), (4) and (9) into (11), one obtains a relationship between the crack propagation speed and the multiplier of the leading term of the crack opening asymptotic expansion (9):

$$\frac{dl}{dt} = \frac{k}{3M} w_0^3(t). \tag{12}$$

This implies that the quality of the numerical estimation of w_0 [(*see estimate* (9)] *should be vital for the accuracy of computations.*

By substituting the Poiseulle equation (2) into the continuity equation (1) one obtains a lubrication (Reynolds) equation for the considered problem, where the net fluid pressure function $p(t, x)$ is eliminated:

$$\frac{\partial w}{\partial t} - \frac{k}{M}\frac{\partial}{\partial x}\left(w^3 \frac{\partial w}{\partial x}\right) + q_l = 0, \quad t \geq t_0, \quad 0 \leq x \leq l(t). \tag{13}$$

In this way the modified formulation of the PKN model includes: (i) the Reynolds equation (13), (ii) the boundary conditions (5)–$(6)_1$, (iii) the asymptotics (9), (iv) the initial conditions (10), (v) the speed equation in the form (12).

The paper is organized as follows: in the next subsection we present the normalized formulation of the problem. Then, two types of self-similar solutions for the PKN model are discussed. These solutions are used in Sect. 2 to investigate a numerical algorithm for a time independent variant of the problem. In Sect. 3, the algorithm is modified to tackle the transient regime. Two alternative integral solvers are developed and their performances and applicability are examined. Section 4 contains the final conclusions.

1.2 Normalized formulation

Following Kusmierczyk et al. (2012), we normalize the problem by introducing dimensionless variables:

$$\begin{aligned}
&\tilde{x} = \frac{x}{l(t)}, \quad \tilde{t} = \frac{t}{t_n}, \quad t_n = \frac{M}{k l_\diamond}, \quad \tilde{w}_\diamond(\tilde{x}) = w_\diamond(x),\\
&\tilde{w}(\tilde{t}, \tilde{x}) = \frac{w(t, x)}{l_\diamond}, \quad L(\tilde{t}) = \frac{l(t)}{l_\diamond}, \quad l_\diamond^2 \tilde{q}_0(\tilde{t}) = t_n q_0(t),\\
&\tilde{l}_\diamond \tilde{q}_l(\tilde{t}, \tilde{x}) = t_n q_l(t, x), \quad l_\diamond^{2/3} \tilde{w}_0(\tilde{t}) / L^{1/3}(\tilde{t}) = w_0(t),
\end{aligned} \tag{14}$$

where $\tilde{x} \in [0, 1]$, $L(0) = 1$.

In the new variables Eq. (13) reads:

$$\frac{\partial \tilde{w}}{\partial \tilde{t}} - \tilde{x}\frac{L'}{L}\frac{\partial \tilde{w}}{\partial \tilde{x}} - \frac{1}{L^2(t)}\frac{\partial}{\partial \tilde{x}}\left(\tilde{w}^3 \frac{\partial \tilde{w}}{\partial \tilde{x}}\right) + \tilde{q}_l = 0, \quad \tilde{t} \geq 0, \quad 0 \leq \tilde{x} \leq 1. \tag{15}$$

The boundary conditions (5)–$(6)_1$ may be rewritten as:

$$-\frac{1}{L(\tilde{t})}\left[\tilde{w}^3 \frac{\partial \tilde{w}}{\partial \tilde{x}}\right]_{\tilde{x}=0} = \tilde{q}_0(\tilde{t}), \quad \tilde{w}(\tilde{t}, 1) = 0. \tag{16}$$

The initial conditions (10) are defined as:

$$L(0) = 1, \quad \tilde{w}(0, \tilde{x}) = \tilde{w}_\diamond(\tilde{x}), \quad \tilde{x} \in [0, 1]. \tag{17}$$

The asymptotic expansion for crack opening (9) takes the form:

$$\tilde{w}(\tilde{t}, \tilde{x}) \sim \tilde{w}_0(\tilde{t})(1 - \tilde{x})^{1/3}, \quad \text{for} \quad \tilde{x} \to 1. \tag{18}$$

For the sake of completeness of the normalization, we also present the global fluid balance equation (7), although it will not be used later on:

$$L(\tilde{t}) \int_0^1 \tilde{w}(\tilde{t}, x)dx - \int_0^1 \tilde{w}(0, x)dx - \int_0^{\tilde{t}} \tilde{q}_0(t)dt + \int_0^{\tilde{t}} L(t) \int_0^1 \tilde{q}_l(t, x)dxdt = 0. \tag{19}$$

Finally, the transformation of the speed equation (12) yields:

$$\frac{d}{dt}L(\tilde{t}) = V_0(\tilde{t}) = \frac{1}{3L(\tilde{t})}\tilde{w}_0^3(\tilde{t}). \tag{20}$$

As shown in Mishuris et al. (2012), Eq. (20) is convenient to trace the fracture front when standard ODE solvers are in use for the dynamic system (DS) describing the problem. On the other hand, the crack length can be computed from (20) by direct integration to give:

$$L(\tilde{t}) = \sqrt{1 + \frac{2}{3}\int_0^{\tilde{t}} \tilde{w}_0^3(\tau)d\tau}, \tag{21}$$

which is useful when an implicit method (for example Crank-Nicolson scheme) is utilised (see Mishuris et al. 2012).

On substitution of (20) into (15) one can rewrite the later to obtain:

$$3L^2\left(\frac{\partial \tilde{w}}{\partial \tilde{t}} + \tilde{q}_l(\tilde{t},\tilde{x})\right) = \tilde{x}\tilde{w}_0^3\frac{\partial \tilde{w}}{\partial \tilde{x}} + 3\frac{\partial}{\partial \tilde{x}}\left(\tilde{w}^3\frac{\partial \tilde{w}}{\partial \tilde{x}}\right). \tag{22}$$

In the following, Eq. (22) will be used as a basic relation to formulate our integral solvers.

From now on, for convenience, we shall omit the tilde symbol in all quantities. In this way all the notations refer henceforth to the normalized formulation.

1.3 Self-similar solutions

Let us assume

$$q_l(t,x) = \gamma e^{\gamma t} q_l^*(x), \tag{23}$$

and look for the similarity solution of the problem in the form:

$$w(t,x) = u(x)e^{\gamma t}, \quad w_0(t) = u_0 e^{\gamma t}, \tag{24}$$

where the asymptotic behaviour (18) holds true, and u_0 is the limiting value of u defined in the same manner as in estimate (18). Thus, Eq. (20) transforms to an identity if one takes:

$$L^2(t) = \frac{2u_0^3}{9\gamma}e^{3\gamma t}. \tag{25}$$

On substitution of (23), (24), and (25) into the equation (22) one can reduce the latter to the following ordinary differential equation:

$$\beta u_0^3(u + q_l^*) = \mathcal{A}(u), \tag{26}$$

with $\beta = 2/3$. Here, the nonlinear differential operator $\mathcal{A}$ is defined by the right-hand side of equation (22) and is equipped with the boundary conditions

$$-3u_0^{-3/2}\left[u^3\frac{du}{dx}\right]_{x=0} = q_0^*, \quad u(1) = 0, \tag{27}$$

where we have introduced an auxiliary notation:

$$q_0^* = \sqrt{\frac{2}{\gamma}}e^{-\frac{5\gamma t}{2}}q_0(t). \tag{28}$$

If q_0^* is constant, then Eq. (26) together with the boundary conditions (27) do not depend on time and constitute a boundary value problem (BVP) degenerated at point $x = 1$. Indeed, the nonlinear coefficient in front of the second order term of the differential operator vanishes at the point $x = 1$ in view of the boundary condition $(27)_2$. This BVP is in fact a self-similar formulation of the original problem with specific, given leak-off regime and the inlet flux.

Other class of similarity solutions can be found, for some $a \geq 0$, by assuming:

$$q_l(t,x) = \gamma(t+a)^{\gamma-1}q_l^*(x), \quad w(t,x) = (t+a)^\gamma u(x),$$
$$w_0(t) = u_0(t+a)^\gamma, \tag{29}$$

$$L^2(t) = \frac{2u_0^3}{3(3\gamma+1)}(a+t)^{3\gamma+1}. \tag{30}$$

As a result, one again obtains the BVP (26)–(27) with $\beta = 2\gamma/(3\gamma+1)$ and

$$q_0^* = \sqrt{\frac{6}{(3\gamma+1)}}(a+t)^{\frac{1-5\gamma}{2}}q_0(t). \tag{31}$$

Thus, for $\gamma = 1/5$ the self-similar solution corresponds to the constant injection flux rate, while the crack propagation speed decreases with time as $L'(t) = O(t^{-1/5})$ for $t \to \infty$. If, however, one takes $\gamma = 1/3$, the crack propagation speed is constant and the injection flux rate increases with time: $q_0(t) = O(t^{1/3})$ for $t \to \infty$.

Note that self-similar solutions do not necessarily satisfy the initial conditions (17) as the normalised initial crack lengths are:

$$L(0) = \sqrt{\frac{2u_0^3}{9\gamma}}, \quad L(0) = \sqrt{\frac{2u_0^3a^{3\gamma+1}}{3(3\gamma+1)}},$$

for the first and the second type, respectively.

Remark 2 As one can see the second type of similarity solution has a physical sense for any $-1/3 < \gamma < \infty$ and thus, can be used to model three different transient regimes of the crack evolution: crack acceleration ($\gamma > 1/3$), crack deceleration ($\gamma < 1/3$), and a steady-state propagation of the fracture ($\gamma = 1/3$). The first type of solution possesses a physical interpretation only for positive values of γ, which restricts its application to the cases of accelerating crack.

The self-similar solutions formulated above are used in the following sections to analyse computational accuracy provided by the developed solvers. So far to this end, the asymptotic models have been usually

employed (Nordgren 1972; Kemp 1989; Kovalyshen and Detournay 2009; Linkov 2011c). However, all of them are restricted to the case of a constant influx, q_0.

2 Numerical solution of the self-similar problem

In this section we will formulate an algorithm of the solution for the self-similar problem defined by Eq. (26) and the boundary conditions (27). The following representation of the sought function $u(x)$ will be accepted:

$$u(x) = u_0(1-x)^{1/3} + \Delta u(x). \tag{32}$$

It results from the asymptotic behaviour (18) and $\Delta u(x) = O((1-x)^{\zeta})$ for $x \to 0$. Parameter $\zeta > 1/3$ depends strongly on the behaviour of the leak-off function q_l near the crack tip. In particular, when q_l vanishes near the crack tip in the same manner as the solution, or faster, ($\eta \geq 1/3$) then $\zeta = 4/3$. One can show that [(compare (3)]

$$\zeta = \min\{4/3, 1+\eta\} \geq 1/2, \tag{33}$$

see also Kusmierczyk et al. (2012) for details.

2.1 Integral solver for the self-similar problem

Below we present an algorithm to solve Eq. (26) by numerical inversion of the operator $\mathcal{A}$. Exploiting the solution representation (32), the inverse operator $\mathcal{A}^{-1}$ defines both components: u_0 and Δu. To derive $\mathcal{A}^{-1}$, we integrate the Eq. (26) twice over the interval $[x, 1]$ taking into account the boundary condition $(27)_2$. Then, after simple transformations one obtains:

$$\begin{aligned} 3u_0^3(1-x)\Delta u = &-\frac{3}{4}\Big[6u_0^2(1-x)^{2/3}(\Delta u)^2 \\ &+4u_0(1-x)^{1/3}(\Delta u)^3 + (\Delta u)^4\Big] \\ &+u_0^3\int_x^1 \Delta u d\xi + 2u_0^3\int_x^1 (\xi - x)u d\xi \\ &-(1-x)u_0^3\int_x^1 u d\xi \\ &+\beta u_0^3\int_x^1 (\xi - x)(u + q_l^*)d\xi. \end{aligned} \tag{34}$$

In short, the latter can be symbolically written in the compact form:

$$\Delta u = G_1(\beta, u_0, \Delta u) + G_2(\beta, u_0, \Delta u, q_l^*), \tag{35}$$

where the operators involved in the right-hand side are defined as follows:

$$\begin{aligned} 3u_0^3(1-x)G_1 = &-\frac{3}{4}\Big[6(u_0^2(1-x)^{2/3}(\Delta u)^2 \\ &+4u_0(1-x)^{1/3}(\Delta u)^3 + (\Delta u)^4\Big] \\ &+(2+\beta)u_0^3\int_x^1 (\xi - x)\Delta u d\xi, \end{aligned} \tag{36}$$

$$\begin{aligned} 3(1-x)G_2 = &\, x\int_x^1 \Delta u d\xi + \frac{3}{28}(3\beta - 1)u_0(1-x)^{7/3} \\ &+\beta\int_x^1 (\xi - x)q_l^* d\xi. \end{aligned} \tag{37}$$

One can conclude from (3) and (18) that for $x \to 1$

$$G_1 = O\left((1-x)^{1+\zeta}\right), \quad G_2 = O\left((1-x)^{\zeta}\right), \tag{38}$$

where ζ has been defined in (33).

The relation to compute u_0 is derived by integration of (26) with respect to x from 0 to 1. Then, taking into account the representation (32) and the boundary condition $(27)_1$, one can formulate the condition:

$$G_3(\beta, u_0, \Delta u, q_0^*) = 0, \tag{39}$$

where

$$\begin{aligned} G_3 = &\frac{3}{4}(\beta + 1)u_0^{5/2} \\ &+u_0^{3/2}\left[(\beta+1)\int_0^1 \Delta u dx + \beta\int_0^1 q_l^* dx\right] - q_0^*. \end{aligned}$$

It is easy to prove that for any $q_0^* > 0$ and $\beta > -1$ there exists a unique positive solution u_0 of equation (39), regardless of the values of the functions Δu and q_l^*.

The inverse operator $\mathcal{A}^{-1}$ is defined, by equations (35) and (39), as:

$$[u_0, \Delta u] = \mathcal{A}^{-1}(\beta, q_l^*, q_0^*). \tag{40}$$

Its numerical execution is based on the following iterative algorithm:

$$\begin{cases} G_3\left(\beta, u_0^{(i+1)}, \Delta u^{(i)}, q_0^*\right) = 0, \\ \Delta u^{(i+1)} = G_1\left(\beta, u_0^{(i+1)}, \Delta u^{(i)}\right) \\ \qquad\qquad +G_2\left(\beta, u_0^{(i+1)}, \Delta u^{(i)}, q_l^*\right), \end{cases} \tag{41}$$

where superscripts refer to the consequent iterations. In the first step we assume that:

$$\Delta u^{(0)} = 0. \tag{42}$$

Of course, if any better approximation is available, it can replace (42). Note the first relation of (41) is a nonlinear algebraic equation that can be solved e.g. by the Newton-Raphson method, while the second equation is a typical iterative relationship.

2.2 Numerical examples and discussions

In this section we investigate the performance of the numerical algorithm (41)–(42). To this end, the power law type self-similar solution (29) is utilized as a benchmark. Apart from the fact that the equations for other self-similar solution (24) look identically, the value of the parameter β appearing in the exponential benchmark is always the same ($\beta = 2/3$). Thus the power law type self-similar benchmark gives us an opportunity to manipulate with the value of β, which will be crucial for further implementation of the algorithm to the transient regime.

Let us utilize the following variant of benchmark solution used previously in Mishuris et al. (2012) [(compare (63) in the "Appendix"]:

$$\begin{aligned} u(x) &= (1-x)^{1/3}(1+s(x)),\\ s(x) &= -\frac{1}{8e}\left(\frac{1}{3} - \frac{2\gamma}{3\gamma+1}\right)(1-x) + 0.05(1-x)^2, \end{aligned} \tag{43}$$

which yields $\eta = 4/3$.

Computations are carried out for different meshes based on $N+1$ nodal points:

$$x_j^{(\varrho)} = 1 - \left(1 - \frac{j}{N}\right)^{\varrho}, \quad j = 0, 1, \ldots, N. \tag{44}$$

When one sets $\varrho = 1$, the points of spatial mesh are uniformly distributed over the whole interval—this mesh will be called the *uniform mesh*. By taking $\varrho > 1$ we obtain increased mesh density while approaching the end point $x = 1$ (the larger ϱ the greater mesh density near the crack tip)—this mesh will be refereed to as the *non-uniform mesh*. In our computations we will use $\varrho = 3$. This choice is motivated by the following reason. To increase the solution accuracy, we compute integral operators from (41) using the classic Simpson quadrature, which gives an error controlled by the fourth derivative of the integrand. Accounting for the asymptotic behaviour of the solution, the transformation (44), for $\varrho = 3$, improves the smoothness of the integrand with respect to the new independent variable near the crack tip (replaces the fraction powers function). In this way, the error of integration can be minimized. We shall confirm this below in numerical tests.

We use two parameters as the measures of computational accuracy: (1)—the maximal point-wise relative error (denoted as: δu) of the complete solution $u(x)$ and (2)—relative error (denoted by δu_0) of the coefficient u_0 defining the leading asymptotic term in (32).

The efficiency of computations will be assessed by the number of iterations needed to compute the solution. The iterative process is stopped in each case when the L_2-norm of the relative difference between two consecutive approximations becomes smaller than $\epsilon = 10^{-10}$. Finally, we analyze how the solution errors depend on the number of nodal points N.

The computations revealed that convergence of the iterative process (41) may only be achieved for some range of β values. For the analyzed benchmark it was: $\beta \in [-1.8, 4.8]$. In fact, this interval is wider than one could expect and fully covers any physically motivated values of $\beta \in (0, 2/3)$ following from the self-similar formulation. Interestingly, there are also solutions for $\beta < -1$ [compare discussions after Eq. (39)]. This information shall be used later on, to construct a solver for transient regimes.

In Fig. 1a the values of δu and δu_0 obtained for the analyzed benchmark (43) are shown. The computations were done for the meshes composed of 100 nodes. As can be seen, the non-uniform mesh gives at least one order better accuracy of the solution, for the same number of nodal points. The error of u_0 is lower than the error of the complete solution $u(x)$, as one could expect, while its distribution is not as smooth as for $\delta u(x)$. The minimum of $\delta u(x)$ is located near $\beta = 1/3$ which corresponds to the steady-state similarity solution. Interestingly, in the case of the uniform mesh the aforementioned minimum is deeper and sharper than for the non-uniform one.

Figure 1b depicts the number of iterations needed to obtain the final solution for different values of β. It shows that the convergence rate almost does not depend on the type of mesh chosen. The best efficiency of computations appears for approximately $|\beta| < 1/2$. Note that this interval corresponds to the values of β which

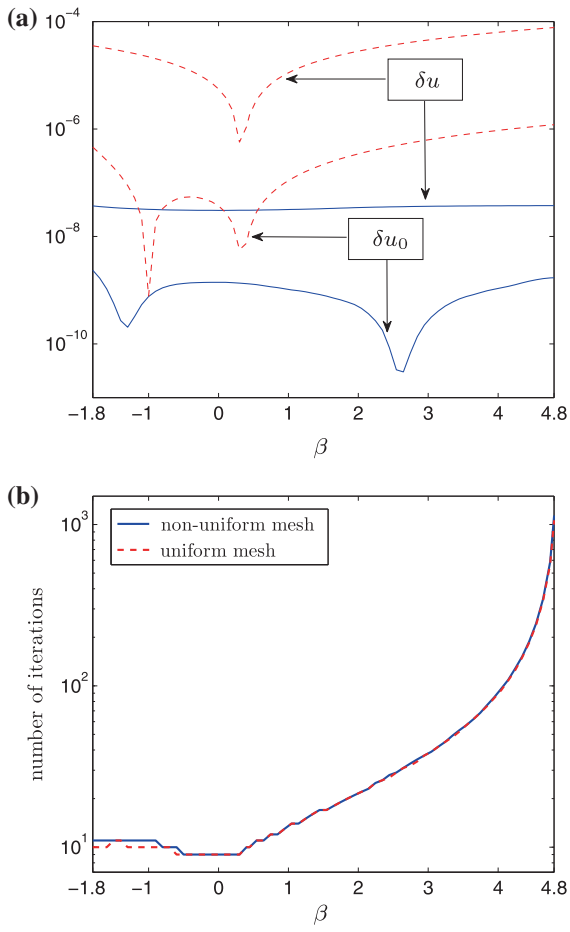

Fig. 1 (**a**) The accuracy of computations: maximal relative error of $u(x)$ and u_0 as a function of β. (**b**) Number of iterations to reach the final solution as a function of β. *Dashed lines* refer to the uniform mesh, *solid lines* to the non-uniform mesh (for $\varrho = 3$)

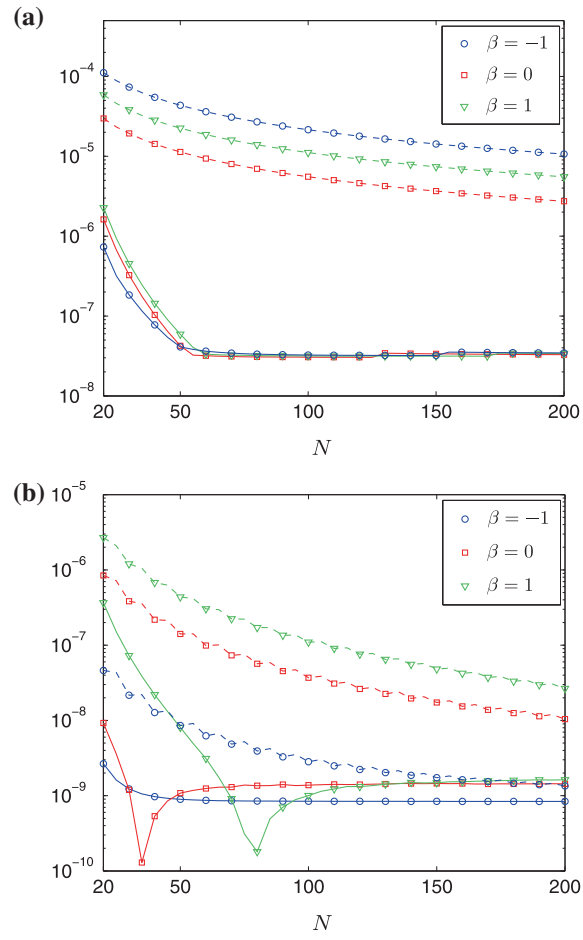

Fig. 2 The maximal relative error of solution as a function of the number of nodal points N. (**a**) δu, (**b**) δu_0. *Dashed lines* refer to the uniform mesh, *solid lines* to the non-uniform mesh (for $\varrho = 3$)

provide the best accuracy of computations. The time of computations (number of iterations) increases with $|\beta|$ growth, however this trend does not exhibit a bilateral symmetry (for negative and positive values of β). Moreover, the cost of computations increases with the solution error.

Figure 2 shows how the accuracy of computations depends on the number of nodal points N. Both chosen meshes are analyzed for three different values of $\beta = -1, 0, 1$. It can be seen that the non-uniform mesh gives at least two orders of magnitude better accuracy than the regular one. For both types of meshes, the values of δu_0 are much lower than respective δu, however the best result for δu_0 does not necessarily correspond to the best δu (e.g. $\beta = -1$ for the regular mesh). The non-uniform mesh provides much lower sensitivity of solution accuracy to the variation of β than the uniform one. Also the maximal level of accuracy is obtained much faster for the non-uniform mesh. In the considered case, it is sufficient to take only 60 nodal points to achieve the maximal possible accuracy.

The last test in this subsection identifies the influence of spatial meshing on the solution accuracy. To this end, we consider the following values of $\varrho = 1, 2, \ldots, 5$ from representation (44). The benchmark taken here accepts $\beta = 1/3$, as it provides the best accuracy and will be important in next subsections. For each of the values of ϱ, a characteristic $\delta u(N)$ was computed. The results are presented in Fig. 3. It shows that, regardless of the mesh under consideration (or

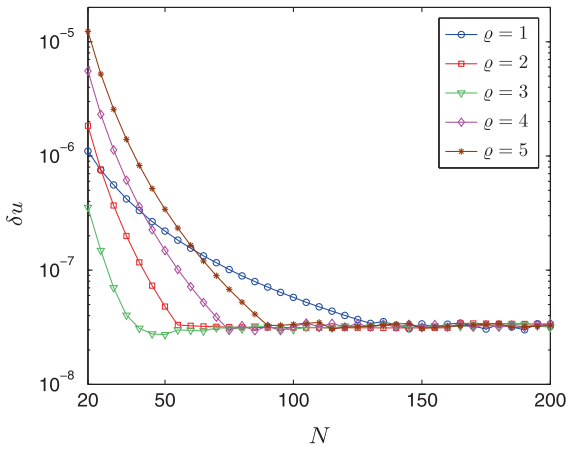

Fig. 3 The maximal relative error of solution δu_0 for various spatial meshes. Computations were done for $\beta = 1/3$

equivalently, the value of the parameter ϱ), the maximal achievable accuracy is the same. However, this ultimate level is reached for different numbers of nodal points, N. The fastest convergence takes place for $\varrho = 3$, which confirms our previous predictions on the optimal choice of the spatial meshing. The slowest convergence to the saturation level manifests the uniform mesh ($\varrho = 1$).

The overall influence of the value of the parameter ϱ on the accuracy of computations results from the following trend: the larger the value of ϱ the lower error near the crack tip and the greater error near the crack inlet. The optimal balance between the local errors is observed for $\varrho = 3$, which confirms our predictions. In the following, only the non-uniform mesh for $\varrho = 3$ will be used in computations.

3 Solution to the transient problem

In this section we adopt the idea of the integral solver, developed for the self-similar formulation, to the transient regime. The basic assumptions of the approach remain the same, however the algorithm has to be modified in some essential aspects. First of all, one has to build the mechanism of temporal derivative approximation into the numerical procedure together with necessary measures to stabilize the algorithm. We will propose two methods of doing this, constructing in fact two different solvers. The second fundamental difference between the self-similar and time-dependent formulations is that in the latter case, the crack length $L(t)$ becomes now an element of the solution, which should be looked for simultaneously with the crack opening $w(t, x)$ and the first term of its asymptotics near the crack tip $w_0(t)$.

The basic system of equations for the transient problem is composed of: the governing equation (22), the boundary conditions (16), the initial conditions (17) and the integral equation defining the crack length (21).

To avoid using multiple subscripts let us adopt the following manner of notation:

$$w(t_j, x) = w(x), \quad w(t_{j+1}, x) = W(x). \tag{45}$$

Consequently, the asymptotic representations of the solution near the crack tip read:

$$\begin{aligned} w(x) &= w_0(1-x)^{1/3} + \Delta w, \\ W(x) &= W_0(1-x)^{1/3} + \Delta W, \end{aligned} \quad x \to 1, \tag{46}$$

where ζ is defined in (33) and

$$\begin{aligned} &\Delta w = O((1-x)^{\zeta}), \\ &\Delta W = O((1-x)^{\zeta}), \quad \text{as} \quad x \to 1. \end{aligned}$$

3.1 Solver using self-similar algorithm (40)–(41)

Below, we show the way to convert the initial boundary value problem defined by equations (16), (21), (22) to the form which may be tackled by the integral solver in the form (41) for the self-similar solution. Obviously, system (41) is to be supplemented with an additional equation defining the crack length.

The main idea of the approach is to use the temporal derivative as one of the dependent variables in the numerical procedure. To achieve this, we compute the derivative of the solution at each time step $t = t_{j+1}$ in the following iterative process:

$$\begin{aligned} \frac{\partial W^{(i+1)}}{\partial t} &= G_4\left(\sigma^{(i+1)}, W^{(i+1)}, \frac{\partial W^{(i)}}{\partial t}\right) \\ &\equiv \sigma^{(i+1)} \frac{W^{(i+1)} - w}{\Delta t} + \left(1 - \sigma^{(i+1)}\right) \frac{\partial W^{(i)}}{\partial t}, \end{aligned} \tag{47}$$

where superscripts refer to the number of iteration, $\Delta t = t_{j+1} - t_j$ and the values of $\sigma^{(i+1)}$ are to be defined later. The first approximation of the temporal derivative is

$$\frac{\partial W^{(1)}}{\partial t} = \frac{\partial w}{\partial t}. \tag{48}$$

Note that the derivative at initial time $t = 0$ can be immediately found from the governing equation (22) by substitution of the initial conditions.

Substituting (47) into (22) one obtains the problem (26) and (27) with respect to the unknown function $W(x)$ by exploiting the following convention:

$$q_l^* = -w + \frac{\Delta t}{\sigma^{(i+1)}}\left[(1-\sigma^{(i+1)})\frac{\partial W^{(i)}}{\partial t} + q_l(t,x)\right],$$

$$\beta W_0^3 = \frac{3\sigma^{(i+1)}}{\Delta t}\left(L^{(i+1)}\right)^2, \tag{49}$$

and

$$q_0^*(t) = 3q_0(t)W_0^{-3/2}L^{(i)}. \tag{50}$$

Finally, the crack length from equation (21) can be iterated as:

$$L^{(i+1)} = G_5(W_0) \equiv \sqrt{\left(L^{(i)}\right)^2 + \frac{\Delta t}{3}\left(W_0^3 + w_0^3\right)}. \tag{51}$$

Here the integral from (21) is computed by the trapezoidal rule, taking into account the information obtained from the previous time steps ($t \le t_j$).

In this way the governing partial differential equation and all the boundary conditions have been transformed to the formulation used previously in the self-similar problem. As a result, respective integral operators (34)–(37) and (39) remain the same with modified arguments (49) and (50), where (51) should be taken into account.

To choose the value of the parameter $\sigma^{(i+1)}$ at each iterative step, let us recall that the best performance of the algorithm described in Sect. 2.2 has been achieved near $\beta = 1/3$. Taking this fact into account in $(49)_2$, one can choose

$$\sigma^{(i+1)} = \frac{\Delta t\left(W_0^{(i+1)}\right)^3}{9\left(L^{(i+1)}\right)^2}. \tag{52}$$

Thus, the inverse operator for $\mathcal{A}$, defining the solution of the transient problem at the time step $t = t_{j+1}$, can be expressed in the following manner [compare (40)]):

$$\left[W_0, L, \Delta W, \frac{\partial W}{\partial t}\right] = \mathcal{A}^{-1}\left(1/3, q_l^*, q_0^*\right). \tag{53}$$

The iterative algorithm of the solver can be described by the system:

$$\begin{cases} G_3\left(1/3, W_0^{(i+1)}, \Delta W^{(i)}, q_0^*\right) = 0, \\ \Delta W^{(i+1)} = G_1\left(1/3, W_0^{(i+1)}, \Delta W^{(i)}\right) \\ \quad + G_2\left(1/3, W_0^{(i+1)}, \Delta W^{(i)}, q_l^*\right), \\ L^{(i+1)} = G_5\left(W_0^{(i+1)}\right), \\ \dfrac{\partial W^{(i+1)}}{\partial t} = G_4\left(\sigma^{(i+1)}, W^{(i+1)}, \dfrac{\partial W^{(i)}}{\partial t}\right), \end{cases} \tag{54}$$

Note that parameters $q_0^* = q_0^{*(i)}$, $q_l^* = q_l^{*(i+1)}$ and $\sigma^{(i+1)}$ are also iterated according to $(49)_1$, (50) and (52). To finalize the algorithm, it is enough to define the initial forms of the crack opening $W^{(1)}$ and the crack length $L^{(1)}$. This is achieved by choosing

$$W^{(1)} = w + \frac{\partial w}{\partial t}\Delta t, \quad L^{(1)} = L(t_i), \tag{55}$$

which finishes the description of the computation scheme for the fixed time step $\Delta t = t_{j+1} - t_j$. Let us recall that the temporal derivative of solution at time $t = 0$ is computed using the initial conditions (17), while for any next $t = t_j$ we utilize representation (47).

We would like to underline here the fact that as the output of the proposed algorithm, one obtains not only the solution of the transient problem, that is, the crack length $L(t)$ and the crack opening $w(t, x)$, but also the temporal derivative of the latter, $w_t'(t, x)$ and the crack propagation speed, $V_0(t)$, from (20).

3.2 Solver based on a modified self-similar algorithm

Temporal derivative of the solution can be taken in the following form:

$$\frac{\partial W}{\partial t} = 2\frac{W - w}{\Delta t} - \frac{\partial w}{\partial t}, \tag{56}$$

which gives the error of approximation of the order $O(\Delta t^2)$. Note that any other two-points finite difference definition yields only $O(\Delta t)$.

Unfortunately, a direct use of the algorithm formulated in Sect. 2.1 is not, generally speaking, possible, as it may fail for small time steps. Indeed if one takes $\sigma = 2$ and sufficiently small value of Δt in the representation $(49)_2$, then the value of the parameter β may be far away from its optimal magnitude $\beta = 1/3$. This in turn, would at least result in a deterioration of the efficiency of computations.

On the other hand, it is quite clear that BVP (26), (27) is solvable for large values of the parameter β.

Indeed, performing asymptotic analysis, one shows that the solution can be represented in the form:

$$u(x) = -q_l^*(x) + b_0(x) + b_1(x), \quad x \in (0,1), \quad (57)$$

where b_0 and b_1 refer to two boundary layers, accounting for the boundary conditions $(27)_1$ and $(27)_2$, respectively.

Thus the algorithm should be modified to be able to deal with large values of β. To achieve this goal, let us look at the original representation (34), where only the last term in the right-hand side depends on β. This term violates the convergence of the iterative process for large β. To prevent this from happening, at each iteration we supplement the term in question with an auxiliary so-called 'viscous' term in the form $\mathcal{V}(x) = \beta u_0^3(1-x)(C_0 + C_1(1-x))$, where the constants, C_0, C_1, are computed by comparing both the original and the viscous terms. In this way one can construct a *modified* algorithm, schematically represented in the following manner

$$\begin{cases} G_3\left(\beta^{(i)}, W_0^{(i+1)}, \Delta W^{(i)}, q_0^*\right) = 0, \\ \mathcal{V}^{(i+1)} = \mathcal{V}(W^{(i)}, \hat{q}_l^{*(i+1)}), \\ \Delta W^{(i+1)} \\ \quad = \hat{G}_{12}\left(\beta^{(i+1)}, W_0^{(i+1)}, \Delta W^{(i)}, \hat{q}_l^*, \mathcal{V}^{(i+1)}\right), \\ L^{(i+1)} = G_5\left(W_0^{(i+1)}\right), \\ \dfrac{\partial W^{(i+1)}}{\partial t} = 2\dfrac{W^{(i+1)} - w}{\Delta t} - \dfrac{\partial w}{\partial t}, \end{cases} \quad (58)$$

where

$$\hat{q}_l^* = -w - \frac{\Delta t}{2}\left[\frac{\partial w}{\partial t} + q_l(t,x)\right],$$

$$\beta^{(i)} = \frac{6}{\Delta t}\left(L^{(i)}\right)^2\left(W_0^{(i)}\right)^{-3}. \quad (59)$$

The function $q_0^*(t)$ is defined in the same way as previously [see (50)]. At each time step the initial value of W is taken in the form (55). We do not show here the precise definition of the operator $\hat{G}_{12}$, as it may be easily derived by merging operators G_1 and G_2 with the viscous term $\mathcal{V}(x)$, where respective constants are computed, for example by the least squares method.

3.3 Analysis of the algorithms performance

The aim of this subsection is to analyze and compare the performances of the algorithms formulated in Sects. 3.1 and 3.2. To this end, the benchmark solution used previously for the self-similar formulation is utilized, for the time dependent term $\psi(t) = (1+t)^\gamma$ [compare (63)]. First tests are performed for $\gamma = 1/5$.

In the following, the notations *solver 1* and *solver 2* are attributed to the algorithms (54) and (58), respectively.

Let us analyze the influence of the spatial mesh density on the accuracy of computations done by both solvers. For this reason, we start with a single time step (in our case $\Delta t = 10^{-2}$) and carry out the computations for different numbers of nodal points N, ranging from 10 to 100. The following parameters are used for the comparison: the maximal relative error of the crack opening, δw, the relative error of the crack length, δL, and the maximal relative error of the temporal derivative of the crack opening, δw_t.

The results of computations are depicted in Fig. 4a. It shows that regardless of the considered parameter, *solver 2* always provides better accuracy (two orders of magnitude for the crack opening, w, and an order for the crack length, L). The only exception is for δw_t, whose values are of the same order and *solver 1* may even give a bit lower errors. However, this does not result in a better accuracy of the two remaining components of the solution: w and L. Note that for the first time step, the accuracy of computation of the crack length, $L(t)$, is of two orders of magnitude better than that for the crack opening, δw, regardless of the solver type and the number of the nodal points, N.

Similarly to the trend observed in the self-similar formulation, computational errors stabilize for some critical value $N = N_*(\Delta t)$. Surprisingly, for the transient regime where the temporal derivative plays a crucial role, the critical N_* appears to be even slightly lower than that for similarity solution. Thus, for both solvers it is sufficient to take only 40 points to achieve the maximal level of accuracy.

Remark 3 From first glance it may be surprising that the accuracy of the temporal derivative, δw_t, is up to four orders of magnitude worse than that for δw in the case of *solver 2*. However, this fact can be easily explained when one analyses the estimation: $\delta w_t \approx 2w\delta w/(\Delta t w_t)$ which follows from (56) for small Δt. Computing the multiplier in the right hand side of the estimation, one obtains the respective value of the order of 10^4. Note that the recalled formula does not account for the error of the method of w_t approximation itself. In some cases (very small time steps) this value may be

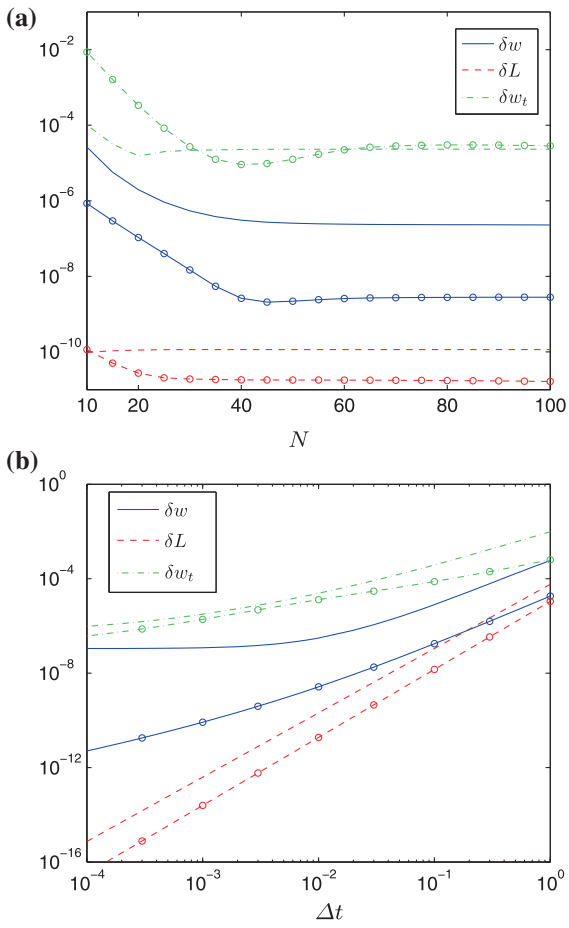

Fig. 4 The errors of computations: $\delta L, \delta w, \delta w_t$ as functions of (**a**) spatial mesh density (number of nodal points, N) for the fixed time step $\Delta t = 10^{-2}$. (**b**) time step Δt for fixed number of the nodal points $N = 40$. *Line with markers* refer to *solver 2*

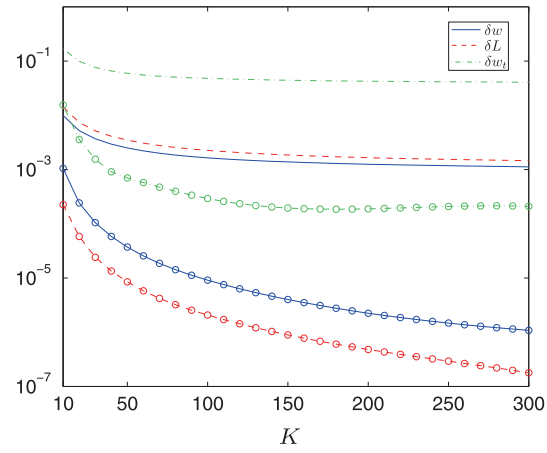

Fig. 5 Influence of the number of time steps, K, within time stepping strategy (60), on accuracy of the computations for the fixed non-uniform spatial mesh ($N = 40, \varrho = 3$). *Lines without markers* correspond to *solver 1*, *lines with markers* refer to *solver 2*

comparable to the former, and thus essentially influence the overall error.

Remark 4 At the first time step, there also exists a direct relationship between the error of the crack length, δL, and the error of multiplier of the leading term of the asymptotics (46), δW_0. For small values of Δt it reads: $2\delta L = W_0^3 \delta W_0 \Delta t$. Having this relation we do not show a separate analysis for W_0 and consequently, V_0.

All the results presented above were obtained for a single time step $\Delta t = 10^{-2}$. In order to illustrate the influence of Δt on the solution accuracy, a number of computations were done in the interval $\Delta t \in [10^{-4}, 1]$ for a fixed number of nodal points, $N = 40$. The results describing the solution errors are shown in Fig. 4b. One can conclude from them that the errors decrease when reducing the time step (at least in the analyzed range of parameters). Simultaneously, one can expect that this tendency should have its own limitation for a fixed number of the nodal points, N, and starting from some small value $\Delta t(N)$, it will reverse to the opposite. For both solvers, a fast decrease of δL is observed as Δt gets smaller ($\delta L \approx a 10^{-4} \Delta t^3$ and $a \sim 1$).

Now let us analyze the accuracy of the solution on a given time interval $t \in [0, t_K]$. As, it follows from the results presented in Kusmierczyk et al. (2012) and Mishuris et al. (2012), after the initial growth to the maximal value, the error of computations stabilizes at some level or even decreases, and thus further extending of the time interval does not contribute to the deterioration of accuracy. Bearing this in mind we take here $t_K = 100$. First we test how the number of time steps K affects the accuracy of computations within the same time stepping strategy. For both solvers we took a fixed number of spatial mesh points, $N = 40$, changing the number of time steps K from 10 to 300. The utilized time stepping strategy was the same as that used in Mishuris et al. (2012) for the Crank-Nicolson scheme. It is given by the following equation ($i = 1, 2, \ldots, K$):

$$t_i = (i-1)\delta t + \frac{t_K - (K-1)\delta t}{(K-1)^3}(i-1)^3, \qquad (60)$$

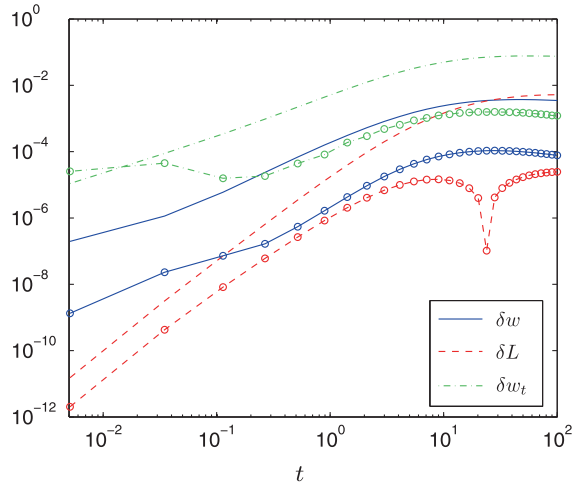

Fig. 6 Distribution of the solution errors in time. *Line with markers* refer to *solver 2*. The time stepping strategy (60) is used for $K = 30$. The number of nodal points, $N = 40$

where δt is a parameter controlling the first time step. Note that by increasing the value of K one distributes the time points near $t = 0$ almost uniformly.

The achieved accuracy, as a function of the number of time steps K, is presented in Fig. 5. It shows that for the same value of K *solver 2* provides much more accurate results again and this advantage increases with growing K. One can expect, as it follows from the discussions after Fig. 4 which refers to a single time step, that each solver gives the solution errors satisfying the estimation: $\delta L << \delta w << \delta w_t$ for any set of the input parameters. Indeed, *solver 2* supports this statement as can be seen Fig. 5, while for *solver 1* surprisingly another trend is observed: $\delta w < \delta L << \delta w_t$.

To explain this apparent paradox, let us recall that all errors in Fig. 5 are taken as the maximal values over the time-space domain. However, since for the assumed time stepping strategy the time steps increase with growing time, one can expect (compare the results in the Fig. 4b) that at some point δL may become greater than δw. Moreover, the error accumulation in successive time steps may additionally influence the relationship between δL and δw, especially if they are close to each other as it is in the case of *solver 1*. Figure 6 depicts the evolution of computational errors in time. For both solvers δw and δw_t reach their maximal values and stabilize or decrease for $t < t_K$. In the case of *solver 1* one can observe an intensive error accumulation for the crack length, δL. Indeed, it can be seen from the Fig. 6, that there exists a moment when the error curves for δL and δw intersect for the *solver 1*. In the case of *solver 2*, respective errors of solution (δL, δw and δw_t) have essentially different values. As a result, this effect does not take place in the time interval under consideration. However, it may be encountered for $t_K > 100$, as suggests the trend observed for t close to t_K.

It is interesting to compare the solution errors generated by the two integral solvers with those obtained in Mishuris et al. (2012) for a DS approach. In the latter case a standard MATLAB solver, *ode15s*, was employed. The best results were obtained for a uniform spatial mesh, the presented data corresponded to $N = 100$. The time stepping strategy chosen automatically by the solver [whose character is approximated by (60)] accepted 242 time steps. The maximal relative error of the crack opening, δw, and the crack length, δL, were 5.0×10^{-3} and 6.8×10^{-3} (see Table 1), respectively.

When analyzing the data collected in Table 1, one concludes that it is sufficient for any of the integral solvers to take only $N = 40$ nodal points and $K = 30$ time steps, as suggested the previous analysis, to have better (but comparable—*solver 1*), or much better (*solver 2*) results. Indeed, the corresponding maximal errors for the integral solvers are: $\delta w = 3.7 \times 10^{-3}$, $\delta L = 5.2 \times 10^{-3}$ for *solver 1*, and $\delta w = 1.1 \times 10^{-4}$, $\delta L = 2.4 \times 10^{-5}$ for *solver 2*. In other words, the first solver provides the same accuracy for the crack opening and the crack length as the DS solver using much greater numbers of nodal points and time steps, while the second one, under the same conditions, improves the results at least one order of magnitude. It also shows that the *solver 2* yields one order of magnitude better accuracy of the crack propagation speed, V_0, than the *solver 1*.

The new algorithms allow us to automatically compute the temporal derivatives in the solution process. The respective errors, δw_t, are: 7.5×10^{-2}—*solver 1* and 1.5×10^{-3}—*solver 2*. We decided to compare these figures, with the ones obtained in postprocessing (here, also the DS approach was examined). To this end two FD schemes (2- and 3-points) were used. This time, the corresponding errors, δw_t, were: 4.6×10^{-2} and 2.0×10^{-2} for *solver 1*, 4.6×10^{-2} and 3.4×10^{-3} for *solver 2* and the same value 4.1×10^{-1} for both schemes in the case of DS. It is worth mentioning that the values obtained for DS appeared at the first time step. Then, the errors decreased with time and stabilized to give

Table 1 Comparison of the results obtained for the solver developed in Mishuris et al. (2012) and two integral solvers: *solver 1* and *solver 2*. The same benchmark solution with the leak-off vanishing near the crack tip is considered

Solver	N	K	δL	δw	δV_0	δw_t	$\delta w_t^{(FD_2)}$	$\delta w_t^{(FD_3)}$
From Mishuris et al. (2012)	100	242	6.8×10^{-3}	5.0×10^{-3}	–	–	4.1×10^{-1} (1.0×10^{-2})	4.1×10^{-1} (4.8×10^{-3})
Solver 1	40	30	5.2×10^{-3}	3.7×10^{-3}	7.1×10^{-3}	7.5×10^{-2}	4.6×10^{-2}	2.0×10^{-2}
Solver 2	40	30	2.4×10^{-5}	1.1×10^{-4}	3.2×10^{-4}	1.5×10^{-3}	4.6×10^{-2}	3.4×10^{-3}
Solver 1	5	30	5.2×10^{-3}	3.7×10^{-3}	7.0×10^{-3}	7.5×10^{-2}	4.1×10^{-2}	2.0×10^{-2}
Solver 2	5	30	8.0×10^{-5}	5.7×10^{-4}	6.3×10^{-4}	5.6×10^{-2}	6.3×10^{-2}	6.2×10^{-2}

their minimal levels of 10^{-2} and 4.8×10^{-3}, correspondingly.

As can be seen, the integral solvers give at least one order of magnitude better accuracy of w_t than the DS. Moreover, while the postprocessing gives smaller (but comparable) error for the *solver 1*, *solver 2* returns more accurate values of w_t than those obtained in the postprocessing, even for the 3-points FD. Finally, apart from the fact that δw and δL for the DS and *solver 1* look comparable in values, the quality of the computation is better for the new solver as is clear from the postprocessing analysis.

Just for comparison we also present in Table 1 the results obtained for the spatial mesh composed of only five nodal points, $N = 5$. It turned out that even for such a drastic reduction of the mesh density, the solution accuracy for most of the parameters is of the same order as for $N = 40$. Interestingly, *solver 1* exhibits almost no sensitivity to this mesh reduction. In fact the distinguishable differences can be observed only for the *solver 2*.

Let us now analyze the distribution of crack opening error, δw, in time and space. The respective results are presented in Fig. 7a—*solver 1*, and Fig. 7b—*solver 2*. In both cases the maximal errors are located at the crack tip while the error distribution in time follows the trend visible in Fig. 6.

Figure 8 shows the distributions of δw_t obtained by the integral solvers. It confirms our previous observation, that *solver 2* always provides better results than *solver 1*. Moreover, the greatest error in the case of *solver 1* is located at the crack inlet and the lowest at the crack tip, while *solver 2* gives approximately the same values of δw_t along the crack length.

In the last test in this subsection we analyze the relation between the regimes of crack propagation and the

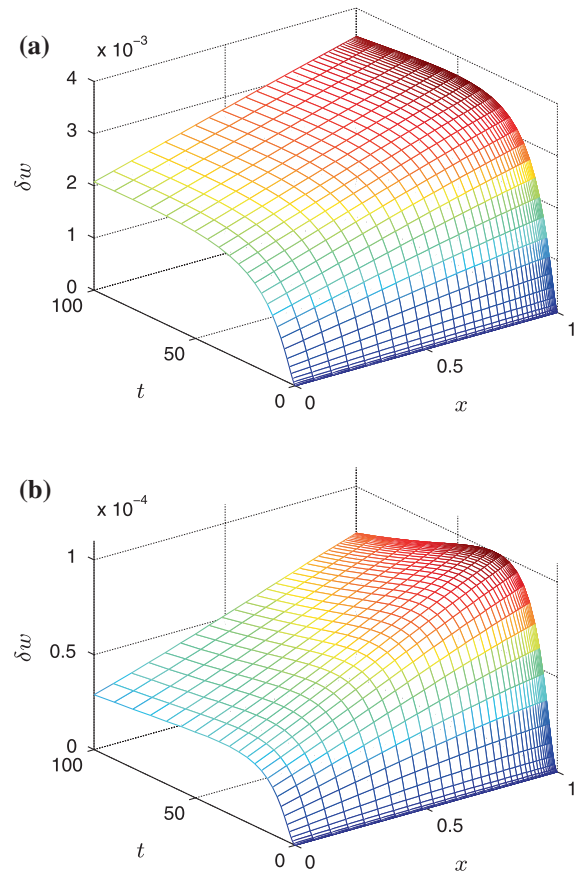

Fig. 7 The relative error of the crack opening obtained for $N = 40$ (nonuniform mesh, $\varrho = 3$) and the time stepping strategy (60) with $K = 30$. (**a**) Corresponds to the *solver 1*, while (**b**) refers to the *solver 2*

performances of respective solvers. As mentioned previously, the benchmark solution in form (29) can be used to imitate various dynamic modes of the crack evolution. So far we have utilized the exponent of the time dependent term of the value $\gamma = 1/5$ which refers

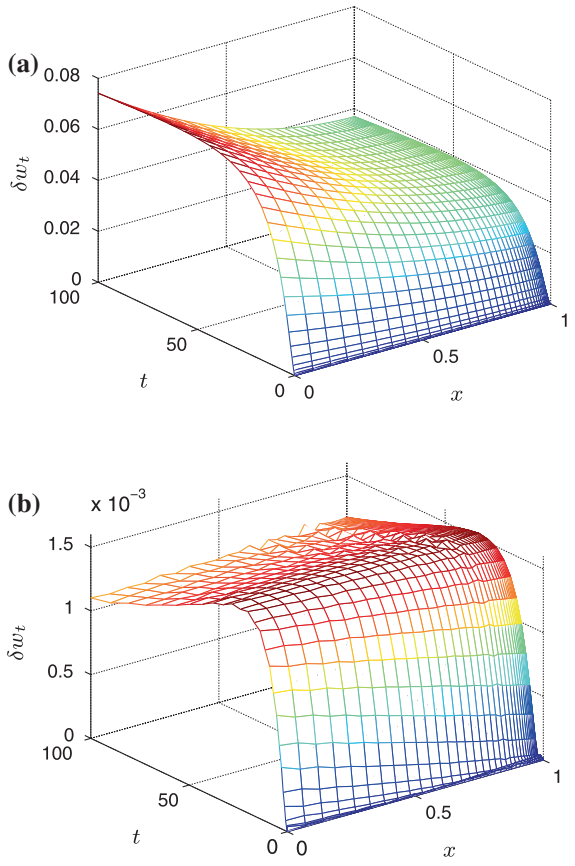

Fig. 8 The relative error of the temporal derivative of the crack opening. Solution obtained by (**a**) *solver 1*, (**b**) *solver 2* for $N = 40$, non-uniform mesh ($\varrho = 3$) and the time step strategy (60) with $K = 30$

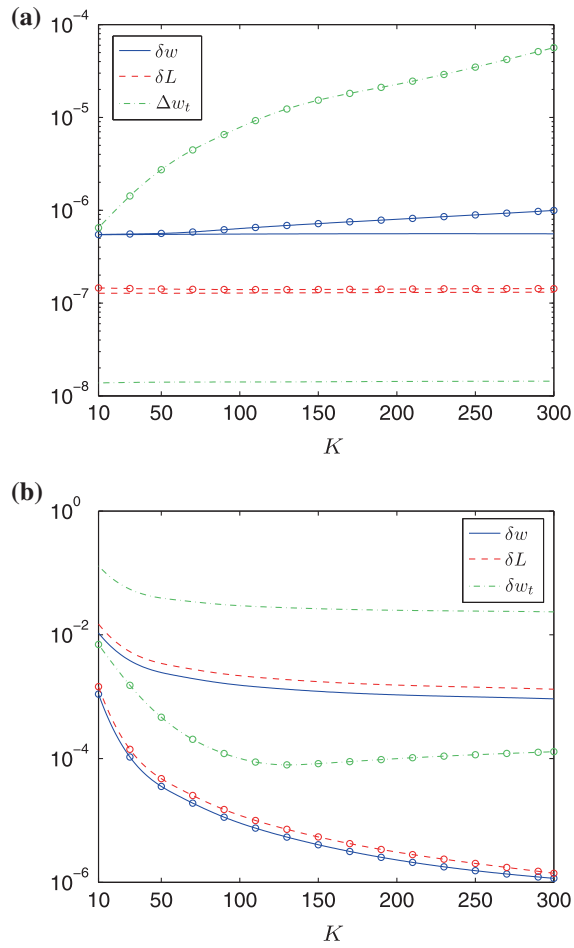

Fig. 9 The errors of solution for two variants of γ. (**a**) $\gamma = 0$, (**b**) $\gamma = 1/3$. *Line with markers* correspond to *solver 2*

to the constant injection flux rate. Now, let us consider two other variants of γ: (1) $\gamma = 0$—for this choice the normalized crack opening is constant in time; (2) $\gamma = 1/3$—this value corresponds to the steady state propagation of the fracture. For the computations the same spatial mesh as before is taken ($N = 40$). The number of time steps accepted within strategy (60), K, ranges from 10 to 300. In this way the graphs (Fig. 9a, b) describing the solution errors in the function of K were prepared for both values of γ (similarly as in Fig. 5). The analyzed accuracy parameters were: relative error of the crack opening, δw, relative error of crack the length, δL, relative error of the crack opening temporal derivative, δw_t, for $\gamma = 1/3$ and absolute error of the crack opening temporal derivative, Δw_t, for $\gamma = 0$.

The results depicted in Fig. 9 show that for $\gamma = 0$ one obtains much more accurate results than for $\gamma = 1/3$ that could have been predicted ($w_t = 0$). However, it is a surprise that for $\gamma = 0$ *solver 1* provides better solution accuracy than *solver 2*. Although the difference is moderate in case of δw and δL, the values of Δw_t vary by at least two orders of magnitude. From Fig. 9a it follows that for this regime of crack propagation the solution accuracy cannot be improved by simple refining the temporal mesh, and for *solver 2* even a reverse relation is observed.

The situation is quite different for $\gamma = 1/3$ (Fig. 9b). This time again *solver 2* proves its advantage over *solver 1* for all the analyzed parameters. For *solver 2*, it is sufficient to take only 30 time steps to have much better results than those provided by *solver 1* for 300 steps. The solution accuracy can be improved by increasing the number of time steps, however it seems that for *solver 1* the saturation level is close to $K = 300$. A similar trend was observed for $\gamma = 1/5$ (see Fig. 5).

A direct conclusion from this test is that for different modes of crack propagation there are different optimal time stepping strategies. This should be properly accounted for especially in the cases when the values of injection flux rate or leak-off to formation change appreciably in the considered time interval.

The aforementioned analysis proves that in terms of accuracy in most of the cases *solver 2* is much better that *solver 1* with respect to all computed components of the solution: the crack length, L, the crack opening, w, its temporal derivative, w_t and the fracture propagation speed, V_0. However, for some regimes of crack propagation (low values of γ) *solver 1* may give comparable or even slightly better results than *solver 2*. The advantage of *solver 1* is better efficiency of computations: the time of computations for this solver was on average one third lower that for *solver 2*.

3.3.1 Example with singular leak-off regime

In (3) we have assumed that the behaviour of the leak-off function near the crack tip can be described by a power law, giving in the worst case a square root singularity. Such a limiting behaviour corresponds to the Carter leak-off model (Carter 1957). As a result, although the leading term of the asymptotic expansion for the crack opening near the fracture tip remains the same, the higher terms change, disturbing the solution smoothness (see Kovalyshen and Detournay 2009). A comprehensive analysis of this case was done in Kusmierczyk et al. (2012) where it was also proved that the deterioration of solution accuracy can be prevented by employing the second asymptotic term in the computational algorithm.

In this subsection we show that the algorithms developed in the paper are capable of tackling this kind of problems *without any additional modifications*. To this end let us consider another benchmark solution $u(t, x)$ [see (63)] defined by the functions:

$$h(x) = (1-x)^{1/3}(1+s(x)), \quad s(x) = \frac{1}{5}(1-x)^{1/6}, \tag{61}$$

with the same as before function $\psi(t)$ for $\gamma = 1/5$.

One can easily check, that the above form of $s(x)$ results in a singular behaviour of q_l, with the leading term of the order $O((1-x)^{-1/2})$ as $x \to 1$. The value of multiplier u_0 in (63) was taken in such a way to make the benchmark comparable with the one used previously, in a sense of an average particle velocity. Indeed, in Mishuris et al. (2012) for a fluid velocity defined as $V = q/w$, a parameter describing its variation along the crack length was introduced:

$$\gamma_v(t) = \left[\max_x (V(x,t)) - \min_x (V(x,t)) \right] \times \left[\int_0^1 V(x,t)dx \right]^{-1}. \tag{62}$$

This parameter reflects indirectly the balance between the flux injection rate and leak-off to formation. It was also shown there that it has a decisive influence on the accuracy of computations (the greater value of γ_v, the greater error of the computations). For the benchmarks with comparable values of γ_v, one can expect similar accuracy of the computations. This trend was also confirmed in Kusmierczyk et al. (2012). In our case, the deterioration of the solution smoothness near the crack tip is an additional factor which contributes to the increase of the computational error.

Note that because of the chosen structure of (63), the value of γ_v is constant in time for all benchmarks considered in this paper. For the benchmark (43) γ_v yields 0.408, while (61) leads to $\gamma_v = 0.411$.

The computations were done for the same spatial ($N = 40$, $\varrho = 3$) and temporal [$K = 30$, strategy (60)] meshes as previously considered. The distributions of the errors δw and δw_t are shown in Figs. 10 and 11, respectively. Comparing these results with those obtained for the finite leak-off regime (see Figs. 7, 8) one can see that the accuracy of computations decreased significantly.

In the case of *solver 1*, the crack opening error, δw, shown in Fig. 10 a) is of one order of magnitude greater than that reported in Fig. 7a. For *solver 2* in turn, δw depicted in Fig. 10b is almost two orders higher than the one obtained previously for q_l vanishing at $x = 1$ (Fig. 7b). A pronounced jump of the error is observed at the crack tip, especially for *solver 2*, which explains the deterioration of accuracy when comparing with the finite leak-off case. However, if one considers the accuracy of the solution away from the crack tip, it is still of the same quality as before.

The same trend can be observed for the temporal derivative of the crack opening, w_t (see Fig. 11). Opposite to the benchmark case (43) with the finite leak-off, the distribution of δw_t for *solver 2* becomes

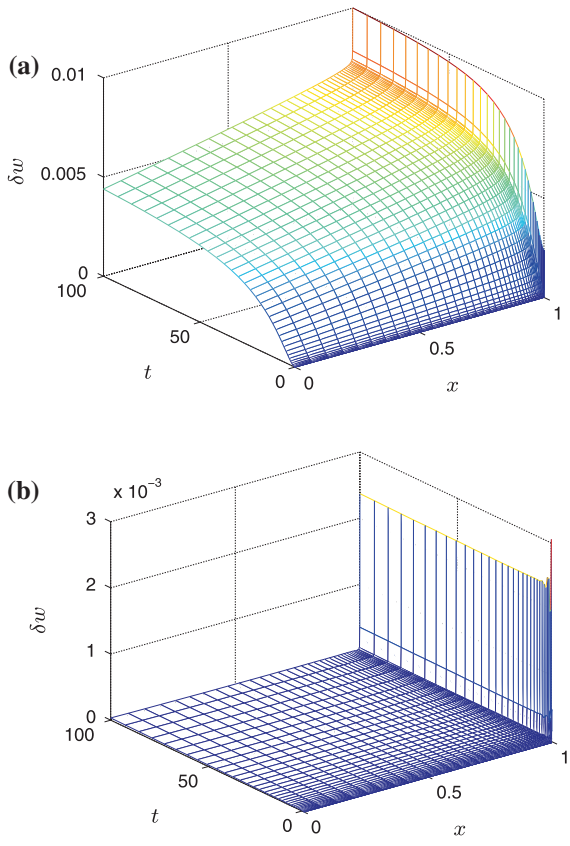

Fig. 10 The relative error of the crack opening. Solution obtained by: (**a**) *solver 1*, (**b**) *solver 2*. Other parameters in the computations were: $N = 40$, $K = 30$, $\varrho = 3$

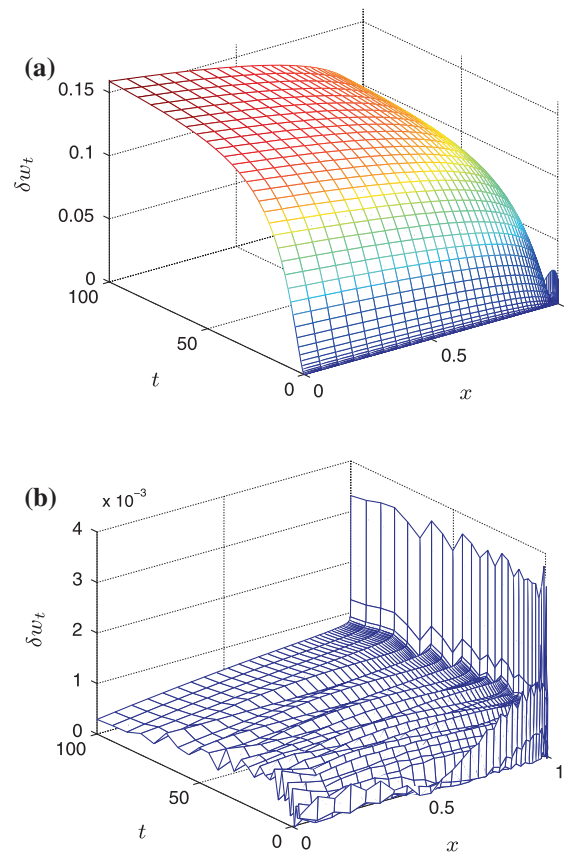

Fig. 11 The relative error of the temporal derivative of crack opening. Solution obtained by (**a**) *solver 1*, (**b**) *solver 2*. Other parameters in the computations are: $N = 40$, $K = 30$, $\varrho = 3$

highly non-uniform, with distinct increase at the crack tip.

This deterioration of the solution accuracy near $x = 1$ for the Carter leak-off model should not be a surprise, as the algorithm described above in (34)–(37) accounted directly only for the first (leading) term of the asymptotic expansion for the crack opening. For the singular leak-off model, the function Δu and other integrands employed in G_1 and G_2 are not sufficiently smooth near the crack tip, which increases the errors of integration. However, one can counteract this tendency by accounting for further asymptotic terms in the algorithm, which will be illustrated in the following with the example of two terms of w expansion.

Let us investigate the evolution of solution errors (δw, δL and δw_t) in time in the way it was done in Fig. 6 for the case of the finite leak-off. In Fig. 12a we show the results obtained by direct execution of the algorithm (34)–(37). Figure 12b depicts the case of a modified algorithm employing also the second asymptotic term of the crack opening. When analyzing the results, one observes again the trend of δL accumulation for both solvers. Interestingly, δL has the same level for both the original and the modified algorithms. In the case of *solver 1* for other analyzed parameters as well only a very slight improvement is reported. However, for *solver 2* one obtains a pronounced reduction of computational errors for w and w_t. To have a further advance here, the next terms of asymptotic expansion of the crack opening may be taken into account. On the other hand, for the Carter leak-off model the number of nodal points for which the accuracy saturation is observed is larger than for a finite leak-off. Thus, if necessary, one can also improve the accuracy by increasing the value of N and by taking a greater number of time steps K.

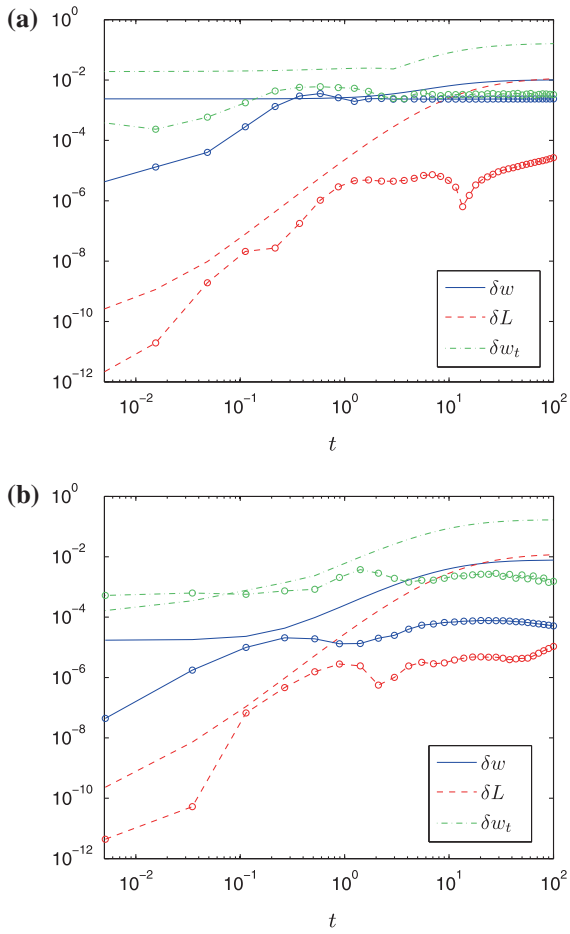

Fig. 12 The relative error of the temporal derivative of crack opening. Solution obtained by (**a**) original algorithms, (**b**) modified algorithms. The *lines with markers* refer to *solver 2*

From the analysis given in this subsection two main conclusions can be drawn. First, that the problems with singular leak-off regimes can be directly tackled by the proposed algorithms. Also in such cases *solver 2* yields more accurate results than *solver 1*. The second conclusion is that in order to have the solution accuracy comparable to that achieved for non-singular leak-off, one has to use a larger number of nodal points and/or employ further terms of asymptotic expansion for w in the algorithm.

4 Conclusions

We would like to itemize the following conclusions as a resume of this paper:

- Presented approach can be efficiently used for tackling the PKN model of hydrofracturing and may be adopted for multifracture systems.
- Both new solvers provide better computational accuracy than the conventional algorithms from Mishuris et al. (2012). Moreover, comparable accuracies can be achieved here at much lower computational cost, as the new solvers enable one to drastically reduce the densities of spatial and temporal meshes.
- New solvers are appropriate for directly tackling the problems with different fluid flow regimes, including various injection flux rates and singular leak-off.
- In order to increase the efficiency and accuracy of computations, it is advisable to employ at least two asymptotic terms of the crack opening, w.
- The developed algorithms do not require any regularization techniques. The boundary conditions are imposed directly into the numerical scheme. The speed equation plays a crucial role in the analysis.

Acknowledgments The authors are grateful to the Institute of Mathematics and Physics of Aberystwyth University and EUROTECH Sp. z o. o. for the facilities and hospitality.

5 Appendix: Numerical benchmarks

Let us define a set of benchmark solutions useful for testing different numerical solvers. Consider a class of positive functions $C_+(0,1)$ described in the following manner:

$$C_+(0,1) = \Big\{ h \in C^2(0,1) \cap C[0,1], \quad \lim_{x\to 1-} (1-x)^{-1/3} h(x) = 1, \quad h(x) > 0, \quad x \in [0,1) \Big\}.$$

By taking an arbitrary $h \in C_+(0,1)$, one can build a benchmark solution for the normalized formulation of the problem as:

$$u(x) = u_0 \psi_j(t) h(x). \tag{63}$$

where functions $\psi_j(t)$ and $h(x)$ are specified below. On substitution of (63) into (26) one finds:

$$q_l(t,x) = \gamma u_0 \left[\frac{1}{\beta} \left(x h^3(x) h'(x) + 3 \left(h^3(x) h'(x) \right)' \right) - h(x) \right] \psi_j^{\alpha}(t),$$
$$w_0(t) = u_0 \psi_j(t), \tag{64}$$

where two sets of the benchmark solutions can be considered. For the first one we choose $\psi_1(t) = e^{\gamma t}$ and $\beta = 2/3, \alpha = 1$, while for the second, $\psi_2(t) = (a+t)^\gamma$ and $\beta = 2\gamma/(3\gamma+1), \alpha = (3\gamma-1)/\gamma$. Corresponding crack lengths are defined in (25) and (30), respectively. Finally, the injection flux rate is computed from the boundary condition $(16)_1$:

$$q_0(t) = -\frac{u_0^4 \psi_j^4(t)}{L(t)} h^3(0) h'(0), \tag{65}$$

while the initial condition reads $W(x) = u_0 \psi_j(0) h(x)$.

Note that when taking the function $h(x)$ from the class $C_+(0, 1)$ in the following representation:

$$h(x) = (1-x)^{1/3}(1+s(x)),$$
$$s \in C^2[0,1], \quad s(x) > -1, \quad x \in [0,1)$$

q_l automatically satisfies the condition $q_l(t, x) = O((1-x)^{1/3}), x \to 1$.

The presented benchmarks allow one to test numerical schemes in various fracture propagation regimes (accelerating/decelerating cracks) by choosing proper values of the parameter γ [see Mishuris et al. (2012)]. Additionally, if one reduces the requirements for the smoothness of $s(x)$ near $x = 1$, assuming that $h \in C^2[0, 1) \cap H^\alpha(0, 1)$ the benchmark can serve to model singular leak-off regimes [compare with (61)].

Note that the zero leak-off case cannot be described by the aforementioned group of benchmarks. However, an analytical benchmark for this regime, represented in terms of a rapidly converging series, has been developed in Linkov (2011c).

References

Adachi J, Detournay E (2002) Self-similar solution of a plane–strain fracture driven by a power-law fluid. Int J Numer Anal Methods Geomech 26:579–604

Adachi J, Peirce A (2007) Asymptotic analysis of an elasticity equation for a finger-like hydraulic fracture. J Elast 90(1): 43–69

Adachi J, Siebrits E, Peirce A, Desroches J (2007) Computer simulation of hydraulic fractures. Int J Rock Mech Min Sci 44:739–757

Carter E (1957) Optimum fluid characteristics for fracture extension. In: Howard G, Fast C (eds) Drilling and production practices. American Petroleum Institute, pp 261–270

Crittendon BC (1959) The mechanics of design and interpertation of hydraulic fracture treatments. J Pet Tech 21:21–29

Detournay E (2004) Propagation regimes of fluid-driven fractures in impermeable rocks. Int J Geomech 4:1–11

Dobroskok AA, Linkov AM (2011) Modeling of fluid flow, stress state and seismicity induced in rock by instant pressure drop in a hydrofracture. J Min Sci 47(1):10–19

Economides M, Nolte K (eds) (2000) Reservoir stimulation, 3rd edn. Wiley, Chichester

Garagash DI, Detournay E (2000) The tip region of a fluid-driven fracture in an elastic medium. J Appl Mech 67:183–192

Geertsma J, de Klerk F (1969) A rapid method of predicting width and extent of hydraulically induced fractures. J Pet Tech 21:1571–1581 [SPE 2458]

Hubbert MK, Willis DG (1957) Mechanics of hydraulic fracturing. J Pet Tech 9(6):153–168

Kemp LF (1989) Study of Nordgren's equation of hydraulic fracturing. SPE Prod Eng 5:311–314

Khristianovic SA, Zheltov YP (1955) Formation of vertical fractures by means of highly viscous liquid. In: Proceedings of the fourth world petroleum congress, Rome, pp 579–586

Kovalyshen Y, Detournay E (2009) A reexamination of the classical PKN model of hydraulic fracture. Transp Porous Media 81:317–339

Kovalyshen Y (2010) Fluid-driven fracture in poroelastic medium. PhD thesis, The University of Minnesota

Kusmierczyk P, Mishuris G, Wrobel M (2012) Remarks on numerical simulation of the PKN model of hydrofracturing in proper variables. Various leak-off regimes. arXiv:1211.6474

Lecampion B, Detournay E (2007) An implicit algorithm for the propagation of a hydraulic fracture with a fluid lag. Comput Method Appl Mech 196(49–52):4863–4880

Linkov AM (2011) Speed equation and its application for solving ill-posed problems of hydraulic fracturing. ISSM 1028-3358. Doklady Phys 56(8): 436–438

Linkov AM (2011) Use of a speed equation for numerical simulation of hydraulic fractures. arXiv:1108.6146

Linkov AM (2011) On efficient simulation of hydraulic fracturing in terms of particle velocity. Int J Eng Sci 52:77–88

Mishuris G, Wrobel M, Linkov A (2012) On modeling hydraulic fracture in proper variables: stiffness, accuracy, sensitivity. Int J Eng Sci 61:10–23

Mitchell SL, Kuske R, Peirce A (2007) An asymptotic framework for finite hydraulic fractures including leakoff. SIAM J Appl Math 67(2):364–386

Mitchell SL, Kuske R, Peirce AP (2007) An asymptotic framework for the analysis of hydraulic fractures: the impermeable case. ASME J Appl Mech 74(2):365–372

Moos D (2012) The importance of stress and fractures in hydrofracturing and stimulation performance: a geomechanics overview. Search and Discovery Article 80255

Nordgren RP (1972) Propagation of a vertical hydraulic fracture. J Pet Tech 253:306–314

Olson JE (2008) Multi-fracture propagation modeling: applications to hydraulic fracturing in shales and tight gas sands. In: The 42nd US rock mechanics symposium (USRMS), June 29–July 2, 2008, San Francisco, CA, USA

Peirce A, Detournay E (2008) An implicit level set method for modeling hydraulically driven fractures. Comput Methods Appl Mech Eng 197:2858–2885

Peirce A, Detournay E (2009) An Eulerian moving front algorithm with weak-form tip asymptotics for modeling hydraulically driven fractures. Numer Methods Eng 25(2):185–200

Perkins TK, Kern LR (1961) Widths of hydraulic fractures. J Pet Tech 13(9):37–49 [SPE 89]

Rubin AM (1995) Propagation of magma filled cracks. Annu Rev Earth Planet Sci 23:287–336

Sneddon IN, Elliot HA (1946) The opening of a Griffith crack under internal pressure. Quart Appl Math 4:262–267

Tsai VC, Rice JR (2010) A model for turbulent hydraulic fracture and application to crack propagation at glacier beds. J Geophys Res 115:1–18

Zhang X, Jeffrey R, Llanos EM (2004) A study of shear hydraulic fracture propagation. In: Gulf rocks, the 6th North America rock mechanics symposium (NARMS), June 5–9, 2004, Houston, TX, USA

DOI 10.1007/978-3-319-04397-5_13
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9858-8

ORIGINAL PAPER

A solution to the parameter-identification conundrum: multi-scale interaction potentials

J. G. M. van Mier

Received: 30 October 2012 / Accepted: 19 March 2013 / Published online: 5 June 2013

Abstract Softening is a structural property, not a material property. Any material will show softening, but in this paper the focus is primarily on cement and concrete, which show this property very clearly owing to their coarse heterogeneity (relative to common laboratory-scale specimen sizes). A new model approach is presented, based on pair-potentials describing the interaction between two neighbouring particles at any desired size/scale level. Because of the resemblance with a particle model an equivalent lattice can be constructed. The pair-potential is then the behavioral law of a single lattice element. This relation between force and displacement depends on the size of the considered lattice element as well as on the rotational stiffness at the nodes, which not only depends on the flexibility of the global lattice to which the element is connected but also on the flexural stiffness of the considered element itself. The potential $F - r$ relation is a structural property that can be directly measured in physical experiments, thereby solving size effects and boundary effects.

Keywords Concrete · Fracture · Model parameters · Interaction potentials ($F - \delta$) · Size effect · Asymptotic behaviour

J. G. M. van Mier
Swiss Federal Institute of Technology(ETH Zurich),
8093 Zurich, Switzerland
e-mail:jvanmier@ethz.ch

1 Introduction

A reliable model for fracture of concrete is helpful for the design of strong and flexible structures that can withstand a variety of complex loadings. Two properties of concrete are of great importance when considering fracture. First of all, the material has a very low tensile strength, much lower than its compressive resistance. The imbalance between tensile and compressive strength becomes even larger when high (compressive) strength concrete is applied. Also in the case of confined compression the relative difference increases. The second, for fracture very important characteristic of concrete is its rough heterogeneity. The heterogeneity is a consequence of economics: reducing the price of concrete through the addition of relatively cheap sand and gravel to the more expensive binder (Portland) cement is common practice. Not only the costs of concrete decrease, the properties improve in comparison to the properties of pure hardened Portland cement, in particular the cement's brittleness is partly overcome. Models for concrete for structural applications are generally based on continuum mechanics. This 400-year old methodology is based on the assumption that material properties can be described using stress and strain as state variables, or stated differently, the mechanical behaviour of the material can be described by means of average properties. Central to developing a sound continuum-based theory is, not surprisingly, the so called Representative Volume Element (RVE), i.e. the smallest material volume needed to define the

average properties of the considered material. Conventional wisdom learns that the RVE should at least be a factor 3 to 5 larger than the largest heterogeneity found in the material. For concrete, but also other materials, the size of the grains is considered when defining the RVE. From a series of uniaxial tensile tests on specimens of varying size we concluded that the RVE should be larger than at least 8 times the maximum aggregate size, see Van Vliet and Van Mier (2000). This conclusion was based on the observation that beyond this threshold the scatter in the experiments decreases and becomes more-or-less constant. Small specimens, or rather, small structures show clearly the effect of heterogeneity. The increase of scatter below the aforementioned threshold increases as one single aggregate may be responsible for the structure's behaviour. Next to this, what we will call 'RVE-based-on-fixed-material-structure', clearly the size of nucleating and actively growing cracks must be considered. If a crack increases to a size of the same order of magnitude as the characteristic specimen size, boundary condition effects and geometry-related effects cannot be ignored any longer. In concrete cracks are not only caused by mechanical loading, but environmental conditions may have a profound influence as well. Differential temperature distributions during the hydration of cement (when concrete hardens) and/or differential moisture content in various part of a structure lead to eigen-stresses and with that, if the strength threshold is exceeded to crack nucleation and growth.

Owing to the coarse heterogeneity of concrete severe stress concentrations are present in the material when external load is applied. The stress-concentrations are the results from E-mismatch between the various material phases in the composites and the material's porosity. An important factor leading to crack growth in concrete at moderate external loading levels is the low tensile strength of the interfacial transition zone (ITZ) between matrix and aggregates. It is noted that when we would scale-down the specimen/structure size below the aggregate size, and zoom-in on a volume of cement matrix, the same observations can be made. Compared to the cement structure small sand grains will be relatively coarse and cause stress-concentrations. Going further down in scale will bring us to the scale of the cement binder. Here we will find un-hydrated cement, and the hydration products, which again can be seen as aggregates (now the un-hydrated cement kernels) embedded in a matrix with a clear smaller material structure (the structure of hydrated cement is found at nm-size/scale).

Thus, obviously, concrete and also cement are highly heterogeneous at various size/scale levels. It would be tempting to address the heterogeneity of concrete (and cement) via a fractal analysis, but it appears that 'jumps' are made along the dimensional scale, which would demand for a multi-fractal approach. We will not discuss these matters further, but rather suggest a different solution that will incorporate heterogeneity at any size/scale level implicitly. The proposed solution is based on a lattice model. Lattice models have been suggested as a tool for analyzing fracture of disordered materials in the last two decades of the past century, see for instance Roux and Guyon (1985), Termonia and Meakin (1986) and Herrmann et al. (1989), among many others. Since 1990 we have applied lattice-type models for simulating fracture of concrete, which has shown to be a valuable tool for obtaining a better understanding of fracture, be it that the approach is most fruitful when at the same time relevant experiments are carried out; see for instance Schlangen and Van Mier (1992) and Lilliu and Van Mier (2003) for 2D- and 3D-versions of the 'Delft' lattice model. Quite essential in our approach has been to incorporate the structure of concrete (or cement) directly into the model. Various methods are available to do so; the interested reader is referred to Van Mier (2012) for an overview. In a lattice model the material is modeled as a regular or irregular network of linear elements. For fracture it appears that the most realistic results are obtained if beam elements are used. By means of a finite element program forces and deformations in the network are calculated, and given a fracture criterion it can be decided which of the elements will break. Fracture is simulated by removing in each load-step just one element, re-calculating the stress-redistributions after removal, after which the next critical element is determined and removed. The disadvantage of this rather coarse way of simulating fracture is that the calculated load-displacement curves are generally too brittle. This can be repaired by including a softening stress-crack opening relation for the lattice elements as proposed by Ince et al. (2003). The obvious disadvantage is that not only the computation becomes more elaborate but also the softening stress-crack opening is in principle 'un-determined. The main advantage of our lattice approach based on element removal is that it is simple and transparent. No complicated iterative

procedures are needed, which is quite essential if the model is used in combination with experiments. The goal is getting a better understanding of fracture mechanisms in disordered materials like concrete; it is certainly not an attempt to develop a simulation model that can be fitted as closely as possible to experimental data. Up till now always the decision to fracture an element was based on a simple 'stress-criterion', i.e. when stresses in a lattice element would exceed a prescribed maximum stress (e.g. normal stress, flexural stress, etc.) the element would fail instantaneously (elastic-purely brittle). This implies that still continuum beam theory is used to decide whether an element will fail or not. In view of the statements regarding the RVE this is quite extraordinary (and probably not correct) since in the aforementioned lattice models the size of a lattice element is in the same order as the material's heterogeneity, for instance in concrete usually the lattice element length is selected 3 to 4 times smaller than the smallest aggregate particle incorporated in the material structure. Clearly there is reason to look at these matters in a different way, which is precisely what we will do in this paper.

The organization of the paper is as follows. In Sect. 2 we will argue that softening is a structural property, not a material property. This distinction is quite essential since it will be necessary to deviate from cohesive fracture models. In Sect. 3 we will show that the consequence of softening as a structural property, multi-scale approaches are deemed to fail. Rather an approach called 'up-scaling' from a pre-defined smallest size/scale level seems a more realistic option, which is then worked out in Sect. 4 for fracture of concrete. The softening relation needed in the 'structural lattice' debated in Sect. 4 is no longer expressed in terms of σ and δ, but rather in F and δ, and must be regarded as the structural property of a lattice element. A consequence is that the sought relationship between F and δ depends on the structural boundary conditions (fixed or pinned support; more general, the rotational stiffness k_r of the supports) and on the size of the lattice element, viz. a structural property. This is precisely what can be measured routinely in basic experiments in concrete technology, namely the uniaxial tensile properties and the (confined) compressive behaviour. A further advantage of using a 'structural' lattice model is that very likely compressive fracture can be captured in a lattice model as well. Buckling of lattice elements can be considered and can be included in the model in a relatively simple and straightforward manner. Again here we find the significant influence of boundary conditions and element slenderness (geometry), which is now incorporated directly into the model. Finally, in Sect. 5 methods are suggested that can be used to validate the new model approach. Without that the model would be incomplete as so many models are today.

2 What can be measured directly in fracture experiments?

In popular cohesive models for concrete the crack-tip closing stress is modeled according to the outcome of a uniaxial tension test between fixed (non-rotating) loading platen. This is what the fictitious crack model developers tell us; see Hillerborg et al. (1976). Since the early 1980s there have been numerous efforts to establish a standard tension or bending test that would help quantifying the properties needed in the fictitious crack model. What all experiments tell, however, is that none of the required properties for the fictitious crack model are true material properties. These parameters are the tensile strength f_t, the fracture energy G_f, the maximum crack opening δ_c and the shape of the softening curve $\sigma(\delta)$. In some way they all depend on specimen size, specimen geometry and boundary conditions. Drawing an equivalent to fracture under (confined) compression shows the same effects, albeit in this case specimen geometry and boundary conditions appear to be even more influential. In the case of tensile fracture researchers often revert to in-direct tests under the argument that uniaxial tension tests are too difficult. Instead seemingly simple experiments are carried out like the Brazilian splitting test or a (3-point of 4-point) bending test on prismatic beams. These latter tests require back-analysis of the results, and thus a number of assumptions are usually needed that are not always realistic. Here we focus on uniaxial tension and draw a parallel to uniaxial compression tests because both tests can deliver the data needed for the 'structural' lattice model, which is presented in Sect. 4. Observations from uniaxial tension and uniaxial compression tests lead to the following overview:

(1) When considering concrete the characteristic specimen size should be larger than 100–150 mm (i.e. 5 to 8 times the maximum aggregate size); for concretes with very large aggregates (dam concrete)

this would lead to even larger specimen sizes. Along the same lines of reasoning: for hardened cement paste the minimum required specimen size would be 500–1000 μm.

(2) In the fictitious crack model it has become common practice to model the pre-peak stress-strain curve as purely linear elastic. Experiments by Evans and Marathe (1968) indicate that already at a relatively low stress-level in the pre-peak regime microcracking starts. This is also the outcome from own experiments and simulations, see Van Mier (2009).

(3) The fictitious crack model tells us that microcracks propagate and widen in the softening regime. This is not true. Not only the experiments by Evans and Marathe (1968) lead to a different conclusion, a survey carried out by Mindess (1991) shows that the extent of a fracture process zone, commonly associated with the length of the crack-tip bridging zone, not only depends on specimen geometry but also on the accuracy of the crack detection methods used. It is also easy to show that in the softening regime a macroscopic crack dominates the behaviour of a test specimen, see Van Mier and Nooru-Mohamed (1990). These finding for uniaxial tension can easily be extended to (confined) compression; see Van Mier (2012).

(4) The specimen size has a significant influence on the tensile strength of a specimen. In general tensile strength will decrease with increasing specimen size (Fig. 1a); see for instance Van Vliet and Van Mier (2000) and Van Mier and Van Vliet (2003). Several (competing) theories describe this decrease of structural strength with specimen/structure size; see for instance Weibull (1939), Bažant (1984) and Carpinteri et al. (2003). The theory by Weibull is the only size effect approach based on sound physics. The main assumption here is that large structures have a larger probability of containing defects and hence are weaker. The other two models, sometimes dubbed 'laws' are exercises in curve-fitting. Quite important in both latter approaches is the notion of the asymptotic behaviour for small and large sizes (see Fig. 1a). Trying to validate the asymptotic behaviour in the small size range is impossible: as soon as the specimen/structure size becomes smaller than the RVE experimental scatter will become larger and a solid conclusion cannot be drawn. The RVE is estimated at $8d_a$ and indicated in Fig. 1a. The large-size asymptotic behaviour cannot be established either because the laboratory facilities will be decisive for the maximum size of a specimen/structure that can be tested. In most labs the characteristic specimen size will not exceed 2 m. A workable option would be to test larger specimens floating on water. The large-size asymptote is considered more important than the lower-size asymptote for the sole reason that such results can be used directly in structural engineering. Note that at the lower-size asymptote, below the RVE, we start testing different materials, namely the individual constituents of concrete: hardened cement paste and aggregate. In Fig. 1a conclusions about models and experiments can only be drawn in the area enclosed by the box of the diagram. Beyond these boundaries only fruitless speculation is possible; many hours have been lost at conferences in the past decade debating the best 'size effect law' on the basis of the behaviour at the extremes. It is obvious that it is impossible to discriminate between the proposed models (Bažant 1984 and Carpinteri et al. 2003) simply because it will never be possible to validate the speculations by means of physical experiment.

(5) With increasing specimen/structure size the fracture energy (i.e. the area under the post-peak softening curve) increases as shown in Fig. 1b. For large sizes (> 1.6 m for 8-mm concrete) the curve seems to level of towards a horizontal asymptote. Therefore it is believed that fracture energy may be a true material property, but very large specimens are needed to make the actual measurements for a given concrete. Models based on local fracture energy g_f (for instance Duan and Hu 2004) are difficult to feed with experimental data simply because at the lower end of the size-scale measurements of fracture energy will not yield realistic values. The complete stress-deformation diagram is affected when specimen/structure size increases. Larger specimens show a more brittle behaviour; just beyond peak even snap-back behaviour may occur depending on the choice of control parameter in the closed-loop testing system (see also references mentioned in the next paragraph).

(6) It is by no means clear for which characteristic specimen size the softening diagram needed in the fictitious crack model must be determined. Obtaining a stable softening diagram is not easy if

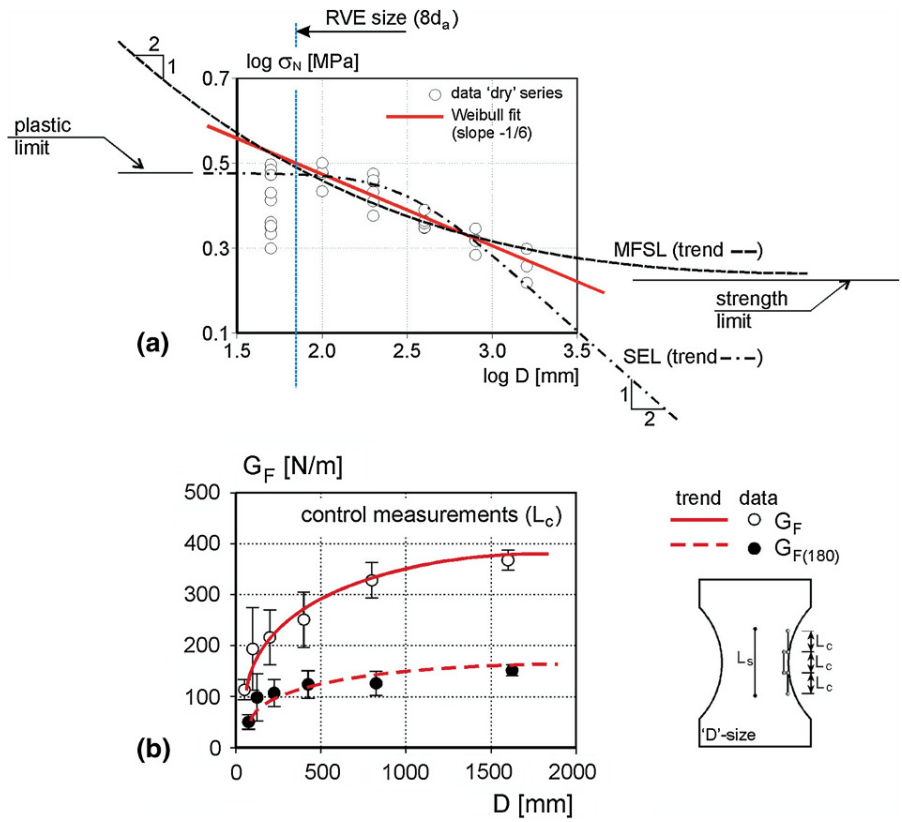

Fig. 1 Size effect on structural strength (**a**) and fracture energy (**b**) from uniaxial tensile tests on dog-bone shaped specimens. Testing below the RVE-size (taken here as $8\,d_a$, and indicated in (**a**)) is not possible due to increasing scatter. Therefore assumed asymptotic behaviour in the small size/scale regime can never be validated experimentally. The fracture energy has been calculated up till 180 μ m crack opening, denoted as $G_{F,180}$ and till full separation G_F. In the latter case measured curves were linearly extrapolated to the point where they intersected with the x-axis. Data are from Van Mier and Van Vliet (2003). The Weibull theory has been fitted to the experimental data; for SEL and MFSL only trends are shown indicating the asymptotic behaviour for small and large sizes

the right equipment is lacking, see Van Mier and Shi (2002). Considering that the fracture energy grows towards an asymptotic value for larger specimens, this might indicate that even larger specimens (2-m range) would be required than based on considering the RVE. At those larger sizes maintaining stability of crack growth in the softening regime is most difficult because often snap-back behaviour may occur. The difficulties can be overcome, however, as shown in Van Vliet and Van Mier (2000) and Van Mier and Shi (2002), but require next to the servo-hydraulic control system some additional electronics.

(7) Boundary conditions have a pronounced influence on the softening behaviour, both in tension and under (confined) compression. More specifically, in tension the rotational freedom at the nodes affects the tensile strength, the pre-peak non-linearity, the shape of the softening curve and the fracture energy. In (confined) compression, in addition to the rotational freedom of the supports the frictional restraint at the specimen-loading platen interface must be considered. In Fig. 2 three different cases are shown: fixed boundaries ($k_r = \infty$) using a slender specimen ($h/d > 2$) in Fig. 2a, pinned boundaries ($k_r = 0$) using a slender specimen in Fig. 2b and in Fig. 2c a stubby specimen ($h/d < 1$) loaded between fixed boundaries. With fixed boundaries two cracks will develop in the softening regime; restraining the rotations at the specimens ends will cause the bump in the diagram as the two cracks develop in sequence from two opposite

sides of the specimen, see Van Mier (1986). In contrast, when a slender specimen is loaded between pinned boundaries the first crack to develop is also the crack leading to complete failure of the specimen, i.e. no secondary cracking can occur. The fracture energy is markedly smaller in the second case as a direct consequence of the reduced crack area; the tensile strength is smaller under pinned boundaries in comparison to fixed boundaries; see Van Mier et al. (1995). When a stubby specimen is used instead of a slender specimen the stress-redistributions occur earlier, around peak, and results by Hordijk (1991) show that the pre-peak part of the diagram becomes more curved, the deformation at peak-load increases and the 'bump' has disappeared. The result is shown schematically in Fig. 2c. More recently, Akita et al. (2007) showed that the specimen shape has a significant effect on tensile strength as well.

(8) As a consequence of the size effect on strength and deformation, and the influence of boundary rotations it is impossible to choose the 'best' or 'most appropriate' type of experiment for determining the softening diagram of concrete in tension and with that the closing stress-profile in a cohesive fracture model. The point of view that comes closest to all results is that softening is a 'structural property', rather than a 'material property'. It is impossible to separate boundary effects from material effects in all these experiments. The main reason is that a crack with a size comparable to the specimen/structure dimensions is developing, and like in classical fracture mechanics a correction for these effects must be incorporated in any model trying to deal with the aforementioned phenomena. Thus, the behaviour measured in a uniaxial tensile test is valid only for the chosen specimen size and the applied boundary conditions. The resulting $F-\delta$ relation should be used directly in a model, as this is the only un-biased result that can be derived from an experiment. We will return to these matters in Sect. 4. It should be noted that the last remaining 'continuum state variable' in the fictitious crack model has been dropped; rather than giving results in terms of average stress over the specimens cross-section it will be an improvement to present matters directly in force and displacement, thereby also incorporating the pre-peak behaviour in the formulation. Note that this is a significant deviation from the fictitious crack model.

For compressive fracture the same situation emerges, see Van Mier (2009, 2012). Next to the chosen boundary rotations also frictional restraint between loading platen and specimen ends will affect the measurements. Higher boundary restraint results in a higher compressive strength, larger deformations at peak stress and a shallower softening branch. The interested reader is referred to my recent book for a complete overview of all factors affecting compressive fracture; see Van Mier (2012). The aforementioned effects will also be found under confined compression, provided the confinement stays below the brittle-to-ductile transition.

In conclusion to this section it can be stated that it does not make sense to continue with cohesive fracture models for concrete. The essential parameter, the $\sigma-w$ relation is not a 'material property' but must be seen as the response of the complete specimen-machine system. Softening is a 'structural property'. In Sect. 4 we

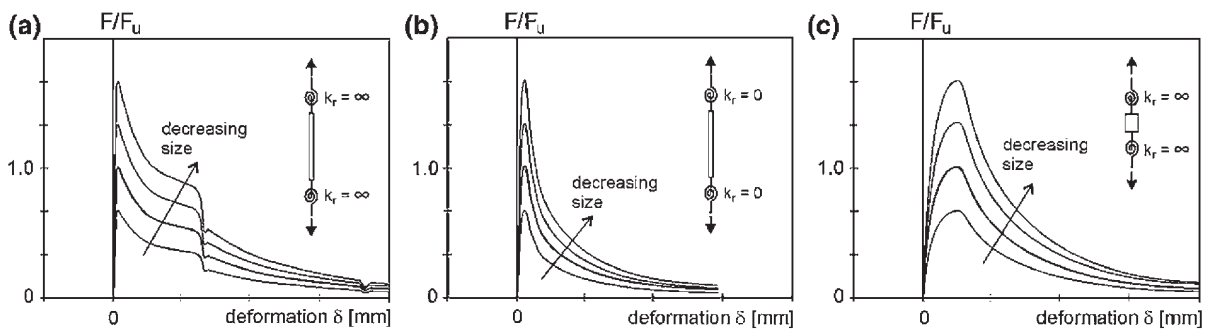

Fig. 2 Effect of boundary rotations on tensile stress-deformation diagram of concrete. In **a** and **b** a slender specimen is loaded between fixed (non-rotating) loading platen and pinned (freely rotating) loading platen, respectively, whereas in **c** fixed specimen ends are used in a test on a stubby specimen

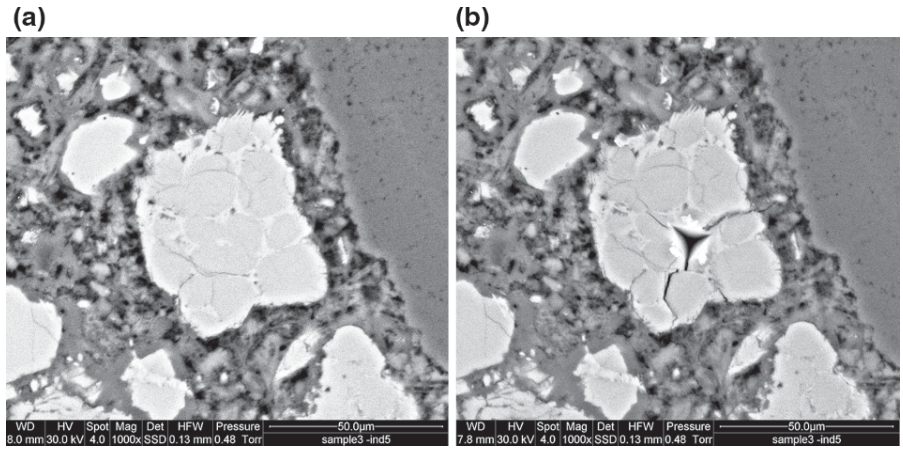

Fig. 3 Two images of a polished surface of hardened cement paste, before (**a**) and after indentation with a Berkovich diamond tip. The almost white particles are the remaining un-hydrated cores of partially hydrated cement grains. *Light gray* and *dark gray* are the low and high-density calcium silicate hydrates (CSH), and *black* is porosity. After Van Mier (2007)

will see how we can still work with such a relationship.

3 Multi-scale modeling?

Recently there has been quite some interest in multi-scale modeling. At the smallest considered size/scale-level the behaviour of the constituents of a composite are determined, for example, and used in sequential analyses at higher size/scale levels, all the way up to the macroscopic (or engineering) level. Does this make sense? It will only work when at the smallest size/scale true material properties are used. In scaling-up to larger size/scale levels boundary effects and size effects can be incorporated, which would lead to correct results at the macroscopic level. The question is thus: are we capable of determining true material properties at the smallest size/scale? In relation to this question immediately a second one can be posed, namely: what is the most appropriate 'smallest' size/scale to start from? Let us assume that we are dealing with concrete and the appropriate lower size/scale-level is the [μm]-level where the structure of hydrated cement can be seen in great detail. In Fig. 3 the structure of cement at the [μm]-size/scale is shown, before and after an indent with a Berkovich diamond tip. The smooth gray area at the right corner of Fig. 3a is part of a sand grain, the more-or-less white particle in the center is an un-hydrated cement grain (approximately 50 μm across), the smaller darker gray patches forming the matrix between the sand and un-hydrated cement particles is hydrated cement, interspersed with porosity, which appears as black specs. The hydrated cement usually comes in two forms: low-density Calcium Silicate Hydrates (in short: CSH) away from the un-hydrated kernel, and high-density CSH directly in contact with the un-hydrated cement grain. In the smallest-scale part of the multi-scale model we need to incorporate all these material phases and, in addition, the interfaces between the various components. A minimum model would require knowledge about the mechanical properties of un-hydrated cement (which is a composite by itself, as can be seen in Fig. 3b after the indentation has been made), low- and high density CSH and at least 3 types of interfaces. The indentation shown in Fig. 3b is one of the few (in-direct) tests available for determining the mechanical properties of these material phases. For instance, Constantinides and Ulm (2004) have attempted to determine the Young's modulus of low- and high-density CSH by means of indentation tests, and reported a higher modulus for high-density CSH in comparison to low-density CSH. Problems in indentation testing are numerous, and just the simple fact that the tests are in-direct makes them suspicious. As an alternative one can try to carry out uniaxial tension tests (see for instance Trtik et al. 2007). Machining tiny specimens of hardened cement paste, or isolating small probes made of the individual cement hydrates (see below) is tedious, and often leads to using larger specimens that contain all the aforementioned material phases. Using a micro-mechanical model one would then have to perform back-calculations and try to estimate the properties of the various material phases. For certain, not a simple task and not a small task either.

An alternative route is to scale-down from the micro-mechanical tensile test and try to obtain specimens consisting of pure cement phases (low-density CSH, high-density CSH, calcium hydroxide and un-hydrated cement), for instance by using a focused ion beam (FIB). This is certainly not an easy task either, but can be done. If we have succeeded in producing the specimens and testing them as well, the same problems that we discussed in the previous Section will return: what to do with size effects and boundary conditions? The answer will not change: again we are measuring structural properties rather than material properties, except perhaps for the Young's modulus and the fracture energy, but these are certainly not sufficient for constructing a fracture model. So, the suggestion to use a lattice model and to feed into the model directly the structural properties of a lattice element might be a workable approach. In the next Section we will explore the advantages and disadvantages of such a model.

4 Structural lattice based on multi-scale interaction potentials (*F-r*)

There is a resemblance between the shape of the attractive part of an atomic potential and the tensile force-deformation diagram for concrete, see Van Mier (2007). A well-known form for the atomic potential for noble gases is the Lennard-Jones (LJ) potential, which may be written as:

$$\frac{V_{LJ}(r)}{\varepsilon} = -4\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{1}$$

where σ and ε are units of length and energy, respectively. The potential describes the balance between attractive and repulsive forces at the level of atoms. Is the distance between two atoms below the equilibrium separation r_0, the repulsion must be overcome and the atoms must be forced to remain at the prescribed distance. With increasing separation distance between the atoms the energy needed decreases. Equation (1) can be rewritten as:

$$\frac{F}{F_u} = \alpha\left[\left(\frac{\sigma}{r}\right)^{m} - \left(\frac{\sigma}{r}\right)^{n}\right], \tag{2}$$

A relation between the force F to keep the atoms at prescribed separation distance $1/r$. The powers n and m can be varied to obtain the required shape of the potential. In Fig. 4 we show the shape for the parameter

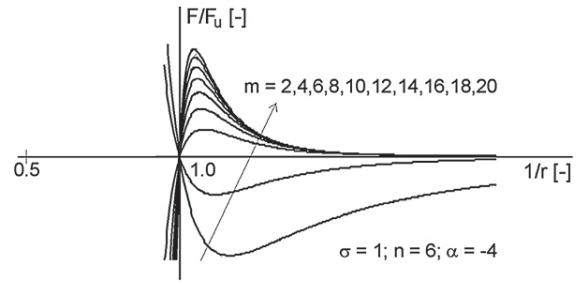

Fig. 4 Variation of the power m in Eq. (2) leads to the family of potentials shown in this Figure. The shape corresponds to the force-deformation diagrams for concrete under uniaxial tension. After Van Mier (2007)

setting $\sigma = 1, n = 6, \alpha = -4$, and m varies between 2 and 20. The curves flip over when $m > 6$. The family of curves above this threshold resembles the shape of a family of force-deformation curves from uniaxial tension tests on concrete, as shown for instance in Fig. 2. It is therefore tempting to investigate whether the simple formulation can be used at higher size/scale-levels as well. Note that the family of curves has been shown relative to the response of a specimen of reference size D_0, having a maximum force F_u.

Beranek and Hobbelman (1992, 1994) showed that a similarity exists between a particle model and a beam lattice model. Starting point was the analysis of a stack of equal-sized spherical particles in contact. The deformations of this model were compared with those of a beam lattice. Lattice beams were assumed to connect the centers of two neighbouring particles. For a certain size of the lattice beams the similarity was perfect, indicating that both models would lead to the same result. The idea is now as follows. A tensile test on a prismatic specimen is interpreted as the potential for a pair of particles of the same size. The so-called pair-potential is thus assumed to apply at larger size/scale levels than the atomic level, even all the way up to the macroscopic size/scale level. Testing a specimen at the required size/scale will yield immediately the required potential. As we show in Fig. 5, we can thus establish the potential at various size/scale-levels such as the nano-, micro-, meso- and macro-levels. Nano would probably be far-fetched when dealing with cement and concrete but the behaviour of the material at the other scales is quite relevant.

Let us now consider the two different interpretations of the meso-structure of concrete as shown in Fig. 6a, b. Following the interpretation of Fig. 6a concrete is seen as a three-phase composite, consisting

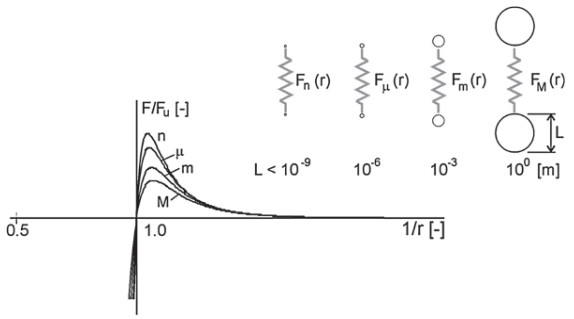

Fig. 5 Pair-potentials for application in a beam-lattice model at various size/scale levels (nano-, micro-, meso- and macro-level). After Van Mier (2012). Note that the potential is active between two particles of identical size; the distance between the particles is here shown exaggerated to indicate that only normal forces are considered between the interacting particles

of cement-matrix in which the various aggregates are embedded. Between the two phases is an interfacial transition zone, which has relatively low strength, and is in fact the weakest part of the concrete structure. Since the matrix is shown here as a continuous phase, we could interpret this visualization as the situation after hydration. The matrix is built up from the smallest sand grains that have not been explicitly included in the model, the Portland cement, and, if present, fly-ash and/or condensed silica. In the lattice model that we built in Delft and Zurich, as a series of consecutive PhD-projects, the visualization of Fig. 6a was taken as a starting point. The regular or random lattice was simply projected on top of the 3-phase material structure and properties were assigned depending on where a certain lattice element would be located, see Fig. 6c. In Fig. 6b concrete is depicted as a stack of spherical particles of varying size. In a way this is the situation before the cement hydrates; only the water needed for hydration is not shown in this Figure (note that the mixing water in concrete is initially absorbed at the particle's surfaces and possible excess water will gather in voids between the particles). All particles sizes are present: from the largest [mm]-size aggregates to the smallest (sub-[μm] size) fly-ash and condensed silica particles with the cement grains of a size falling between these extremes. With such a hierarchical system the densest possible material structure can be obtained, which will have the highest possible strength. The material interpretation of Fig. 6b can also be turned into a lattice. By simply connecting the centers of neighbouring particles a lattice is constructed. Each lattice element represents the interaction between the two neighbouring particles. Interpreting the system as just a bunch of pair-interactions is probably too simple, and higher order interactions may be included, at the cost of a loss in transparency of the model. In Fig. 6d the connectivity between the particle centers is shown, in Fig. 6e the remaining lattice.

The potential law, Eq. (2) describes the behaviour of a lattice element, but an essential adjustment must be made. The potential used in our lattice model depends on the actual size of the individual lattice elements, and on the rotational support stiffness at both nodes, i.e. the connectivity to the other lattice elements. The rotational stiffness at the supports depends not only on the connectivity to the rest of the lattice but also on the flexural stiffness of the lattice element itself. So, rather than descending to a so-called material level, and describing the properties of the lattice elements via constitutive equations, we remain at the 'structural level' and describe the properties of each lattice element directly as a function of size and support conditions. The complication that arises in conventional cohesive fracture models is solved, namely the dependence of cohesive fracture properties on element size and boundary conditions is now implicitly included in the model. Figure 7 shows examples of (size-dependent) potentials for three different boundary conditions and varying specimen slenderness. Specific characteristics of a lattice element are included in the $F - r$ potential. For instance the 'bump' in the softening curve when a slender lattice element is tested between fixed boundaries is included in the potential function. It is not seen as an 'inconvenience' that at all costs must be removed from the model. No, it is just part of the behaviour of that particular lattice element when the element's ends are fixed against rotations. Likewise we will have to use the smooth curve for a lattice element between pinned supports, which is actually a condition that will not be found in a beam lattice model, and the increased pre-peak deformations for a stubby lattice beam between fixed supports. The latter case may appear frequently since in the beam lattices explored to date always relatively stubby elements have been used, which came from the demand to justify the elastic lattice properties to those of a real concrete specimen or structure; see for instance in Schlangen and Mier (1994).

Returning to the model of Fig. 6d, e, which was derived from the concrete material structure of Fig. 6b,

we can easily see the implications of this model. For compression the effects from bc on lattice element response are even more pronounced in comparison to the effects shown here for tension; see the overview in Chapter 8 in Van Mier (2012). The same approach as sketched above can be applied, however. As an example consider the particle stack of equal-sized disks (2D is considered for explaining the matter here) subjected to external compression in Fig. 8a. The contact forces between the particles are either compressive or tensile (splitting forces will occur between horizontal oriented particle pairs), as shown in the equivalent lattice model. If the disks are of varying size the lattice element sizes vary correspondingly and the potential functions describing the relation between lattice element force F and deformation r will vary as well.

5 Which experiments are essential?

In the model approach suggested in the previous section the behaviour of a lattice element is described directly in terms of force and deformation. The consequence is that for each lattice element size and bc the behaviour must be estimated from experiments. How can this be done? In the first place the task will be simplified by considering a limited set of bc's, which could be identified from elastic analyses of a corresponding beam lattice. The size of the lattice elements can be set using the approach in Schlangen and Mier (1994). After the rotational stiffnesses of the two nodes of an element have been established, we know how large the specimen should be as well as the rotational support stiffnesses. After that it is rather simple and straightforward to conduct the experiment. The main advantage of the

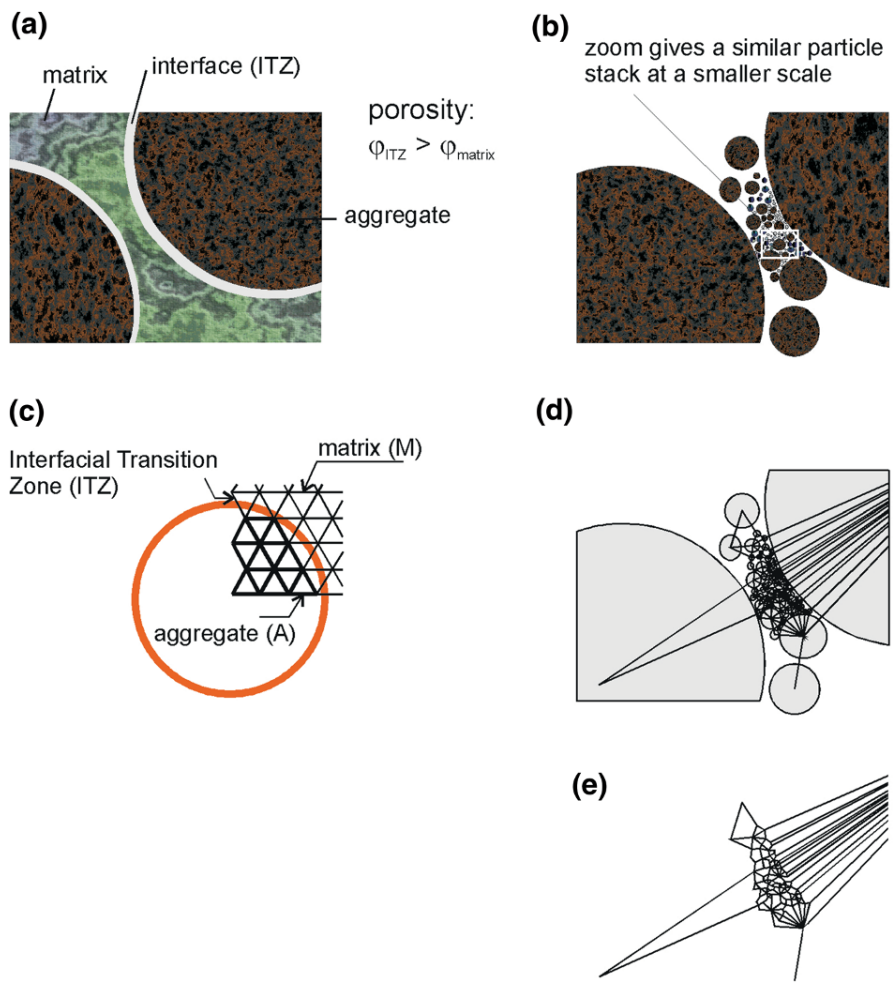

Fig. 6 Two different interpretations of the meso-structure of concrete: **a** concrete as a three-phase composite of matrix, aggregate and interfacial transition zone (ITZ), and **b** as a stack of particles of different size. The model of (**a**) has been used frequently in the past: a regular or random lattice was projected on top of the material structure and the lattice elements would be assigned properties according their location on the material structure as shown in (**c**). In the model of (**b**) particles can be anything from gravel, sand, condensed silica, Portland cement, fly-ash, and so on. The particle stack of (**b**) can form the basis for the construction of a lattice as shown in (**d**) and (**e**), which would form the basis of the 'structural lattice' discussed in this paper

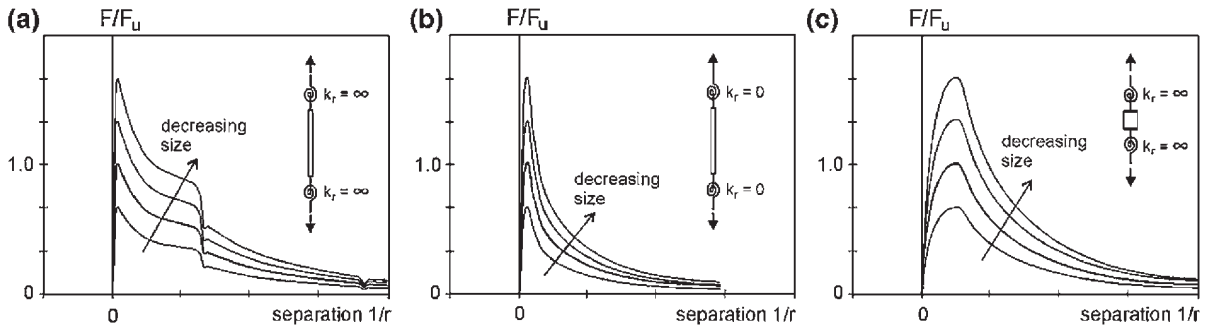

Fig. 7 Examples of pair-potentials $F/F_u - r$ for three different structural conditions of the lattice element: **a** slender lattice element loaded between fixed (no-rotating) ends, **b** slender lattice element loaded between pinned ends, and **c** stubby lattice element between fixed ends

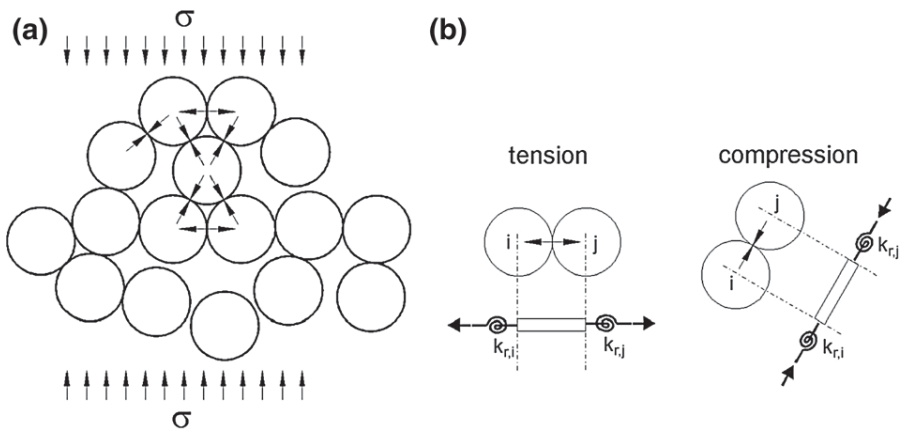

Fig. 8 Stack of equal-sized disks subjected to external compression (**a**). Tensile splitting forces develop between horizontal pairs and compressive stresses are transmitted between vertically arranged particle pairs or inclined pairs. In (**b**) the equivalent lattice element and its boundary conditions are shown. The input potential is simply what is measured for such a lattice element in a physical experiment

procedure sketched here is that the basis of the model approach is defined from the constraints set by laboratory experiments. The model is defined based on what we can actual derive from physical experiments and no ad-hoc assumptions are necessary. As we have discussed in Sect. 2 properties for cohesive crack models are dependent on specimen size and bc, and in that respect cohesive crack models are rather useless. The situation is more severe for compression than for tension. With the new approach one could argue that many different specimen sizes and bc's must be tested. Indeed this is the case, but probably the number of cases can be reduced significantly by simplifying the material structure to some extent, for instance leaving out all particles below a certain size-threshold. Such simplifications were also made in the lattice model mentioned before (Schlangen and Van Mier 1992 and Lilliu and Van Mier 2003), but the model outcomes will still be useful for obtaining a profoundly improved insight in the fracture behaviour of cement and concrete. Unfortunately the simplicity and transparency of the original lattice model are lost to some extent. Yet, the fit of the load-displacement curve should improve substantially. Since any type of structure is calculated as such (i.e. reconstructing the exact boundary conditions), there is no need to assume that some properties are 'material properties'. We simply calculate the effects from structure size and boundary conditions and the effort will be rewarded because the model starts from the behaviour of single lattice elements that can actually be measured in laboratory experiments. In principle the fictitious crack model also predicts size effects; yet there

an un-resolved problem remains, i.e. softening is not a material property.

Next to the tests aimed at establishing the potential-law for the individual lattice elements, large scale tests on larger material volumes are quite essential, in particular the determination of the evolution of the crack populations under a variety of loading paths. These large-scale tests are needed to establish fracture mechanisms. A useful approach is the use of X-ray tomography, which can provide the evolution of the full three-dimensional crack population in time; see for instance Trtik et al. (2007) and Meyer et al. (2009).

6 Conclusion

In this paper we discuss an alternative to cohesive crack models, namely a lattice model based on $F - r$ potentials. The potential function to be used depends on the size of the lattice element (slenderness h/d) and the rotational stiffness at the nodes. The rotational stiffness at the nodes of each lattice element depends on the connectivity of an element to neighbouring lattice elements as well as on the flexural stiffness of the lattice elements themselves. The implication is that for a variety of boundary conditions and a variety of lattice element sizes the potential function must be measured, but the enormous advantage is that what is needed as input in the model can actually be measured in physical experiments. The potential describes the structural behaviour of a lattice element. Material properties do not exist, which is the main deviation here from assumptions in cohesive models where the softening curve is considered as a material property. Basically we start from what can be measured in a fracture experiment, in stead of trying to measure haphazardly proposed parameters. The model approach explained in this short paper is an elegant manner to overcome the parameter-identification conundrum which seems to affect most fracture models today.

References

Akita H, Koide H, Mihashi H (2007) Specimen geometry in uniaxial tension test of concrete. In: Carpinteri A et al (eds) Proceedings FraMCoS-6. Taylor & Francis, London, pp 243–248

Bažant ZP (1984) Size effect in blunt fracture: concrete, rock, metal. J Eng Mech 110:518–535

Beranek WJ, Hobbelman GJ (1992) Handboek voor het mechanisch gedrag van metselwerk (Glossary of the mechanical behaviour of masonry). Report C-77, CUR Foundation, Gouda (in Dutch)

Beranek WJ, Hobbelman GJ (1994) Constitutive modelling of structural concrete as an assemblage of spheres. In: Mang H, Bicanic N, de Borst R (eds) Computational models of concrete structures, proceedings EURO-C. Pineridge Press, Swansea, pp 535–542

Carpinteri A, Chiaia B, Cornetti P (2003) On the mechanics of quasi-brittle materials with a fractal microstructure. Eng Fract Mech 70:2321–2349

Constantinides G, Ulm FJ (2004) The effect of two types of C-S-H on the elasticity of cement-based materials: results from nano-indentation and micro-mechanical modeling. Cem Conc Res 34(1):67–70

Duan K, Hu X (2004) Specimen boundary induced size effect on quasi-brittle fracture. Strength Fract Complex 2:47–68

Evans RH, Marathe MS (1968) Microcracking and stress-strain curves for concrete in tension. Mater Struct (RILEM) 1(1):61–64

Herrmann HJ, Hansen A, Roux S (1989) Fracture of disordered elastic lattices in two dimensions. Phys Rev B 39(1): 637–648

Hillerborg A, Modeér M, Peterson P-E (1976) Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements. Cem Conc Res 6:773–782

Hordijk DA (1991) Local approach to fatigue of concrete. PhD thesis, Delft University of Technology

Ince R, Arslan A, Karihaloo BL (2003) Lattice modeling of size effect in concrete strength. Eng Fract Mech 70:2307–2320

Lilliu G, Van Mier JGM (2003) 3D lattice type fracture model for concrete. Eng Fract Mech 70(7/8):927–942

Meyer D, Man H-K, Van Mier JGM (2009) Fracture of foamed cementitious materials: a combined experimental and numerical study. In: Zhao H, Fleck NA (eds), Mechanical properties of cellular materials. New York: IUTAM Bookseries 12, Springer Science+ Business, Media, 115–123

Mindess S (1991) Fracture process zone detection. In: Shah SP, Carpinteri A (eds) Fracture mechanics test methods for concrete. Chapman & Hall, London, pp 231–261

Roux S, Guyon E (1985) Mechanical percolation: a small beam lattice study. J Phys Lett 46:L999–L1004

Schlangen E, Van Mier JGM (1992) Experimental and numerical analysis of the micro-mechanisms of fracture of cement-based composites. Cem Conc Comp 14(2):105–118

Schlangen E, Van Mier JGM (1994) Fracture simulations in concrete and rock using a random lattice. In: Siriwardane H, Zaman MM (eds) Computer methods and advances in geomechanics. Balkema, Rotterdam, pp 1641–1646

Termonia Y, Meakin P (1986) Formation of fractal cracks in a kinetic fracture model. Nature 320:429–431

Trtik P, Stähli P, Landis EN, Stampanoni M, Van Mier JGM (2007) Microtensile testing and 3D imaging of hydrated portland cement. In: Carpinteri A, Gambarova P, Ferro G, Plizzari G (eds), Proceedings 6th international conference on fracture mechanics of concrete and concrete structures (FraMCoS-VI). London: Taylor & Francis, 1277–1282

Van Mier JGM (1986) Fracture of concrete under complex stress. HERON 31(3):1–90

Van Mier JGM (2007) Multi-scale interaction potentials ($F - r$) for describing fracture of brittle disordered materials like cement and concrete. Int J Fract 143(1):41–78

Van Mier JGM (2009) Mode II fracture localization in concrete loaded in compression. J Eng Mech (ASCE) 135(1):1–8
Van Mier JGM (2012) Concrete fracture—multiscale approach. Taylor & Francis/CRC Press, Boca Raton
Van Mier JGM, Nooru-Mohamed MB (1990) Geometrical and structural aspects of concrete fracture. Eng Fract Mech 35(4/5):617–628
Van Mier JGM, Schlangen E, Vervuurt A (1995) Lattice type fracture models for concrete. In: Mühlhaus H-B (ed) Continuum models for materials with micro-structure. Wiley, Chichester, pp 341–377
Van Mier JGM, Shi C (2002) Stability issues in uniaxial tensile tests on brittle disordered materials. Int J Solids Struct 39:3359–3372
Van Mier JGM, Van Vliet MRA (2003) Influence of microstructure of concrete on size/scale effects in tensile fracture. Eng Fract Mech 70(16):2281–2306
Van Vliet MRA, Van Mier JGM (2000) Experimental investigation of size effect in concrete and sandstone under uniaxial tension. Eng Fract Mech 65(2/3):165–188
Weibull W (1939) A statistical theory of strength of materials. Roy Swedish Inst Eng Res 151:1–45

DOI 10.1007/978-3-319-04397-5_14
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9867-7

ORIGINAL PAPER

Remarks on application of different variables for the PKN model of hydrofracturing: various fluid-flow regimes

P. Kusmierczyk · G. Mishuris · M. Wrobel

Received: 24 November 2012 / Accepted: 25 June 2013 / Published online: 11 July 2013

Abstract The problem of hydraulic fracture for the PKN model is considered within the framework presented recently by Linkov (Doklady Phys 56(8):436–438, 2011). The modified formulation is further enhanced by employing an improved regularized boundary condition near the crack tip. This increases solution accuracy especially for singular leak-off regimes. A new dependent variable having clear physical sense is introduced. A comprehensive analysis of numerical algorithms based on various dependent variables is provided. Comparison with know numerical results has been given.

Keywords Hydraulic fracture · Numerical simulations · Various leak-off regimes · Carter model

This work has been done in the framework of the EU FP7 PEOPLE Project under Contract Number PIAP-GA-2009-251475-HYDROFRAC.

P. Kusmierczyk · G. Mishuris (✉) · M. Wrobel
Institute of Mathematical and Physical Sciences, Aberystwyth University, Ceredigion, Wales SY23 3BZ, UK
e-mail:ggm@aber.ac.uk

P. Kusmierczyk
e-mail:pkk8@aber.ac.uk

P. Kusmierczyk · G. Mishuris · M. Wrobel
EUROTECH Sp. z o.o., ul. Wojska Polskiego 3, 39-300 Mielec, Poland
e-mail:miw15@aber.ac.uk

Mathematics Subject Classification 74F10 · 74H10 · 74H15

1 Introduction

In its broadest definition, hydraulic fracturing refers to a problem of a fluid driven fracture propagating in a brittle medium. The process has been known for at least 50 years (Crittendon 1959; Desroches and Thiercelin 1993; Desroches et al. 1994; Detournay 2004; Geertsma and de Klerk 1969; Harrison et al. 1954; Hubbert and Willis 1957; Khristianovic and Zheltov 1955). The respective technology has been utilized in the petroleum industry to intensify the extraction of hydrocarbons for decades.

Recently, due to economical reasons, it has been revived for exploiting non-conventional hydrocarbon deposits. The process has many other technological applications (e.g. disposal of waste drill cuttings underground (Moschovidis and Steiger 2000), geothermal reservoirs exploitation (Pine and Cundall 1985) or in situ stress measurements (Desroches and Thiercelin 1993). Hydrofracturing also appears in nature (e.g. geological processes, like magma-driven dykes (Rubin 1995; Lister 1990) or a subglacial drainage of water Tsai and Rice 2010). Due to the complexity of this multiphysical phenomenon, mathematical and numerical simulation of the process still represents a challenging task, in spite of the fact that immense progress has been made since the first algorithms were developed.

The mathematical model of the problem should account for coupled mechanisms driving the process, which are: (i) solid mechanics equations, describing the deformation of the rock induced by the fluid pressure; (ii) equations for the fluid flow within the fracture and the leak-off to the rock formation; (iii) fracture mechanics criteria defining the conditions for fracture propagation. Further development of the model involves incorporation of mass transport for the proppant movement, fluid diffusion to account for the rock saturation by the leak-off flux, thermal effects affecting rheological properties of the fluid and others.

The computational challenges of the hydrofracturing models result from several factors: (i) strong nonlinearity introduced by the Poiseuille equation describing the fluid flow; (ii) in the general case, a non-local relationship between the fracture opening and the net fluid pressure; (iii) moving boundaries of the fluid front and the fracture contour; (iv) degeneration of the governing PDE at the fracture front; (v) possible lag between the crack tip and the fluid front.

The first simplified mathematical descriptions of hydraulic fracture were summarized in the following three main classical models. The so-called PKN model was considered in Perkins and Kern (1961), where the authors adopted Sneddon's solution (Sneddon and Elliot 1946) which was further enhanced by Nordgren (1972) to account for the fluid loss effect and fracture volume change. As a result, the crack length was determined as a part of the solution. The so-called KGD (plain strain) model was developed independently by Khristianovic and Zheltov (1955), and Geertsma and de Klerk (1969). Finally, the radial or penny-shaped model was introduced by Sneddon (1946) with constant fluid pressure and was extended for the general case by Spence and Sharp (1985).

Different variations of the aforementioned models were used for treatment designs for decades, despite the fact that each of them is valid only under very specific assumptions (like elliptic cross-section of the fracture, and fracture half-length much greater than its constant height for the PKN model). Recently, the classical models have been largely replaced by the pseudo 3D models (Mack and Warpinski 2000). A comprehensive review of the history and techniques of hydrofracturing simulation can be found in Adachi et al. (2007).

Although the classical models have been superseded in most of the practical applications, they still play a crucial role when developing and analyzing new computational algorithms. The models enable one to understand the nature of, and reasons for, computational difficulties, find the remedies for them and to extend these ideas to the general case.

Thus, in the pioneering work by Nordgren (1972), main peculiarities of the model were analyzed and a numerical algorithm was proposed to deal with the problem. Asymptotic analysis of the solution near the crack tip for impermeable rock model was presented in Kemp (1989) and an approximate solution for the zero leak-off case, with an accuracy to the first four leading asymptotic terms, was given.[1] Here, probably for the first time, the speed equation was efficiently implemented in the model. The fourth degree of the crack opening was considered as the proper variable and used in the numerical computations. Finally, Kemp (1989) suggested to use a special tip element, compatible with the asymptotic behaviour of the solution, within the finite volume (FV) scheme.

For the leak-off function defined by the Carter law the two leading terms of asymptotics can be found in Kovalyshen and Detournay (2009), where the PKN problem was revisited to take into account the multiscale arguments in the spirit of Garagash et al. (2011). The authors also used the FV scheme with a special tip element to tackle the transient regime.

Recently, the classical models of Nordgren and Spence and Sharp have been revisited again in Linkov (2011a,b,c,d). The author discovered that in some formulations the hydraulic fracture problem may exhibit ill-posed properties. To eliminate the difficulties resulting from these facts, a number of measures were proposed: (i) speed equation to trace the crack front instead of the usually applied total flux balance condition;[2] (ii) the so-called ε-regularization technique which consists of imposing the computational domain boundary at a small distance behind the crack tip; (iii) new boundary condition to be imposed in the regularized formulation; (iv) new dependent variables: the particle velocity and the crack opening taken in a degree to exploit the asymptotic behaviour of the solution; (v) the spatial coordinates moving with the crack front and evaluation

[1] Extension of this solution to the full series representation was given in Kovalyshen and Detournay (2009), while other form, in terms of fast converging series, was obtained in Linkov (2011c).

[2] Probably, for the first time, this idea was recalled in indirect way in Spence and Sharp (1985) and utilized by Kemp (1989), but later was abandoned as the particle velocity at the crack tip is difficult to compute numerically.

of the temporal derivative under fixed values of these coordinates.

The advantages of the modified formulation were clearly demonstrated on the basis of developed analytical benchmarks in Linkov (2011d). An immense improvement of solution accuracy, computation efficiency and stability was shown. In Mishuris et al. (2012) a further step in employing the modified formulation was done by analyzing the stiffness of the system of differential equations arising after spatial discretization. An efficient modification of the algorithm has been proposed to trace the fracture propagation. Finally, the investigation of solution sensitivity to some process parameters has been considered.

However, the aforementioned analysis is concerned with the case when the leak-off vanishes near the crack tip and, as a result, the fluid velocity does not change much in this region.

The primary aim of this paper is to verify the recipes delivered in Linkov (2011a,b,c,d) and Mishuris et al. (2012) for an arbitrary leak off regime. In particular, the analysis includes: (i) Investigation of performance of numerical algorithms for hydrofracturing based on different formulations (different dependent and independent variables); (ii) Utilization of a new dependent variable defined as an integral of the crack opening. Such a variable has a clear physical and technological sense and is not related to any specific type of solution asymptotic behaviour near the crack tip; (iii) Modification of the way to impose boundary conditions in the framework of the ε-regularization technique[3]; (iv) Identification of optimal ranges of the various technique parameters, accuracy of computations and stiffness of the resulting dynamical systems; (v) Comparison of the solution for the Carter leak-off model with the numerical results from Kovalyshen and Detournay (2009) and discussion on the solution sensitivity.

The structure of this paper is as follows. In the next section we collect known results for the PKN model in various formulations. We restrict ourselves only to the information which is absolutely necessary to understand the paper. Section 2.3.2 contains an alternative formulation in terms of a new proper dependent variable—fracture volume. In Sect. 3 we discuss in details the numerical procedures and present main results of the computations comparing performances of different solvers under considerations. The solution obtained for the Carter leak-off model is compared with the available numerical data from Kovalyshen and Detournay (2009). The main findings of this paper are summarized in the conclusion section. Some new results concerning the asymptotic behaviour of the solutions for different leak-off regimes are presented also in "Appendices A and B".

[3] Note that the problem regularization is the important issue. It can be done by various techniques. Another type of the direct regularisation is shown in Wrobel and Mishuris (2013)

2 Problem formulation and preliminary results

2.1 Physical fundamentals and basic equations

Consider a symmetrical crack of the length $2l$ situated in the plane $x \in (-l, l)$, where the length $l = l(t)$ is one of the solution components changing as a result of the fluid flow inside the crack. The initial crack length is assumed to be nonzero: $l(0) = l_* > 0$. There are reasonable motivations behind this assumption. Namely, for the initial (unstable) stage of the crack propagation the acceleration of the process is too large to neglect the inertial terms. For this reason, any classical model not accounting for this effect is not credible. On the other hand, in many cases the hydrofracturing process is associated with the so-called 'perforation technique'. The latter consists of the creation of a number of finger-shaped initial fractures by detonations of shaped charges spaced along the wellbore (Economides and Nolte 2000). In this way, hydrofracturing starts simultaneously from a number of non-zero length cracks. Moreover, in many rock formations (e.g. shale reservoirs) cracks already exist but are closed by the confining stress (Economides and Nolte 2000). Finally, the uncertainties involved in this complex multiphysics problem itself do not allow one to make any reliable modelling of the crack nucleation in the rock formation.

By convention, we assume that the crack is fully filled by a Newtonian liquid injected at known rate $q_0(t)$ at the crack mouth $x = 0$.

The Poiseulle equation for the Newtonian liquid flow in a narrow channel is written in the form:

$$q = -\frac{1}{M} w^3 \frac{\partial p}{\partial x}, \tag{1}$$

where $q = q(t, x)$ is the fluid flow rate and $w = w(t, x)$ is the crack opening, while $p = p(t, x)$ is the net fluid pressure, that is, the difference between the

fluid pressure p_f inside the fracture and the confining stress σ_0 ($p = p_f - \sigma_0$). The constant M involved in the equation is defined as $M = 12\mu$, where μ stands for the dynamic viscosity [see for example Economides and Nolte 2000].

The continuity equation, accounting for the crack expansion and the leak-off of the fluid, may be expressed as:

$$\frac{\partial w}{\partial t} + \frac{\partial q}{\partial x} + q_l = 0, \quad t > 0, \quad 0 < x < l(t), \tag{2}$$

where $q_l = q_l(t, x)$ is the volume rate of fluid loss to formation in the direction perpendicular to the crack surfaces per unit length of fracture.

Numerical algorithms for the PKN model, with and without leak-off, have been considered in Nordgren (1972), Kemp (1989), Kovalyshen (2010) and others. Improvements based on the speed equation and the ε-regularisation technique have been introduced in Linkov (2011c), Linkov (2011d), Mishuris et al. (2012), for the case when leak-off vanishes near the crack tip. Below, we discuss the effectiveness of this approach on the three most popular leak-off models: Carter law (Carter 1957), modified law incorporating pressure difference (Clifton and Wang 1988) and bounded leak-off near the crack tip. An extensive discussion on possible behaviour of the leak-off function can be found in Kovalyshen (2010).

In our numerical simulations, we utilise one of the following leak-off variants:

$$q_l(t, x) = q_l^{(j)}(t, x) + q_j^*(t, x), \quad j = 1, 2, 3, \tag{3}$$

where

$$\begin{aligned} q_l^{(1)} &= \frac{C_1(t)}{\sqrt{t - \tau(x)}}, \quad q_l^{(2)} = \frac{C_2(t)p}{\sqrt{t - \tau(x)}}, \\ q_l^{(3)} &= C_{31}(t)p + C_{32}(t), \quad 0 < x < l(t). \end{aligned} \tag{4}$$

Here $C_1 = C_L$ is usually assumed to be a known constant defined experimentally (Carter 1957). Recently it was estimated analytically for a poro-elastic material in Kovalyshen (2010). The function $\tau(x)$ contains information on the history of the process. It defines the time at which the fracture tip reaches the point x and can be computed as the inverse of the crack length:

$$\tau(x) = l^{-1}(x), \quad x > l_*. \tag{5}$$

For $x \le l_*$ we conventionally set $\tau(x) \equiv 0$. Other constants in (4), $C_j(t) = C_j(t, w, p)$, $(j = 2, 3)$ may depend on the solution itself but are bounded functions in time. Finally, we assume that the terms q_j^*, $(j = 1, 2, 3)$ in (4) are negligible in comparison with $q_l^{(j)}$ near the crack tip. Note that application of the Carter leak-off law (Carter 1957) which is a simplified model of established fluid diffusion through the fracture walls, may be not justified at some stages of the process (Nordgren 1972; Lenoach 1995; Mathias and Reeuwijk 2009).

In this paper, we are aiming to build a general numerical framework for the problem under consideration. Thus, the collection of possible leak-off representations given in (4) covers the whole spectrum of possible bahaviours used in the hydrofracturing simulations (Kovalyshen 2010).

The system of Eqs. (1)–(2) should be supplemented by the elasticity equation. We consider the simplest relationship used in the PKN formulation

$$p = kw, \tag{6}$$

with a known proportionality coefficient $k = \frac{2}{\pi h}\frac{E}{1-\nu^2}$ found from the solution of a plane strain elasticity problem for an elliptical crack of height h (Nordgren 1972). Constants E and ν are the Young modulus and the Poisson ratio, respectively. In physical interpretation, this condition refers to the case when the fracture resistance of the solid is so small, that the energy dissipated by the fracture extension is negligible compared to the energy dissipated in the viscous fluid flow (Adachi and Detournay 2002). However, it turns out that, even in the models where the toughness dominated regime can be accounted for, it may be of a minor importance. For example, in Savitski and Detournay (2002) it has been proven that *radial hydraulic fractures in impermeable rocks generally propagate in the viscosity regime, and that the toughness regime is relevant only in exceptional circumstances* (for the average values of the field parameters the fracture would remain in the viscosity dominated regime for many years).

On substitution of the Poiseulle Eq. (1) and elasticity relationship (6) into the continuity Eq. (2), one obtains a well known lubrication (Reynolds) equation defined in the trapezoidal domain ($t > 0,\ 0 < x < l(t)$):

$$\frac{\partial w}{\partial t} - \frac{k}{M}\frac{\partial}{\partial x}\left(w^3\frac{\partial w}{\partial x}\right) + q_l = 0. \tag{7}$$

Since the system has its natural symmetry with respect to variable x and the equations are local, it is convenient to consider only half (symmetrical part) of the interval $[0, l(t)]$ instead of the full crack length $[-l(t), l(t)]$.

Following the discussion on the initial crack length above, the initial conditions for the problem are:

$$l(0) = l_*, \quad w(0,x) = w_*(x), \quad x \in (0, l_*). \tag{8}$$

The boundary conditions include: known fluid injection rate at the crack mouth, q_0, zero crack opening and zero fluid flux rate at the crack tip:

$$q(t,0) = q_0(t), \quad w(t,l(t)) = 0, \quad q(t,l(t)) = 0. \tag{9}$$

Note that the problem formulated in this way looks overdetermined as the governing Eq. (7) is of the second order with respect to spatial variable. This issue shall be discussed later.

Finally, by consecutive integration of Eq. (7) over time and then space, one can also derive the standard formula for the global fluid balance in the form:

$$\int_0^{l(t)} [w(t,x) - w_*(x)]dx - \int_0^t q_0(t)dt + \int_0^{l(t)}\int_0^t q_l(t,x)dtdx = 0, \tag{10}$$

where it is accepted that $w_*(x) = 0$ when $x > l_*$ and $l'(t) \geq 0$.

As has been shown in Linkov (2011d), the crucial role in the analysis of the problem plays the particle velocity defined in the following manner:

$$V(t,x) = \frac{q}{w}, \quad t > 0, \quad 0 \leq x \leq l(t), \tag{11}$$

which indicates the average velocity of fluid flow through the cross-sections of the fracture.

Under the assumption that the crack is fully filled by the fluid and sucking, ejection or discharge through the front can be neglected, the fluid velocity defines the crack propagation speed and the following speed equation is valid (Kemp 1989; Linkov 2011a,b)[4]:

$$l'(t) = V\,(t, l(t))\,, \quad t > 0. \tag{12}$$

Moreover, for physical reasons, one can assume that the fluid velocity at the crack tip is finite

$$0 \leq V\,(t,x) < \infty, \quad t > 0, \quad x \leq l(t) \tag{13}$$

Note that, allowing the crack propagation speed to be infinite, one has to simultaneously include the inertia term in the equations. Thus, the estimate (13) is a direct consequence of neglecting the acceleration terms.

[4] In fact, the speed equation in this form is valid only under the assumption of zero spurt loss at the crack tip (Nordgren 1972; Clifton and Wang 1988; Adachi et al. 2007)

2.2 Asymptotic behaviour of the solution and its consequences

As was mentioned in Spence and Sharp (1985), the fact that both w and q are present in (11), creates serious difficulties when trying to use the fluid velocity as a variable. However, as shown in Linkov (2011a,b,c,d), proper usage of fluid velocity may be extremely beneficial. First, it allows one to replace two boundary conditions at the crack tip $(9)_{2,3}$ with a single one additionally incorporating information from the speed Eqs. (12), (13).

Indeed, the boundary conditions $(9)_{2,3}$ in view of (1) and (6) lead to the estimate

$$w(t,x) = o\left((l(t) - x)^{\frac{1}{4}}\right), \quad x \to l(t), \tag{14}$$

which does not necessarily guarantee (13). However, further analysis of the problem, for different leak-off functions (see Kemp 1989; Kovalyshen and Detournay 2009 and "Appendix B" of this paper), shows that the particle velocity is bounded near the crack tip and the crack opening exhibits the following asymptotic behaviour:

$$w(t,x) = w_0(t)\big(l(t) - x\big)^{\frac{1}{3}} + w_1(t)\big(l(t) - x\big)^{\alpha} + o\big((l(t) - x)^{\alpha}\big), \quad \text{as} \quad x \to l(t), \tag{15}$$

with some $\alpha > 1/3$. For the classical PKN model for an impermeable solid (or when leak-off vanishes near the crack tip at least as fast as the crack opening) the exponent $\alpha = 4/3$ was found in Kemp (1989). For the case of the singular Carter's type leak-off, the exponent $\alpha = 1/2$ was determined in Kovalyshen and Detournay (2009).

Note that the asymptotics (15) shows that fluid velocity is indeed bounded near the crack tip. Moreover,

$$V(t,x) = V_0(t) + V_1(t)\big(l(t) - x\big)^{\beta} + o\big((l(t) - x)^{\beta}\big), \tag{16}$$

as $x \to l(t)$, where $\beta = \alpha - 1/3$ and

$$V_0 = \frac{k}{3M} w_0^3(t), \quad V_1 = \frac{k}{M}\left(\alpha + \frac{2}{3}\right) w_0^2(t) w_1(t). \tag{17}$$

As follows from "Appendix B", $V(t,x)$ may not be so smooth near the crack tip as one could expect and the exponent β in (16) plays an important role for this. Indeed, if $\beta \geq 1$ then $V(t,\cdot) \in C^1[0, l(t)]$ and the

particle velocity function is smooth enough near the crack tip. However, this happens only in the special case of $\alpha = 4/3$ when $V_x(t, x)$ is bounded near the crack tip. In case of singular leak-off ($0 < \beta < 1$), the particle velocity near the crack tip is only of the Hölder type $V(t, \cdot) \in C^1[0, l(t)) \bigcap H^\beta[0, l(t)]$. In "Appendix B" we present an exact form of the asymptotic expansion (15), which yields the aforementioned smoothness deterioration of V near the crack tip for the singular leak-off models.

Note that estimate (15) [or (16)] is equivalent to the condition (13). Thus, in view of (11), the pair of conditions $(9)_2$ and $(9)_3$ is equivalent to $(9)_2$ and (15). This discussion clearly illustrates why accounting for asymptotic behaviour of the solution in form (15) is of crucial importance for effective numerical realisation of any algorithm utilised in hydrofracturing (Adachi and Peirce 2007; Garagash et al. 2011; Mitchell et al. 2007). On the other hand, the fact that the particle velocity function is smooth enough near the crack tip has been one of the important arguments to use *the speed equation and proper variable approach* as the basis for improvement of the existing numerical algorithms (Linkov 2011a). It should be emphasized that behaviour of $V(t, x)$ near the crack tip may have serious implications when using ε-regularization technique.

Therefore, the main aim of this paper is to show that, regardless of possible smoothness of the particle velocity near the crack tip, the approach proposed in Linkov (2011a,b,c,d) *and* Mishuris et al. (2012) *is still efficient.*

2.3 Normalised formulation

Let us normalize the problem by introducing the following dimensionless variables:

$$\tilde{x} = \frac{x}{l(t)}, \quad \tilde{t} = \frac{t}{t_n}, \quad t_n = \frac{M}{kl_*}, \quad \tilde{w}_*(\tilde{x}) = w_*(x),$$
$$\tilde{w}(\tilde{t}, \tilde{x}) = \frac{w(t, x)}{l_*}, \quad \tilde{V}(\tilde{t}, \tilde{x}) = \frac{t_n}{l_*} V(t, x), \quad L(\tilde{t}) = \frac{l(t)}{l_*},$$
$$l_*^2 \tilde{q}_0(\tilde{t}) = t_n q_0(t), \quad l_* \tilde{q}_l(\tilde{t}, \tilde{x}) = t_n q_l(t, x), \tag{18}$$

where $\tilde{x} \in (0, 1)$ and $L(0) = 1$.

Using this notation, one defines the normalised particle velocity as:

$$\tilde{V}(\tilde{t}, \tilde{x}) = -\frac{\tilde{w}^2}{L(\tilde{t})} \frac{\partial \tilde{w}}{\partial \tilde{x}}. \tag{19}$$

The conservation law (2) in the normalised domain is rewritten in the following manner:

$$\frac{\partial \tilde{w}}{\partial \tilde{t}} = \frac{1}{L(\tilde{t})} \left[\left(\tilde{x} \tilde{V}(\tilde{t}, 1) - \tilde{V}(\tilde{t}, \tilde{x}) \right) \frac{\partial \tilde{w}}{\partial \tilde{x}} - \tilde{w} \frac{\partial \tilde{V}}{\partial \tilde{x}} \right] - \tilde{q}_l(\tilde{t}, \tilde{x}), \tag{20}$$

The leading terms of the asymptotic estimate of the leak-off function from (4) are now:

$$\tilde{q}_l^{(1)}(\tilde{t}, \tilde{x}) = \frac{\tilde{C}_1(\tilde{t}) D(\tilde{t})}{\sqrt{1 - \tilde{x}}}, \quad \tilde{q}_l^{(2)}(\tilde{t}, \tilde{x}) = \frac{\tilde{C}_2(\tilde{t}) D(\tilde{t})}{\sqrt{1 - \tilde{x}}} \tilde{w}(\tilde{t}, \tilde{x}),$$
$$\tilde{q}_l^{(3)}(\tilde{t}, \tilde{x}) = \tilde{C}_{31}(\tilde{t}) \tilde{w}(\tilde{t}, \tilde{x}) + \tilde{C}_{32}(\tilde{t}). \tag{21}$$

Here, the function

$$D(\tilde{t}) = \sqrt{\frac{L'(\tilde{t})}{L(\tilde{t})}}, \tag{22}$$

is introduced in the "Appendix A", where the remainder between the normalised total flux and the leading term (21) has been effectively estimated. Thus the normalised term $\tilde{q}_j^*(\tilde{t}, \tilde{x})$ vanishes near the crack tip faster than the solution itself.

Finally, normalised initial conditions (8) and boundary conditions (9) are:

$$L(0) = 1, \quad \tilde{w}(0, \tilde{x}) = \tilde{w}_*(\tilde{x}), \quad x \in (0, 1), \tag{23}$$

and

$$-\frac{1}{L(\tilde{t})} \tilde{w}^3 \frac{\partial \tilde{w}}{\partial \tilde{x}}(\tilde{t}, 0) = \tilde{q}_0(\tilde{t}), \quad \tilde{w}(\tilde{t}, 1) = 0. \tag{24}$$

The global fluid balance (10) can be written in the form:

$$L(\tilde{t}) \int_0^1 \tilde{w}(\tilde{t}, x) dx - \int_0^1 \tilde{w}(x, 0) dx - \int_0^{\tilde{t}} \tilde{q}_0(t) dt + \int_0^{\tilde{t}} L(t) \int_0^1 \tilde{q}_l(t, x) dx dt = 0. \tag{25}$$

For convenience, from this point on we will omit the "$\sim$" symbol for all dependent and independent variables and will only consider the respective dimensionless values.

Note that the particular representation (20) of the Reynolds equation highlights an essential feature of the problem - it is singularly perturbed near the crack tip. Indeed, both coefficients in front of the spatial derivatives on the right-hand side of the Eq. (20) tend to

zero at $x = 1$. Thus, the asymptotic behaviour of the solution near the crack tip ($x \to 1$)

$$w = w_0(t)\,(1-x)^{\frac{1}{3}} + w_1(t)\,(1-x)^{\alpha} + o\left((1-x)^{\alpha}\right), \tag{26}$$

$$V = V_0(t) + V_1(t)\,(1-x)^{\alpha-\frac{1}{3}} + o\left((1-x)^{\alpha-\frac{1}{3}}\right), \tag{27}$$

represents nothing but the boundary layer. Moreover, normalizing $(17)_1$ one obtains:

$$V_0(t) = \frac{1}{3L(t)} w_0^3(t). \tag{28}$$

The terms $w_0,\ w_1$ and $V_0,\ V_1$ in (26) and (27) are different from those in (15) and (16). In fact, the former should be written with "$\sim$" symbol.

On substitution of (19) into (20), one eliminates the particle velocity function from Reynolds equation:

$$\frac{\partial w}{\partial t} = \frac{1}{L^2(t)}\left[\frac{1}{3} w_0^3 x \frac{\partial w}{\partial x} + 3\tilde{w}^2 \left(\frac{\partial w}{\partial x}\right)^2 + w^3 \frac{\partial^2 w}{\partial x^2}\right] - q_l. \tag{29}$$

Here w_0 is the multiplier of the first term of the asymptotic expansion (15). This form of lubrication equation exhibits the same degenerative properties as (20). Also the coefficients appearing in front of the leading terms tend to zero near the crack tip.

The speed Eq. (12) defining the crack propagation speed is given in the normalised variables as:

$$L'(t) = V_0(t), \quad t > 0. \tag{30}$$

Taking into account (28), the latter can be rewritten in the following form

$$\frac{d}{dt} L^2 = \frac{2}{3} w_0^3(t), \qquad t > 0. \tag{31}$$

This equation serves us to determine the unknown value of the crack length $L(t)$. As it has been shown in Mishuris et al. (2012), such an approach has clear advantages over the standard one based on the global fluid balance Eq. (25).

As a result of the foregoing transformations, one can formulate a system of PDEs describing the hydrofracturing process. The system is composed of two operators:

$$\frac{d}{dt} w = \mathcal{A}_w\left(w, L^2\right), \quad \frac{d}{dt} L^2(t) = \mathcal{B}_w(w), \tag{32}$$

where $\mathcal{A}_w$ is defined by the right-hand side of Eq. (29) with the boundary conditions $(24)_{1,2}$, while the second operator $\mathcal{B}_w$ is given by (31). The system is equipped with the initial conditions:

$$L(0) = 1, \quad w(0, x) = w_*(x), \quad x \in (0, 1). \tag{33}$$

2.3.1 Reformulation of the problem in proper dependent variables. First approach

In Linkov (2011d) and later in Mishuris et al. (2012) it has been shown that the dependent variable

$$U(t, x) = w^3(t, x) \tag{34}$$

is more favorable for the solution of the system (32), (33) than the crack opening itself. This idea is based on the fact that, according to the asymptotics of the solution near the crack tip, the dependent variable U is much smoother than w. In the case of an impermeable solid, the solution U is analytic in the closed interval [0, 1] (see Linkov 2011d). However, the type of leak-off function is of significant importance here. Thus, adopting asymptotic representation (26), one can see that for $x \to 1$

$$U = U_0(t)(1-x) + U_1(t)(1-x)^{\frac{2}{3}+\alpha} + o((1-x)^{\frac{1}{3}+2\alpha}), \tag{35}$$

where the coefficients $U_0(t)$ and $U_1(t)$ are directly related to those appearing in the crack opening formulation:

$$U_0(t) = w_0^3(t), \quad U_1(t) = w_0^2(t) w_1(t). \tag{36}$$

Depending on the type of leak-off described in (4), the exponent in the second asymptotic term $\frac{2}{3}+\alpha$ may take value $3/2,\ 11/6$ or 2, respectively. Thus in the first two cases, the transformation (34) no longer results in polynomial representations of asymptotic expansion for U. For this reason, the advantage of the approach using variable U in more general cases, when the leak-off is singular near the crack tip, should still be confirmed. This is one of the aims of this paper. On the other hand, at least two factors work in favor of this formulation. First, the spatial derivative of U is not singular and, second, the particle velocity is given by a linear relationship

$$V(t, x) = -\frac{1}{3L(t)} \frac{\partial U}{\partial x}. \tag{37}$$

The governing Eq. (20) in terms of the new variable can be written in the normalized domain $x \in (0, 1)$ as:

$$\frac{\partial U}{\partial t} = \frac{1}{L(t)}\left[(xV(t,1) - V(t,x))\frac{\partial U}{\partial x} - 3U\frac{\partial V}{\partial x}\right] - 3U^{\frac{2}{3}}q_l, \tag{38}$$

Similarly to (29) the particle velocity function may be eliminated from the lubrication equation:

$$\frac{\partial U}{\partial t} = \frac{1}{3L^2(t)}\left[xU_0\frac{\partial U}{\partial x} + \left(\frac{\partial U}{\partial x}\right)^2 + 3U\frac{\partial^2 U}{\partial x^2}\right] - 3U^{\frac{2}{3}}q_l. \tag{39}$$

Note that Eqs. (38)–(39) are of a very similar structure to those evaluated for the crack opening w. They exhibit the same degenerative nature near the crack tip.

Finally, boundary conditions (24) transform to:

$$-\frac{\sqrt[3]{U(t,0)}}{3L(t)}\frac{\partial}{\partial x}U(t,0) = q_0(t), \quad U(t,1) = 0, \tag{40}$$

while the speed Eq. (31) takes the following form:

$$\frac{d}{dt}L^2 = \frac{2}{3}U_0(t), \quad t > 0. \tag{41}$$

The system of PDEs equivalent to (32) is now defined as:

$$\frac{d}{dt}U = \mathcal{A}_U(U, L^2), \quad \frac{d}{dt}L^2(t) = \mathcal{B}_U(U). \tag{42}$$

The operator $\mathcal{A}_U$ is described by (39) with boundary conditions (40), while the second operator $\mathcal{B}_U$ is given by (41). Finally, the initial conditions are similar to those in the previous formulation (23):

$$L(0) = 1, \quad U(0,x) = w_*^3(x), \quad x \in (0,1). \tag{43}$$

2.3.2 *Reformulation of the problem in proper dependent variables. Second approach*

The aforementioned formulation of the problem in terms of the dependent variable U has one considerable drawback. It is well known that for different elasticity models and different hydrofracturing regimes one has various asymptotic behaviours of the solution near the crack tip (Adachi and Detournay 2002). For example, for exact equations of elasticity theory and the zero toughness condition ($K_{IC} = 0$), the exponent of the leading term of w varies from 2/3, for the Newtonian fluid, to 1, for the ideally plastic fluid. Thus, the same reformulation to the type of the proper variable might be inconvenient, or even impossible.

For this reason, we introduce another dependent variable. Although it does not transform the asymptotic behaviour of the solution in such a smooth manner as it has been done previously when adopting U, this variable has its own advantages. Namely, let us consider a new dependent variable Ω defined as follows:

$$\Omega(t,x) = \int_x^1 w(t,\xi)d\xi. \tag{44}$$

This variable is not directly related to any particular asymptotic representation of $w(x,t)$, however it assumes that $w \to 0$ for $x \to 1$. As a result the form of governing equations for Ω remains the same regardless of $w(x,t)$ asymptotics, i.e. this formulation has a general (universal) character. Note, that in case of U the optimal way of transformation for the lubrication equation essentially depends on the exact form of asymptotic expansion (the leading term) for w.
Another advantage of Ω comes from the fact that it has clear physical and technological interpretation. Namely, it reflects the crack volume measured from the crack tip.

Asymptotics of the function Ω near the crack tip takes the following form:

$$\Omega(t,x) = \Omega_0(t)(1-x)^{\frac{4}{3}} + \Omega_1(t)(1-x)^{\alpha+1} + o((1-x)^{\alpha+1}), \quad x \to 1, \tag{45}$$

where the coefficients $\Omega_0(t)$ and $\Omega_1(t)$ are related to those in (26):

$$\Omega_0(t) = \frac{3}{4}w_0(t), \quad \Omega_1(t) = \frac{1}{\alpha+1}w_1(t). \tag{46}$$

Thus, similarly to U, the new variable is smoother than the crack opening, w, and the first singular derivative of Ω is that of the second order.

By spatial integration of (20) from x to 1 and substitution of (44) one obtains:

$$\frac{\partial \Omega}{\partial t} = -\frac{1}{L(t)}\left[\left(V(t,x) - xV(t,1)\right)\frac{\partial \Omega}{\partial x} + V(t,1)\Omega\right] - Q_l, \tag{47}$$

where the monotonicity of $L'(t) > 0$ has been taken into account and

$$Q_l(t,x) = \int_x^1 q_l(t,\xi)d\xi.$$

Here, the particle velocity (19) is computed in the manner:

$$V(t,x) = \frac{1}{3L(t)}\frac{\partial}{\partial x}\left(\frac{\partial \Omega}{\partial x}\right)^3. \tag{48}$$

By eliminating $V(x,t)$ from the Eq. (47) we derive a new formula for the lubrication equation:

$$\frac{\partial \Omega}{\partial t} = -\frac{1}{L^2(t)}\left[\left(\frac{\partial \Omega}{\partial x}\right)^3 \frac{\partial^2 \Omega}{\partial x^2} + \frac{64}{81}\Omega_0^3\left(\Omega - x\frac{\partial \Omega}{\partial x}\right)\right] - Q_l. \tag{49}$$

The boundary conditions (24) are expressed in the following way:

$$-\frac{1}{L(t)}\left(\frac{\partial \Omega}{\partial x}\right)^3 \frac{\partial^2 \Omega}{\partial x^2}(t,0) = q_0, \quad \frac{\partial \Omega}{\partial x}(t,1) = 0. \tag{50}$$

Interestingly, the first boundary condition, directly substituted into the lubrication Eq. (47) can be equivalently rewritten in the form

$$\frac{\partial \Omega}{\partial t}(t,0) = -\frac{64}{81^2 L(t)}\Omega(t,0)\Omega_0^3(t) + \frac{q_0(t)}{L(t)} - Q_l(t,0), \tag{51}$$

This condition, in turn, represents nothing but the local (in time) flux balance condition. To verify this, it is enough to apply the time derivative to the Eq. (25). Furthermore it appears much easier for implementation into a numerical procedure than $(50)_1$ itself, but may lead to some increase of the problem stiffness, as we will show later.

It is easy to check, by using the governing Eq. (47) and limiting values of all its terms for $x \to 1$, that a weaker boundary condition

$$\Omega(t,1) = 0 \tag{52}$$

is equivalent to the original one $(50)_2$. Finally, the speed Eq. (31) in the Ω formulation assumes the following form:

$$\frac{d}{dt}L^2(t) = \frac{128}{81}\Omega_0^3(t). \tag{53}$$

In this way we obtain another system of PDEs that is composed of two operator relations:

$$\frac{d}{dt}\Omega = \mathcal{A}_\Omega(\Omega, L^2), \quad \frac{d}{dt}L^2(t) = \mathcal{B}_\Omega(\Omega), \tag{54}$$

where, as previously, $\mathcal{A}_\Omega$ is defined by (29) with boundary conditions $(50)_{1,2}$ or (51) and (52). The second operator, $\mathcal{B}_\Omega$, is given by Eq. (53). Here the initial conditions are obtained from (23):

$$L(0) = 1, \quad \Omega(0,x) = \Omega_*(x) \equiv \int_x^1 w_*(\xi)d\xi. \tag{55}$$

2.4 ε-Regularization and the respective boundary conditions

In our analysis we are going to use the so-called ε-regularization technique. It was originally introduced in Linkov (2011a) for the system of spatial coordinates moving with the fracture front. In Mishuris et al. (2012), the authors efficiently adopted the approach for the normalised coordinate system.

The reason to separate the domain from the end point $x = 1$ by a small distance of ε and to introduce ε-regularisation has been thoroughly described in Linkov (2011a). It consists of replacing the Dirichlet boundary condition $(40)_2$ with an approximate one:

$$U(t, 1-\varepsilon) = 3\varepsilon L(t) V(t,1), \tag{56}$$

emerging from deep physical arguments. The value of the crack propagation speed $V(t,1)$ (and simultaneously the particle velocity at a fracture tip) was suggested to be computed from the speed Eq. (41) in its approximated form:

$$V(t,1) = -\frac{1}{3L(t)}\frac{\partial U}{\partial x}(t, 1-\varepsilon). \tag{57}$$

The pair of conditions (56) – (57) has shown an excellent performance in terms of solution accuracy and, as has been proven in Mishuris et al. (2012), reduced the stiffness of dynamic system in case of leak-off function vanishing near the crack tip. One can check that for such a leak-off model numerical error introduced by using the approximate conditions, instead of the exact ones, is of the order $O(\varepsilon^2)$. In view of all improvements following from utilisation of the regularized conditions (Linkov 2011d; Mishuris et al. 2012) such a strategy is fully justified and in fact inevitable.

The conditions can be written in an equivalent form. Indeed, one can merge (56) and (57) into a single condition of the third type:

$$U(t, 1-\varepsilon) + \varepsilon\frac{\partial U}{\partial x}(t, 1-\varepsilon) = 0. \tag{58}$$

Interestingly, the latter condition is nothing but the consequence of a direct utilization of the information about

the leading term of asymptotics of the solution near the crack tip [compare with (35)].

Analogously, one can define the respective pairs of boundary conditions in the regularized formulations. Considering the dependent variable w one should take (31) together with the condition

$$w(t, 1-\varepsilon) + 3\varepsilon \frac{\partial w}{\partial x}(t, 1-\varepsilon) = 0, \tag{59}$$

while analysing the system based on the dependent variable Ω, the speed Eq. (53) should be accompanied by

$$4\Omega(t, 1-\varepsilon) + 3\varepsilon \frac{\partial \Omega}{\partial x}(t, 1-\varepsilon) = 0. \tag{60}$$

To conclude this subsection, one can make a prediction that in the case of a singular leak-off function, even when using the ε-regularization technique, the accuracy of the solution should be worse than that presented in Linkov (2011a) and Mishuris et al. (2012). However, it is always possible to use information on accurate asymptotic behaviour of the solution (employing higher order terms) and in this way improve the accuracy of computations.

3 Numerical solution of the dynamic systems

In this section, three alternative systems of PDEs (32), (42) and (54) describing the problem of hydrofracruing are transformed into the corresponding non-linear dynamic systems of the first order. Then, on the basis of respective analytical benchmarks, we analyze their stiffness properties, the accuracy and efficiency of computations. The benchmark solutions in question are described in "Appendix C".

3.1 Representation of the boundary conditions and the speed equation

Consider a spatial domain of the problem reduced in accordance with the ε-regularization technique to the interval $x \in [0, 1-\varepsilon]$, where ε is a small parameter. Let the mesh, $\{x_j\}_{j=1}^{N}$, be composed of N nodes with $x_1 = 0$ and $x_N = 1-\varepsilon$.

For each of the problem formulations, two boundary conditions should be accounted for: one specified at the crack inlet and a regularized boundary condition at $x = 1-\varepsilon$. In the following we present a brief description of how these conditions are introduced to the numerical scheme.

From now on, for the dependent variables discussed above ($w(t,x)$, $U(t,x)$, $\Omega(t,x)$), we use common notation $f(t,x)$ together with a convention $f_k = f(t, x_k)$.

To discretize the first boundary condition (depending on the formulation: $(24)_1$, $(40)_1$ or $(50)_1$) we exploit the smooth character of the solution near the point $x = 0$. Thus, accepting a polynomial approximation of $f(x,t)$ on the interval $x \in [x_1, x_3]$, the respective nonlinear relation between f_1, f_2 and f_3 may be derived:

$$A_1(f_1, t) f_1 + A_2(f_1, t) f_2 + A_3(f_1, t) f_3 = q_0. \tag{61}$$

As mentioned in 2.4, the regularized boundary condition in the ε-regularization technique proposed in Linkov (2011a) is equivalent to a mixed boundary condition based on the leading term of the asymptotic expansion [see (58), (59), (60)]. Below we propose a modification of this approach which consists in employing two terms of the asymptotics. We will show that such a strategy prevents the deterioration of accuracy when the solution is not so smooth as in the cases originally considered in Linkov (2011a) and Mishuris et al. (2012).

According to (26), (35) and (45), the following asymptotics approximation is acceptable in the proximity of the crack tip ($x \in [x_{N-2}, 1]$):

$$f(t,x) = e_1^{(f)}(t)(1-x)^{\alpha_1} + e_2^{(f)}(t)(1-x)^{\alpha_2}. \tag{62}$$

The values of α_1 and α_2 are known in advance and depend, as has been discussed above, on the chosen variable and the behavior of the leak-off function. Assuming that the last three points of the discrete solution (x_{N-2}, f_{N-2}), (x_{N-1}, f_{N-1}) and (x_N, f_N) lie on the solution graph $(x, f(t,x))$, one can derive a formula combining all these values in one equation:

$$f_N + b_{N-1}^{(f)} f_{N-1} + b_{N-2}^{(f)} f_{N-2} = 0, \tag{63}$$

where $b_j^{(f)} = b_j^{(f)}(x_{N-2}, x_{N-1}, x_N)$. Relation (63) is consequently used to represent the regularised boundary condition at $x = 1-\varepsilon$.

Remark 1 In the authors opinion, the presented approach is a direct generalization of that proposed in Linkov (2011a). Indeed, if one takes $e_2 = 0$ in the representation (62) then the pair of the Eqs. (58) and (57) follows immediately. If $\alpha_2^{(f)} - \alpha_1^{(f)} = 1$, which means that the leak-off function q_l is bounded near the crack

tip, the second asymptotic term provides a small correction. However, in the case of the Carter law, when $\alpha_2^{(f)} - \alpha_1^{(f)} = 1/6$, it brings an important contribution and improves the accuracy of the computations, as will be shown later.

Finally, coefficient e_1 from (62) is substituted into the pertinent form of the speed Eq. (31), (41) or (53) to give the ordinary differential equations for the crack length:

$$\frac{d}{dt}L^2 = \frac{2}{3}\left(e_1^{(w)}\right)^3, \quad \frac{d}{dt}L^2 = \frac{2}{3}e_1^{(U)},$$
$$\frac{d}{dt}L^2(t) = \frac{128}{81}\left(e_1^{(\Omega)}\right)^3. \tag{64}$$

Note that the right-hand sides of the equations define the boundary operators $\mathcal{B}_w$, $\mathcal{B}_U$ and $\mathcal{B}_\Omega$ from $(32)_2$, $(42)_2$ or $(54)_2$, respectively.

Remark 2 As it follows from this analysis, the ε-regularization is, in a sense, equivalent to the introduction of a special tip element in the discrete solution. Thus, one can see a complementarity with the approach utilised in Kovalyshen (2010). However, and this is crucial for the analysis, only the speed equation together with ε-regularization allows to take into account both the local and global phenomena, and to do this in the most efficient way from the numerical point of view.

Remark 3 In the case of the dependent variable U, apart from the representations (62) of the boundary condition near the crack tip in the linear form

$$U_N = b_1^{(U)}U_{N-1} + b_2^{(U)}U_{N-2}, \tag{65}$$

one can use a nonlinear one, adopting the relationship between this dependent variable and the crack opening w:

$$U_N = \left(b_1^{(w)}\sqrt[3]{U_{N-1}} + b_2^{(w)}\sqrt[3]{U_{N-2}}\right)^3. \tag{66}$$

Note that the two terms representation (65) of the function U is less informative than the same representation for the functions w (or Ω) and thus, using the modified condition (66), one can expect a better solver performance.

3.2 Spatial discretization of the Reynolds equation: corresponding dynamic systems

Let us consider the Reynolds equations written in different dependent variables [(32), (42) or (54), respectively]. By representing the spatial derivatives in the right-hand sides of the corresponding equation by central three point finite difference schemes, we obtain a nonlinear system of $N-2$ ordinary differential equations for the values $f_i(t)$ at each internal point of the spatial domain $(x_2, \ldots, x_{N-1})$. The respective boundary conditions are embedded into the system through Eqs. (61) and (63).

Supplementing the system with the pertinent form of the speed Eq. (64), we obtain a non-linear dynamic system of first order describing the process of hydrofracturing which can be written in the form:

$$\mathbf{F}' = \mathbf{A}^{(f)}\mathbf{F} + \mathbf{G}^{(f)}, \tag{67}$$

where $\mathbf{F} = \mathbf{F}(t)$ is a vector of unknown solution $[f(t, x_1),\ f(t, x_2), \ldots, f(t, x_N), L^2(t)]$ of dimension $N-1$. Note that matrix $\mathbf{A}^{(f)}$ and vector $\mathbf{G}^{(f)}$ depend, generally speaking, on the solution. Matrix $\mathbf{A}$ is the so-called mass matrix of the system, in the case of which a tri-diagonal form prevails (however the boundary conditions and the last equation for $L^2(t)$ disturb the tri-diagonal structure).

Remark 4 In the case when the boundary condition in formulation (51) is in use, the dimension of the dynamic system is N. Indeed, this condition has the form of an ODE, and thus can be substituted directly into the system as an additional equation.

In our numerical computations two different types of spatial meshes are used. The first one is a regular mesh with uniformly spaced nodes, while the second one gives an increased nodes density when approaching the crack tip. Both types of meshes can be described by the formula:

$$x_m(\delta) = 1 - \left(1 - \left(1 - \varepsilon^{\frac{1}{\delta}}\right)\frac{m}{N}\right)^{\delta}, \quad m = 1, ..., N. \tag{68}$$

In the foregoing, the parameter δ defines the mesh type. Namely, for $\delta = 1$ one has the uniform mesh (henceforth denoted as $x^{(I)}$), since any $\delta > 1$ gives the nodes concentration near the crack tip (this mesh will be referred to as $x^{(II)}(\delta)$). Mesh $x^{(II)}(\delta)$ allows one to choose appropriate parameter δ to suppress the stiffness of dynamic system or to increase the solution accuracy.

The stiffness of a dynamic system may be described by the condition number or the condition ratio (Aiken 1985) of a mass matrix $\mathbf{A}^{(f)}$. In general, the values given by various measures are different (see some consequences in Mishuris et al. 2012). In this paper we

use the condition ratio as the measure of the system stiffness. Some rough estimation of this parameter may help to choose an optimal value of δ from the stiffness point of view. In the case under consideration, one can analyse the condition ratio of a simplified variant of the system (39), where only the leading term (with the second order derivative) is preserved and the nonlinear multiplier is substituted by the first term of the asymptotic expansion for U. It turns out that $\delta = 2$ gives the lowest possible stiffness. Naturally, in the general case, the optimal value of δ can be different. We have checked however that, for three alternative problem formulations, there are three different optimal values of the parameter, but each of them is very close to 2. Thus, in the following section all results concerning nonuniform mesh are presented for $\delta = 2$.

3.3 Stiffness analysis

In our analysis, we quantify the system stiffness by a condition ratio κ_A defined in (69). Since the problem is nonlinear, our investigation is to be done for the linearized form of matrix $\mathbf{A}^{(f)}$. It is obvious that for all six variants of benchmark solutions under consideration (see "Appendix C") one has different values of $\mathbf{A}^{(f)}$. Computations are carried out for two types of meshes (the uniform and non-uniform one). For each of the benchmark solutions, one obtains a constant value for the condition ratio (independent on time), apart from the fact that the matrix $\mathbf{A}^{(f)}$ depends on time. Those values of the condition ratio κ_A are, generally speaking, different for various benchmarks and chosen meshes.

Before comparing the results for various dependent variables, we have checked that for the dynamic system based on U the stiffness is almost identical for both forms of the regularized boundary condition at $x = 1 - \varepsilon$ [(58) and (63) respectively]. Thus, for the rest of the dependent variables (w and Ω) we restrict our stiffness investigation only to the formulation (63). Remarkably, the situation changes dramatically when one considers the accuracy of computations, which shall be discussed later on.

Computing the condition ratio for the next variants of the matrix linearization, we have confirmed the following estimate valid for large values of N for all cases under investigation:

$$\kappa_A^{(f)} = \frac{|\lambda_{max}|}{|\lambda_{min}|} \sim \varpi^{(f)} N^2, \quad N \to \infty. \tag{69}$$

Here $|\lambda_{max}|$, and $|\lambda_{min}|$ are the largest and smallest absolute values among the $\mathbf{A}^{(f)}$ matrix eigenvalues, while the constant $\varpi^{(f)}$ is to be estimated numerically. Its values for all six benchmark cases are shown in Table 1.

Although the qualitative character of the stiffness behaviour (N^2) is rather obvious, its quantitative measure described by ϖ can be used to select the optimal (in terms of the stiffness properties) variant of the dynamic system.

The following analysis includes investigation of stiffness sensitivity to: (i) the solution (benchmark) type, (ii) choice of the dependent variable, (iii) choice of the independent variable (spatial mesh), (iv) value of the regularization parameter ε.

Remark 5 As mentioned previously, in case of the variable Ω, there are two alternative ways to introduce the boundary condition at $x = 0$ to the system -formulations (51) and (61), respectively. In this way one can construct two alternative dynamic systems of different dimensions (N and $N - 1$). The results in Table 1 show that the stiffness properties of the system corresponding to the boundary condition (61) are slightly better than those of the system utilizing (51). Indeed, the respective parameter ϖ is about two times smaller. One of the possible explanations is the aforementioned difference in the systems' sizes: $\dim \mathbf{A}_{(1)}^{(\Omega)} = \dim \mathbf{A}_{(2)}^{(\Omega)} + 1$ (see Remark 3). We have checked that the accuracy of computations remains practically the same regardless of the system variant. Taking this fact into account, we restrict ourselves in the analysis only to the system which employs the boundary condition at point $x = 0$ in the form (61). Thus, from now on all the investigated dynamic systems (for all dependent variables) will be based on the same mechanisms for incorporation of the boundary conditions.

Influence of the value of ε on the condition ratio is analyzed in Fig. 1. As an example, we present here the benchmark $q_l^{(1)}$ for $Q_l/q_0 = 0.9$ (see "Appendix C"). The results were obtained for the uniform mesh. For other combinations of the benchmark solutions and different meshes the graphs have similar character. As anticipated, the estimation (69) holds true only for sufficiently large N. The threshold value of N depends on the chosen ε. Thus for a fixed number of grid points N, there is a critical value of the regularization parameter $\varepsilon_s(N)$ for which the stiffness characteristics changes its

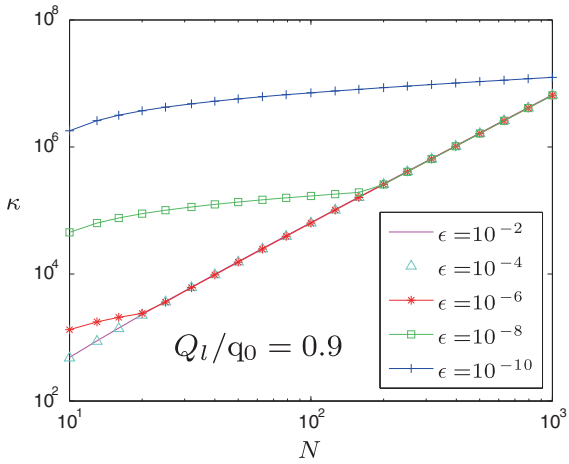

Fig. 1 Condition ratio $\kappa = \kappa^{(w)}(N)$ for the dynamic system based on w variable and different values of the regularization parameter ε. The case of the uniform mesh is analyzed

behaviour. By taking $\varepsilon < \varepsilon_s(N)$ one increases appreciably the system stiffness.

The results of the stiffness investigation are collected in the Table 1 and Fig. 1 The following conclusions can be drawn from this data:

(i) The nonuniform mesh reduces the stiffness approximately up to five times regardless of the solution type (Table 1);

(ii) The most important parameter affecting the stiffness properties is the relation between the injection flux rate and the leak-off to the formation Q_l/q_0 as can be clearly seen in Table 1. The value of this parameter is more important than a particular distribution of the leak-off function (and its behaviour near the crack tip);

(iii) When comparing systems for various dependent variables, the lowest condition ratio gives the system built for U (one order of magnitude lower than the others). The worst stiffness performance takes place for the system corresponding to the Ω-variable. However, in some cases Ω may produce lower stiffness than w;

(v) A value of the regularization parameter ε essentially affects the stiffness of a dynamic system.

3.4 Accuracy of the computations

In this section, we analyze the accuracy of computations by the solvers based on different dynamic systems corresponding to the respective dependent variables. To solve the systems, we use MATLAB ode15s subroutine dedicated for stiff dynamic systems which utilizes the implicit numerical differentiation formulas.

Before we compare different approaches in terms of their accuracy, let us recall two alternative ways to

Table 1 Values of the parameter $\varpi^{(f)}$ from the approximation of the condition ratio (69) for the different dynamic systems (67) and different benchmarks. The computations were provided for $\varepsilon = 10^{-3}$

	$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
	ϖ estimated for the system based on variable w					
$x^{(I)}$	6.5e+0	6.6e+0	6.8e+0	1.8e+1	1.8e+1	1.8e+1
$x^{(II)}$	1.7e+0	1.7e+0	1.7e+0	4.6e+0	4.7e+0	4.7e+0
	ϖ estimated for the system based on variable U					
$x^{(I)}$	3.0e+0	3.0e+0	3.2e+0	6.0e+0	6.1e+0	6.2e+0
$x^{(II)}$	7.5e−1	7.7e−1	8.1e−1	1.5e+0	1.5e+0	1.6e+0
	ϖ estimated for $\Omega_{(1)}$ based on condition (51)					
$x^{(I)}$	4.8e+1	4.8e+1	4.9e+1	1.7e+1	1.7e+1	1.7e+1
$x^{(II)}$	1.2e+1	1.2e+1	1.3e+1	4.3e+0	4.3e+0	4.3e+0
	ϖ estimated for $\Omega_{(2)}$ based on condition (61)					
$x^{(I)}$	2.3e+1	2.2e+1	1.9e+1	9.6e+0	1.0e+1	1.3e+1
$x^{(II)}$	5.8e+0	5.7e+0	4.7e+0	2.5e+0	2.6e+0	3.5e+0

Table 2 Comparison of the accuracy of the solution of dynamical system based on variable U. The results depicted by U_* refer to the regularized boundary condition based on one asymptotic term, while those denoted by U_l and U_n correspond to two terms approximation (linear (65) and nonlinear (66), respectively). Other problem parameters: $N = 100$, $\delta = 2$ for the mesh $x^{(II)}$

		Comparison of conditions (58), (65), (66)					
		$q_l^{(1)}$ $Q_l/q_0 = 0.9$			$q_l^{(3)}$ $Q_l/q_0 = 0.5$		
	ε	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
δU_*	$x^{(I)}$	1.6e−1	1.4e−1	1.3e−1	6.1e−3	3.7e−3	3.7e−3
	$x^{(II)}$	1.4e−1	7.6e−2	6.3e−2	4.5e−3	9.3e−5	8.9e−5
δU_l	$x^{(I)}$	5.0e−2	1.4e−2	1.7e−2	1.2e−5	1.1e−5	1.1e−5
	$x^{(II)}$	5.0e−2	1.7e−3	2.0e−3	4.9e−5	8.2e−5	8.7e−5
δU_n	$x^{(I)}$	4.4e−2	1.2e−2	1.3e−2	2.2e−5	1.3e−5	1.3e−5
	$x^{(II)}$	4.4e−2	9.9e−4	1.8e−3	4.2e−5	8.2e−5	8.7e−5

define the regularized boundary condition at the end point $x = 1 - \varepsilon$. The first one is based on the ε-regularization technique, as it was defined in Linkov (2011d) (see also Mishuris et al. 2012). The second approach to formulate the regularized condition is to take into account the first two terms of asymptotics as described in Sect. 3.1 [compare Eqs. (62) and (63)]. Finally, in the case of U, this condition may be implemented in the non-linear form (66). One can expect that the two term conditions would have a clear advantage, at least in cases when the solution smoothness near the crack tip deteriorates due to the singularity of the leak-off function.

The results of the computations presented in Table 2 confirm such a prediction. We compare only conditions for U, as originally the ε-regularization technique was introduced for this variable. Indeed, the relative errors of the solutions δU_l or δU_n are at least one order of magnitude lower than that in the case of δU_*, corresponding to the formulation based on (58).[5] Surprisingly, for the variants of the non-singular leak-off function, the improvement is even more pronounced (especially for a uniform mesh).

We also made the computations for three different benchmarks reported in Mishuris et al. (2012). They correspond to the leak-off function vanishing near the crack tip. It turned out that computational error corresponding to the modified form of the regularized conditions (based on two terms of asymptotics) was always two orders of magnitude lower than that reported in the previous paper.

On the other hand, there is no difference observed between the solutions δU_l but δU_n at least for those two benchmarks and the choice of the parameters ($N = 100$). However, as we will show later, for large numbers of nodal points, or more severe leak-off function relationship ($Q_l/q_0 \sim 1$), the nonlinear formulation of the condition clearly manifests its advantage.

From now on only the regularized conditions based on two asymptotic terms (63) will be utilized. Additionally, for variable U, two different forms, linear (65) and nonlinear (66), will be adopted. For w and Ω two formulations, (59) and (60), which are equivalent to the condition (58) could not compete with their more accurate analogue (63) in terms of solution accuracy and will not be considered.

Graphs presented in Fig. 2 illustrate some peculiarities of the computational process. Here, the maximal relative errors of the solutions (over the time and space) as functions of the number of mesh points N ($\delta f = \delta f(N)$), are presented for different variables. The considered benchmark assumes $Q_l/q_0 = 0.9$ (see "Appendix C"). Three different values of the regularization parameter $\varepsilon = 10^{-3}$, 10^{-4} and 10^{-5} were chosen.

In Fig. 2 two basic tendencies can be observed. The first one is the monotonous error decrease with growing N, up to some stabilization level. This level is different for different dependent variables and values of ε, and in some cases is reached for $N > 1{,}000$ (and thus

[5] Here and everywhere later, by δf we understand the maximal value of the relative error of the function f over all discretized independent variables ($\delta f \equiv \|\delta f\|_\infty$).

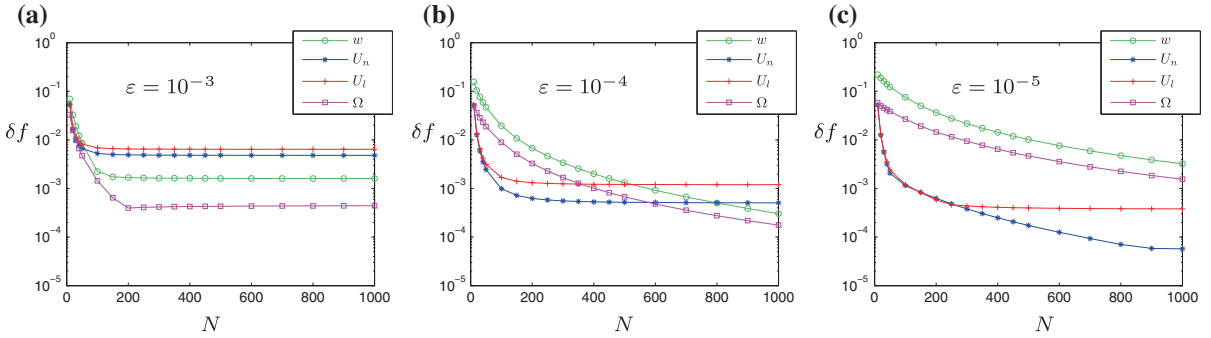

Fig. 2 Maximal relative errors of the solutions computed in different variables δw, δU and $\delta\Omega$, for various number of the grid points N in case of the nonuniform mesh $x^{(II)}(\delta = 2)$. Different values of ε have been considered. All computations were performed for the benchmark $q_l^{(1)}$ for $Q_l/q_0 = 0.9$. Solutions U_l and U_n obtained by unitization of the linear and nonlinear regularized conditions (65) and (66), respectively

cannot be identified in the figure). The second trend is discernible when comparing results for different values of ε. Namely, it turns out that for each dependent variable there exists an optimal ε minimizing the solution error. This value however depends on N. It is not a surprise that the optimal stiffness properties and the maximal solution accuracy are not achieved for the same values of the regularization parameter ε. To increase computational accuracy one needs to decrease ε and increase number of the nodal points N. However, both of these leads to increase of the condition ratio.

Note that the relative errors of respective dependent variables cannot be compared directly. Indeed, even if the errors for w and U are interrelated via the evident relationship $\delta U = 3\delta w$, their comparison with $\delta\Omega$ necessitates an additional postprocessing of the latter. This process, in turn, may introduce its own error. On the other hand, there exists a common component of the solutions, the crack length δL, which can be naturally used for such comparison.

Below we adopt the following strategy for performance test for different dynamic systems. First, we set the number of nodal points, N, to 100. Next, for each of the dependent variables we accept optimal (for $N = 100$) values of the regularization parameter ε. It turned out that the optimal ε differs slightly depending on the type of mesh chosen and the benchmark variant. The general trend for ε can be identified for different meshes (for $x^{(I)}$ it is always smaller than for $x^{(II)}$). However, the sensitivity to the benchmark type is low. The results of computations described by various accuracy measures are collected in Tables 3, 4, 5 and 6 (the optimal values of ε are specified in the captions). We present there: the relative error of solution δf, the absolute error of solution Δf and the relative error of the crack length δL. The following conclusions can be drawn from this data:

(i) Similarly as in case of the stiffness properties, the solution accuracy is affected more by the value of Q_l/q_0 than by the leak-off function behaviour near the crack tip. There is a trend of simultaneous increase of the ratio Q_l/q_0 and the relative errors of dependent variables δf. However this tendency is not in place (or may be even reversed) when analyzing δL.

(ii) In case of the dependent variable U, the way in which the regularized boundary condition is introduced (linear or non-linear) does not play an essential role for the benchmarks and ranges of the parameters under consideration in the Tables 3, 4, 5 and 6. However, there are exceptions to this rule. One of them can be seen in Fig. 2c) for large values of N, where the non-linear condition proves its superiority. Another case will be presented in the end of this section.

(iii) When comparing the common accuracy parameter δL, the dynamic system for Ω gives the best results. The dynamic system for w is the worst performing scheme and comparable to the one for U only in a few cases.

(iv) Since Ω vanishes near the crack tip faster than other variables, one could expect the worst relative error in this case. Surprisingly, even when contrasting the relative (incomparable) errors of the respective dependent variables with each other, the system for Ω seems to be the best choice. The advantage of

Table 3 Performance of the solver based on the dependent variable w for number of nodal points $N = 100$ and various benchmarks. Values of the regularized parameter are $\varepsilon = 5 \cdot 10^{-3}$ and $\varepsilon = 10^{-3}$ for the meshes for $x^{(I)}$ and $x^{(II)}$, respectively

		Dynamic system built on the variable w					
		$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
		$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
δw	$x^{(I)}$	8.5e−3	5.4e−3	5.6e−3	5.2e−3	4.0e−3	3.5e−3
	$x^{(II)}$	2.2e−3	2.6e−3	2.9e−3	1.8e−3	1.9e−3	2.0e−3
Δw	$x^{(I)}$	7.4e−3	9.1e−3	8.8e−3	4.3e−3	4.6e−3	4.7e−3
	$x^{(II)}$	2.8e−3	3.0e−3	3.2e−3	2.1e−3	2.1e−3	2.2e−3
δL	$x^{(I)}$	1.2e−3	1.3e−3	1.1e−3	5.2e−3	5.3e−3	5.2e−3
	$x^{(II)}$	4.0e−4	3.8e−4	3.1e−4	1.8e−3	1.8e−3	1.8e−3

Table 4 Numerical results for the system built on the dependent variable U with the linear regularized condition (65) for $N = 100$ and different ε for the uniform and nonuniform meshes ($\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-5}$, respectively)

		System built on U_l and condition (65)					
		$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
		$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
δU	$x^{(I)}$	1.4e−2	1.0e−2	1.2e−4	2.0e−3	1.4e−3	1.1e−5
	$x^{(II)}$	1.2e−3	6.0e−4	2.5e−4	2.2e−4	1.7e−4	8.6e−5
ΔU	$x^{(I)}$	7.1e−2	4.4e−2	2.0e−3	6.6e−3	4.5e−3	3.9e−4
	$x^{(II)}$	3.1e−2	2.9e−2	7.9e−3	3.5e−3	3.2e−3	7.9e−4
δL	$x^{(I)}$	4.4e−4	2.8e−4	4.4e−6	4.3e−4	2.9e−4	5.6e−6
	$x^{(II)}$	2.6e−4	2.4e−4	1.2e−4	9.5e−5	8.3e−5	4.3e−5

Table 5 Results for the solver based on the dependent variable U for nonlinear regularized condition for $N = 100$ and various benchmarks. Values of the regularized parameter are $\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-5}$ for the meshes for $x^{(I)}$ and $x^{(II)}$, respectively

		System built on U_n and condition (66)					
		$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
		$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
δU	$x^{(I)}$	1.2e−2	9.2e−3	4.3e−5	1.9e−3	1.4e−3	1.3e−5
	$x^{(II)}$	1.2e−3	6.0e−4	2.5e−4	2.0e−4	1.7e−4	8.6e−5
ΔU	$x^{(I)}$	6.4e−2	4.1e−2	1.7e−3	6.5e−3	4.4e−3	4.0e−4
	$x^{(II)}$	3.1e−2	2.9e−2	7.9e−3	3.5e−3	3.2e−3	7.9e−4
δL	$x^{(I)}$	4.1e−4	2.7e−4	1.5e−6	4.2e−4	2.9e−4	6.3e−6
	$x^{(II)}$	2.6e−4	2.4e−4	1.2e−4	9.5e−5	8.3e−5	4.3e−5

Ω over w and U is especially pronounced for the benchmarks variants with a higher ratio Q_l/q_0.

(v) Better solution accuracy is obtained for the nonuniform mesh in almost every case.

We have not observed any significant difference between the time step strategies chosen by the ode15s solver for the different dynamic systems. The number of steps and the main trends were similar. For

Table 6 Computation accuracy for the solver based on the dependent variable Ω for $N = 100$ and $\varepsilon = 10^{-2}$ for $x^{(I)}$ and $\varepsilon = 5 \cdot 10^{-3}$ for $x^{(II)}$

		Dynamic system built on variable Ω					
		$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
		$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
$\delta\Omega$	$x^{(I)}$	2.5e−3	8.7e−4	3.0e−4	5.6e−4	3.3e−4	4.2e−4
	$x^{(II)}$	2.0e−3	7.3e−4	3.6e−4	4.4e−4	2.7e−4	3.1e−4
$\Delta\Omega$	$x^{(I)}$	9.4e−5	1.7e−5	5.9e−5	1.1e−4	1.3e−4	1.5e−4
	$x^{(II)}$	1.9e−4	2.1e−4	2.3e−4	1.3e−4	1.4e−4	1.4e−4
δL	$x^{(I)}$	2.7e−6	5.0e−7	1.7e−6	2.1e−5	2.5e−5	2.9e−5
	$x^{(II)}$	5.5e−6	6.2e−6	6.7e−6	2.6e−5	2.7e−5	2.8e−5

this reason we have not presented any details in the tables.

To visualize the results reported in the Tables 3, 4, 5 and 6, and to complement those presented in Fig. 2, in Fig. 3 we show the relative errors of the crack length δL computed by different dynamic systems (built on different variables). The same benchmark and the values of all other problem parameters as previously discussed in Fig. 2 were considered. If the trends for the different relative errors of the solution δf and the crack length δL in case of $\varepsilon = 10^{-3}$ look similar, the results for the smaller value of the parameter are rather surprising. Indeed, the relative errors δU_l and δU_n are smaller than δw and $\delta\Omega$, while the error δL computed for U no longer follows this trend.

This paradox needs an explanation. A trivial one could be that the maximal error for the solution (δf) is not situated near the crack tip but inside the computational domain. To verify this hypothesis and to give a prospective reader a clear picture of the distribution of the solution error in time and space, we present, in Figs. 4 and 5, the corresponding absolute and relative errors computed for the nonuniform mesh built on $N = 100$ nodal points with the corresponding optimal regularized parameters discussed after Fig. 2. As it follows from Fig. 5, the maximum of the relative error is always achieved near the crack tip ($\max_t \delta f(t, 1-\varepsilon)$). Hence, the initial guess has not been confirmed. On the other hand, the values of the relative errors δf and the respective $\delta L^{(f)}$ are directly interrelated. The following analysis identifies this relationship.

Using (62), after some algebra, one has the estimate:

$$\delta f \approx \delta e_1^{(f)} + \left(\delta e_2^{(f)} - \delta e_1^{(f)}\right) \frac{e_2^{(f)}}{e_1^{(f)}} \varepsilon^{\alpha_2 - \alpha_1}. \tag{70}$$

For the benchmark $q_l^{(1)}$ and $Q_l/q_0 = 0.9$ (see "Appendix C") which always provides the worst accuracy in our computations, one can conclude

$$\delta w \approx \frac{1}{3}\delta U \approx \delta w_0 + \frac{1}{10}(\delta w_1 - \delta w_0)\sqrt[6]{\varepsilon}, \tag{71}$$

$$\delta\Omega \approx \delta w_0 + \frac{8}{90}(\delta w_1 - \delta w_0)\sqrt[6]{\varepsilon}. \tag{72}$$

Finally, from (31) we can derive

$$\delta L \approx \frac{3}{2}\delta w_0. \tag{73}$$

The last relationship has also been verified numerically by evaluating the values of the constant w_0 in the postprocessing procedure using the computed solution (w, U or Ω) and the corresponding regularized boundary condition (compare (62) and (63)).

It is clear from relations (71)–(72) that the relative errors of the respective dependent variables also depend on the quality of approximation of the second term in the regularized boundary condition (62). This explains the surprising relationship between δL and the respective δf.

Interestingly, the results presented in Fig. 3 show that the value of ε which provides the lowest relative error, δf, of the dependent variable f does not necessarily give the best accuracy of the crack length δL. Moreover, the relation $\varepsilon = \varepsilon_L(N)$ is much more sensitive to the variation of N than $\varepsilon = \varepsilon_f(N)$. Indeed, one can observe sharp minima (see Fig. 3a, b) while there is no such phenomenon in the respective graphs for δf (see Fig. 2). To demonstrate that the peaks are not computational artifacts, we also include a small zoom of the corresponding area of the figure Fig. 3b.

To complete the accuracy analysis, let us consider some critical regime of crack propagation. Namely,

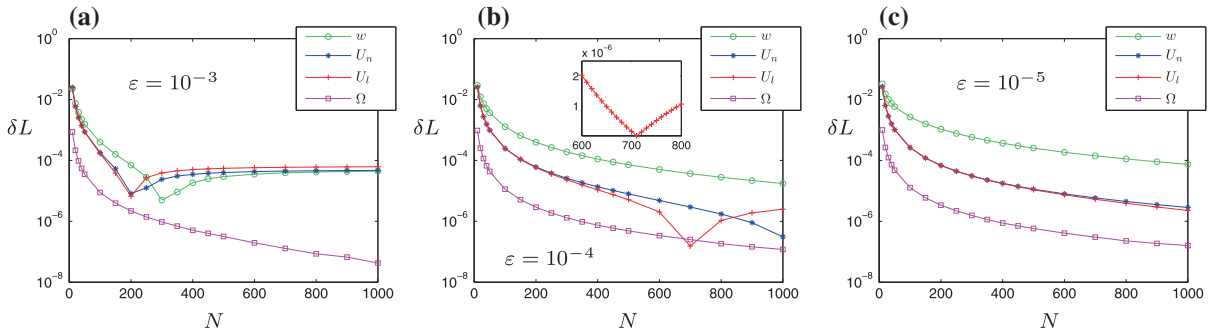

Fig. 3 Distribution of relative errors of the fracture length δL computed by solvers based on different dependent variables (w, U and Ω). When dependent variable U is considered, two different regularised boundary conditions are in use: (65) for U_l and (66) for U_n. Other parameters are the same as in Fig 2. Zoom picture within the Fig. 3b corresponds to the sharp minimum of δL for the variable U_l

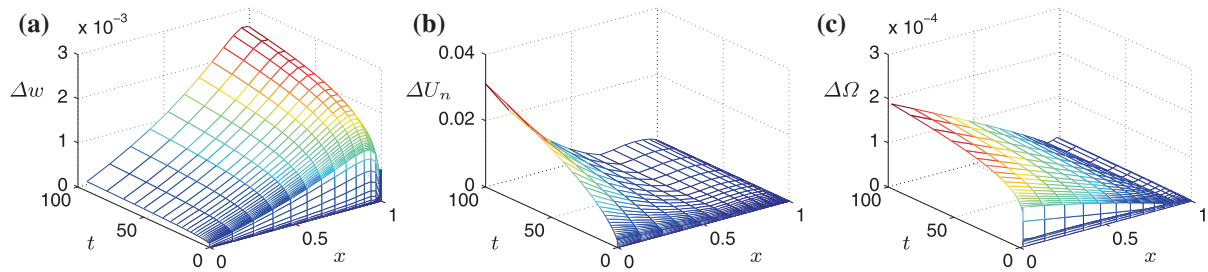

Fig. 4 Absolute error for solutions w, U_n and Ω computed for benchmark $q_l^{(1)}$ with ratio $Q_l/q_0 = 0.9$ and nonuniform mesh $x^{(II)}(\delta = 2)$ with $N = 100$ nodal points. Other parameters: $\varepsilon = 10^{-3}$ for w, $\varepsilon = 5 \cdot 10^{-3}$ for Ω, and $\varepsilon = 10^{-5}$ for U_n

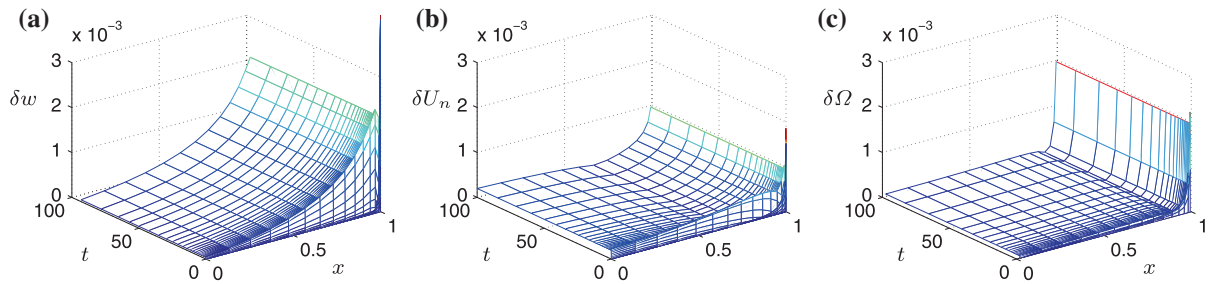

Fig. 5 Relative error of the solutions w, U_n and Ω computed on the corresponding solvers for the same parameters as in Fig. 4

assume that the leak-off flux almost entirely balances the volume of fluid injected into the crack. Indeed, when taking the Carter type benchmark (91) $b_1 = b_2 = 1$, one obtains the fluid balance ratio $Q_l/q_0 = 0.9857$. This gives a very strong variation of the particle velocity function along the crack length [$\gamma_v = 2.07$—see (95)].

In view of the previous conclusion on the influence of the ratio Q_l/q_0 on the solution accuracy (which in fact confirms the observations from Mishuris et al. 2012), one can predict that the solution error will increase appreciably in comparison with the figures shown in Tables 3, 4, 5 and 6. In order to verify this assertion the computations were made for respective dynamic systems (the system for U was analyzed again for two forms of the regularized boundary condition). Both types of meshes, the uniform and the nonuniform, were utilized, each composed of 100 nodal

Table 7 Accuracy parameters for the limiting (critical) variant of the benchmark solution ($Q_l/q_0 = 0.9857$, $\gamma_v = 2.07$) computed for different meshes composed of $N = 100$ nodal points. The blank positions in the table correspond to the case when the solver ode15s could not complete the computations in a reasonable time

	$\epsilon = 10^{-2}$		$\epsilon = 10^{-3}$		$\epsilon = 10^{-4}$		$\epsilon = 10^{-5}$	
	$x^{(I)}$	$x^{(II)}$	$x^{(I)}$	$x^{(II)}$	$x^{(I)}$	$x^{(II)}$	$x^{(I)}$	$x^{(II)}$
δw	1.6e−2	1.6e−2	3.4e−2	2.5e−3	6.8e−2	2.1e−2	–	8.2e−2
δU_l	1.9e−1	1.9e−1	8.2e−2	8.1e−2	1.0e−1	4.3e−2	1.5e−1	2.3e−2
δU_n	4.8e−2	4.8e−2	1.1e−2	6.3e−3	1.9e−2	3.2e−3	2.0e−2	9.1e−3
$\delta\Omega$	2.6e−3	5.0e−3	3.0e−3	1.7e−3	3.9e−3	9.9e−3	4.0e−3	3.0e−2
δL_w	2.0e−4	2.6e−4	2.0e−3	1.8e−4	3.4e−3	3.9e−4	–	6.2e−4
δL_l	1.2e−3	1.1e−3	2.7e−4	1.3e−4	7.5e−4	2.8e−4	1.0e−3	3.2e−4
δL_n	2.5e−4	1.2e−4	1.8e−4	2.6e−4	2.5e−4	3.0e−4	2.6e−4	3.2e−4
δL_Ω	8.1e−7	4.7e−7	1.1e−6	1.1e−6	1.2e−6	1.2e−6	1.2e−6	1.2e−6

points ($N = 100$). Four different values of the regularization parameter ε , ranging from 10^{-5} to 10^{-2}, were analyzed. The results of the computations described by respective accuracy parameters are presented in Table 7. Here, the symbols δU_l and δU_n stand for the relative error of U obtained for the conditions (65) and (66), respectively. The subscript of δL informs us which dynamic system the corresponding result was obtained for.

The data in the table shows that the solution error increased at least one order of magnitude, as compared to the values from Tables 3 , 4, 5 and 6. The lowest deterioration of the solution accuracy was obtained for Ω, which proves the best overall performance of the system built for this variable. Especially impressive is its advantage when comparing the errors of crack length estimation. In all considered cases δL_Ω is at least two orders of magnitude lower than δL for other dependent variables.

In this critical variant of benchmark solution, the non-linear regularized boundary condition (66) for U gives, in most cases, much better performance than its linear counterpart (65) (compare with the discussion after the Tables 2 and 3, 4, 5 and 6). Finally, the non-uniform mesh seems to be a better choice from the point of view of accuracy.

In the last test in this subsection we discuss the sensitivity of respective algorithms to the variation of the crack propagation regime. To this end, consider again the benchmark solution (90) for the critical value of the ratio $Q_l/q_0 = 0.9857$ ($\gamma_v = 2.07$). Now, we analyze a range of parameters $\gamma > -1/3$, motivated by the physical sense of the solution. By changing this value, one simulates different modes of crack propagation (see "Appendix B"). The non-uniform mesh composed of 100 nodes nodes was utilized. For each of the dependent variables an optimal value of the regularization parameter, ε, was taken: $\varepsilon = 10^{-3}$ for w, $\varepsilon = 10^{-5}$ for U, and $\varepsilon = 5 \cdot 10^{-3}$ for Ω. The results of the computations illustrated by the relative errors of the crack length and the maximal relative errors of corresponding dependent variables are shown in Figs. 6 and 7, respectively.

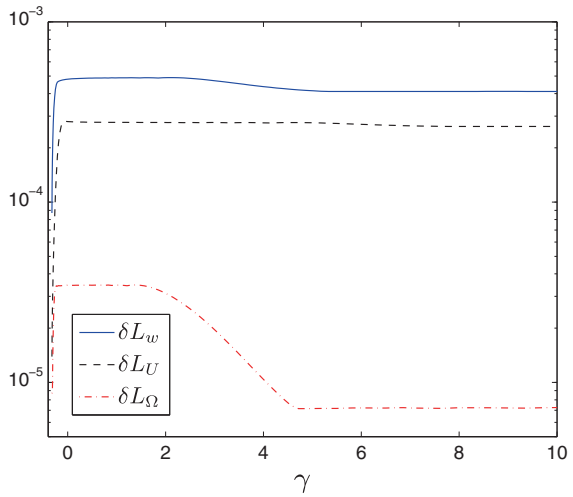

Fig. 6 The relative errors of the crack length for different dependent variables as functions of γ

As can be seen in Fig. 6, for all dependent variables the crack length error rapidly decreases for $\gamma \to -1/3$.

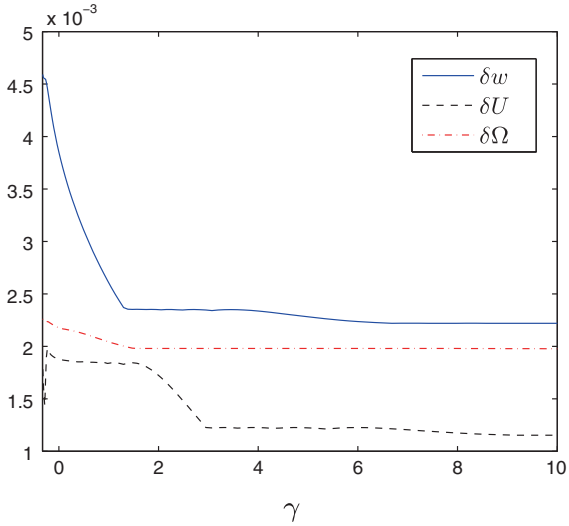

Fig. 7 The maximal relative errors of respective dependent variables as functions of γ

Indeed, this is the case when $L(t) \sim L_0$. For w and U solvers, δL remains very stable over most of the analyzed interval. The solver based on Ω exhibits quite different behaviour. For γ greater than approximately 1.4, the error decreases to achieve the level of its ultimate accuracy, the same as for $\gamma \to -1/3$. Depending on the crack propagation regime, this solver can produce up to two orders of magnitude better accuracy of $L(t)$ than others.

Figure 7 shows, that respective dependent variables themselves are much less sensitive to the changes of γ that the crack length. In the considered interval each solver provides a relatively stable level of accuracy (within the same order of magnitude).

This test proves that using the solver based on Ω is especially beneficial when dealing with the problems of fast propagating fractures (large values of γ).

3.5 Comparison with known numerical results

Although the benchmark solutions utilized in the previous subsections incorporate among others the leak-off term with a square root singularity, there is no analytical solution for the Carter leak-off. In Kovalyshen and Detournay (2009), one can find the numerical results for such a case. This data may be utilized as a reference solution. Unfortunately, the authors provide only some rough estimation of the solution error. Surprisingly, they do not even verify their numerical scheme against the early time asymptotic model (considered as an analytical benchmark) to establish quantitatively the accuracy of computations for the zero leak-off case.

The numerical method used in Kovalyshen and Detournay (2009) is based on an implicit FV algorithm. The data collected in their Table 1 (p. 332) describes the normalized values of the crack length, the crack propagation speed and the crack opening at $x = 0$, at a number of times steps in the interval $t \in [10^{-5}, 5 \cdot 10^2]$. There is no precise information on the utilized number of control volumes and the time stepping strategies (the mentioned number of 10 control volumes refers to the presented graphs, but it is not clear if the data from table was obtained for the same parameters).

In the following we compare our numerical solution (see Table 8) with that by Kovalyshen and Detournay (2009). Note that, due to different normalizations, our normalized crack length, L, is two times greater than respective value in their paper. Our data was obtained by the solver based on U variable for $N = 1{,}000$ nodal points. Although from the previous analysis it emerges that the system for Ω can provide better accuracy of L, we do not use it here to avoid an additional post-processing (numerical differentiation) when computing $w(0, t)$. On the other hand, in the light of previous investigations, the system for U for $N = 1{,}000$ can give the accuracy of L up to 10^{-6}.

First, we present the graphs for evolution of the crack length, $L(t)$,—Fig. 8, and the crack aperture at zero point, $w(t, 0)$—Fig. 9. They depict the data for early time and large time asymptotic models (respective formulae can be found also in Kovalyshen and Detournay (2009)), and the numerical results for a transient regime connecting these asymptotes. The solution by Kovalyshen and Detournay (2009) is indicated by markers. A figure, equivalent to Fig. 8, has been also published in Nordgren (1972), however there is no data available for comparison.

We analyze the time interval $t \in [10^{-8}, 10^8]$ where the initial conditions correspond to the early time asymptote for $t = 10^{-8}$. The same initial time was taken by Kovalyshen and Detournay (2009), but the authors presented their data starting from $t = 10^{-5}$. In order to increase the legibility of the graphs, we have truncated the time axis to the range $t \in [10^{-6}, 5 \cdot 10^3]$, while the complete data is presented in Table 8.

Table 8 Numerical solution of the PKN fracture

$\log(t)$	$L(t)$	$w(t,0)$	$u(t) \times 10$
−7	3.283747e−6	3.988347e−2	7.9701
−6	2.049209e−5	6.298786e−2	7.9355
−5	1.265786e−4	9.915967e−2	7.8716
−4	7.660018e−4	1.551088e−1	7.7536
−3	4.456291e−3	2.397462e−1	7.5185
−2	2.412817e−2	3.629593e−1	7.1173
−1	1.163591e−1	5.326638e−1	6.5267
0	4.849863e−1	7.541837e−1	5.8885
1	1.779508e0	1.037495e0	5.4408
2	6.035529e0	1.403522e0	5.1993
3	1.968511e1	1.883411e0	5.0847
4	6.308563e1	2.518338e0	5.0378
5	2.006370e2	3.362113e0	5.0146
6	6.360179e2	4.485636e0	5.0071
7	2.013373e3	5.982935e0	5.0029

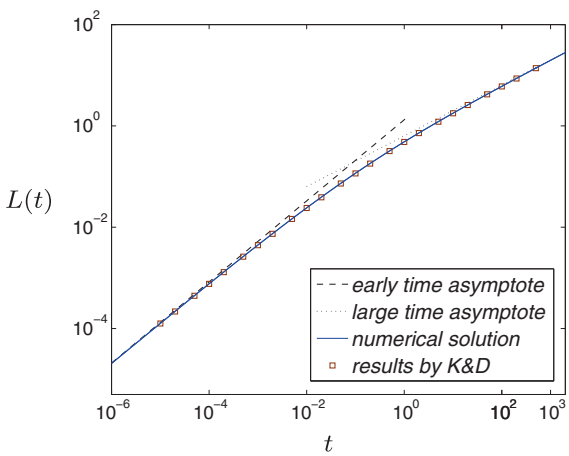

Fig. 8 The crack length evolution in time

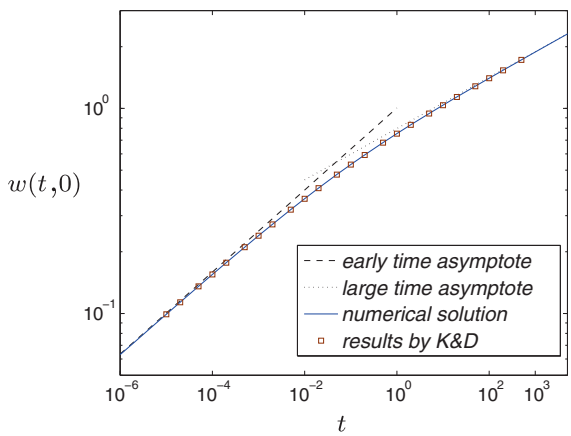

Fig. 9 The evolution of crack opening at zero point, $w(t, 0)$

In Fig. 10 we show the normalized crack propagation speed, defined in a manner introduced by Kovalyshen and Detournay (2009).

As can be seen in the Figs. 8 and 9, in the presented scale, our solution is undistinguishable from that by Kovalyshen and Detournay (2009) in terms of $L(t)$ and $w(t, 0)$. However, the normalized crack propagation speeds differ appreciably from each other. It shows that our solution fits the asymptotes very well, which suggests its good quality. We cannot examine, how the solution of Kovalyshen and Detournay (2009) approaches the asymptotic values due to the shortage of data for time intervals $t < 10^{-5}$ and $t > 5 \cdot 10^3$ in their table.

In the analyzed case, the value of the parameter $Q_l(t)/q_0(t)$ changes continuously with time from zero to unit. From the data presented in Mishuris et al. (2012) and in this paper, one can conclude that for $N = 100$ nodal points, the relative error of the crack length changes from 10^{-6} to 10^{-4} with the increase of the parameter Q_l/q_0. On the other hand, analyzing the data from Figs. 2 and 3 ($Q_l/q_0 = 0.9$), one can expect the achievable level of accuracy of the order

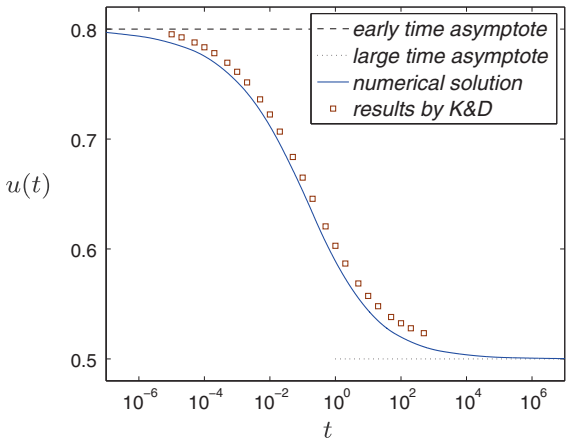

Fig. 10 The evolution of the normalized crack propagation speed

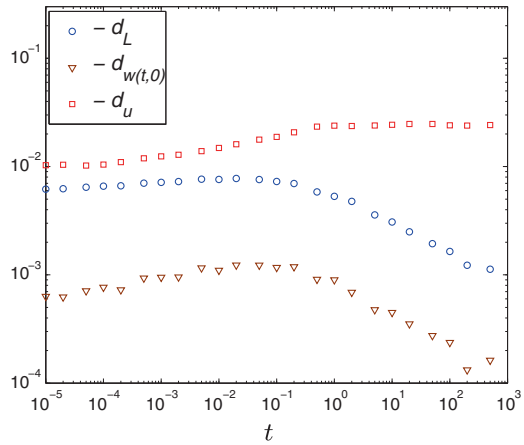

Fig. 12 Relative deviations of the solution by Kovalyshen and Detournay (2009) from the results reported in Table 8

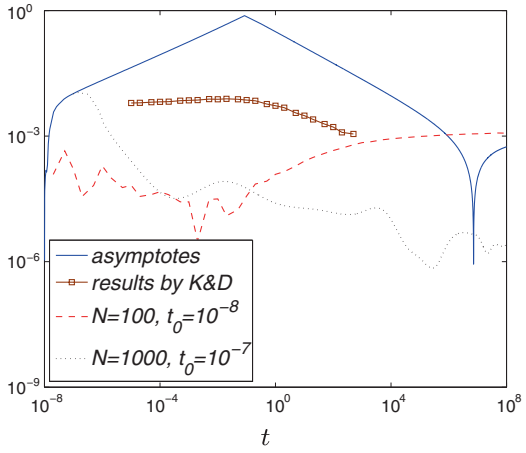

Fig. 11 Relative deviations from the numerical solution for L

10^{-7} for $N = 1{,}000$. This suggests that, in our computations, the relative error of L varies between 10^{-4} and 10^{-6}.

In order to additionally asses the credibility of our solution (computed for $N = 1{,}000$ and presented in the Table 8), we show in Fig. 11 the relative deviations between it and other solutions. Namely, we analyze the crack lengths L provided by: (a) early and large time asymptotes; (b) the solution by Kovalyshen and Detournay (2009); (c) the solution obtained for 100 nodal points and (d) the solution obtained for 1,000 nodal points at another starting point $t_0 = 10^{-7}$.

When tracing the data from Fig. 11 we can see that the deviations of L from the early and large time asymptotes at the ends of the considered interval are of the order 10^{-4}. Moreover, the relative deviation of the solution obtained for 100 points is of the same order in almost entire time range, which corresponds very well to the figures from Table 7. The discrepancy between the reference solution and the solution for $t_0 = 10^{-7}$ decreases rapidly with time. All these observations confirm the credibility of the reference solution.

We do not present respective graph for $w(t, 0)$. However, it is worth mentioning that in this case the deviations from the asymptotes were even lower than for L, while the deviation of the solution for $N = 100$ did not exceed the value of 10^{-4} on the substantial part of the interval.

The relative discrepancies between the components of our solution and the solution by Kovalyshen and Detournay (2009) are shown in Fig. 12. Here d_L, $d_{w(0,t)}$ and d_u refer to the deviations of the crack length, L, the crack opening, $w(t, 0)$ and the normalized crack velocity, u, respectively.

In the light of the presented results, we believe that the data collected in Table 8 provide the accuracy **at least** of the order 10^{-4} for both, the crack length, L, and the crack opening at $x = 0$. Moreover, in the considerable time range ($10^{-6} < t < 10^{6}$), one can expect the error lower by up to two orders of magnitude. The normalized crack propagation speed u is computed with accuracy of one to two orders of magnitude lower.

Summarizing the above discussions, the level of accuracy for the results tabulated by Kovalyshen and

Detournay (2009) can be estimated as $10^{-3} \div 10^{-2}$ for L, $10^{-4} \div 10^{-3}$ for $w(t, 0)$ and of the order 10^{-2} for u.

3.6 Remarks on the sensitivity of the Carter leak-off model

It is well known that applicability of the empirical Carter law $(4)_1$ in the vicinity of the fracture tip is questionable (see for example, Economides and Nolte 2000; Kovalyshen 2010; Mitchell et al. 2007). Moreover, when combining Carter's leak-off with some nonlocal variants of elasticity models (for example, KGD model of hydrofracturing), one obtains an infinite particle velocity at the crack tip. As a result, the speed Eq. (12) cannot be applied in such a case. One of the ways to eliminate the negative consequences of this fact is to assume that the Carter law becomes valid at some distance away from the fracture tip (see for example Mitchell et al. 2007).

The PKN model, which does not exhibit such a drawback, gives however a unique opportunity to assess how the solution is affected by a modification of the classic Carter law in the neighborhood of the fracture tip.

To this end, let us consider two ways of modification of the law. The first one assumes that leak-off function equals zero over some distance from the crack front ($d > \varepsilon$). The second one accepts a constant value of q_l in the same interval. This value is taken in such a manner to preserve the continuity of the leak-off function. Note, that both of these modifications change the volume of fluid loss to the rock formation, with respect to the original state.

The relative deviations of the crack lengths for these modifications from the original one are shown in Fig. 13. Results for two values of d: $d = \varepsilon$, $d = 10\varepsilon$ (for $\varepsilon = 10^{-5}$) are depicted. The symbol q_{ld} in the legend refers to the cases when the leak-off function is complimented by the constant value over $1 - d \le x \le 1$.

One can see from this picture that the maximal relative discrepancies (of the level of 1 %) appear at the initial and large time ranges. To explain this phenomenon we can easily compute the additional volume of fluid retained in the fracture as a result of the Carter law amendments. Taking into account (74), these values are $\Delta Q_l(t) = 2D(t)\sqrt{d}$ and $\Delta Q_l(t) = D(t)\sqrt{d}$, for the respective modifications [see (22) for $D(t)$]. Note that $D(t) = \sqrt{u(t)/t}$, which explains the level of deviation

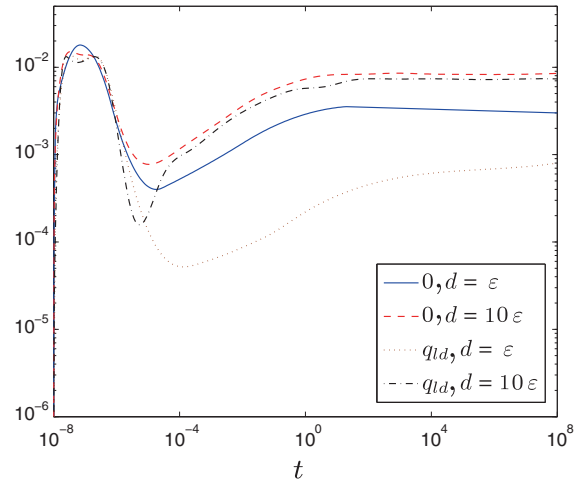

Fig. 13 Relative deviations of the crack lengths for different variants of truncated Carter law

for the small time. For large time, the effect of accumulation of the difference of the fluid loss, $\int_0^t \Delta Q_l(\tau)d\tau = O(\sqrt{td})$, plays a crucial role.

The above test proves that the application of the Carter law modified in the aforementioned ways, is acceptable in terms of both the stability and the accuracy of computations. This allows one to use such an approach in the cases when the law leads to unphysical results (see Mitchel et al. 2007). Note, however, that the intermediate asymptotics related to the carter law still holds true and thus should be taken into account.

On the other hand, appreciable sensitivity of the solution to the slight modifications of the Carter law calls the validity of the law near the crack tip into question.

4 Conclusions

In this paper, we have revisited the PKN model of hydrofracturing providing a comprehensive overview of the known results together with:

(i) analysis of various leak-off regimes (vanishing, bounded and singular near the fracture tip) supplemented with the full asymptotic expansion of the solution for the Carter model;
(ii) introduction of a new dependent variable, Ω, and the resulting problem reformulation;
(iii) implementation of a new form of the regularized boundary condition;

(iv) analysis of different aspects of application of various dependent and independent variables, including the stiffness properties, accuracy and efficiency of the computations;
(v) comparison of the results for the Carter leak-off with the known numerical benchmark from Kovalyshen and Detournay (2009).

The main conclusions of the paper are:

- The approach proposed in Linkov (2011d) and Mishuris et al. (2012) is efficient even in the cases when the smoothness of the particle velocity near the crack tip is disturbed by the singular leak-off function;
- For the best performance of a solver, the regularized boundary condition should incorporate at least two leading terms of asymptotics;
- The stiffness properties of the dynamic systems and the accuracy of the computations can be effectively controlled by the choice of dependent and independent variables.
- The new independent variable, Ω, leading to a slightly greater stiffness of the dynamic system, considerably improves the accuracy of computations.

The value of the aforementioned findings was demonstrated on various analytical benchmarks and for the classic Carter law. Although, the presented analysis concerns the 1-D PKN formulation, at least some of the findings may be utilised in the more advanced cases.

Acknowledgments The authors are grateful to the reviewers for their numerous comments and remarks allowing significant enhancement of the manuscript. The paper was completed during P.K. and G.M. secondments to the industrial partner EUROTECH. The facilities and hospitality of the Partner is gratefully acknowledged.

Appendices

A Carter's leak-off function in the normalised formulation

Consider the transformation of the Carter law described by (4) when applying the normalization (18). Assume that:

$$\frac{1}{\sqrt{t-\tau(x)}} = \frac{D(t)}{\sqrt{1-\tilde{x}}} + R(t,\tilde{x}), \tag{74}$$

where function $D(t)$ is defined in (22) while the remainder R is estimated later in (77).

To find function $D(t)$, and thus to obtain an exact form of Eq. (22), it is enough to compute the limit

$$D^2(t) = \lim_{\tilde{x}\to 1} \frac{1-\tilde{x}}{t-\tau(x)}. \tag{75}$$

This can be done by utilising L'Hopital's rule with taking into account that $x \to L(t)$ as $\tilde{x} \to 1$,

$$\tau(x) = \tau\left(L(t)\tilde{x}\right) = L^{-1}(L(t)\tilde{x}), \tag{76}$$

and that the crack length is a smooth function of time ($L \in C^1$ at least). The last fact immediately follows from the problem formulation in terms of evolution system (32).

Having the value of $D(t)$ we can estimate the remainder $R(t,\tilde{x})$ when $\tilde{x} \to 1$, or, what it is equivalent to when $x \to l(t)$ (or $t \to \tau(x)$). For this reason, we search for a parameter $\xi \neq 0$ which guarantees that the limit

$$\begin{aligned} A &= \lim_{\tilde{x}\to 1} \frac{R(t,\tilde{x})}{(1-\tilde{x})^{\xi}} \\ &= \lim_{\tilde{x}\to 1} \frac{1}{2\xi(1-\tilde{x})^{\xi-1}} \left(\frac{D(t)}{(1-\tilde{x})^{3/2}} - \frac{L(t)\tau'(x)}{(t-\tau(x))^{3/2}} \right) \end{aligned}$$

does not turn to zero or infinity. Due to this assumption, we can write

$$\frac{1}{\sqrt{t-\tau(x)}} = \frac{D(t)}{\sqrt{1-\tilde{x}}} + A(1-\tilde{x})^{\xi} + o\left((1-\tilde{x})^{\xi}\right), \tag{77}$$

when $\tilde{x} \to 1$, or equivalently $x \to l(t)$. Taking the last estimate into account A can be expressed as:

$$\begin{aligned} A = \lim_{\tilde{x}\to 1} \frac{1}{2\xi(1-\tilde{x})^{\xi-1}} \left(\frac{D(t)}{(1-\tilde{x})^{3/2}} - \frac{L(t)\tau'(x)}{t-\tau(x)} \frac{D(t)}{\sqrt{1-\tilde{x}}} \right) \\ - \frac{AL(t)}{2\xi} \lim_{\tilde{x}\to 1} \frac{(1-\tilde{x})\tau'(x)}{t-\tau(x)} \big(1+o(1)\big). \end{aligned}$$

Now, on substitution of $\tau'(x) = 1/L'(t)$ at $x = L(t)$ and (75) into the limit one has:

$$A = \lim_{\tilde{x}\to 1} \frac{D(t)}{2\xi(1-\tilde{x})^{\xi-1/2}} \left(\frac{1}{1-\tilde{x}} - \frac{L(t)\tau'(x)}{t-\tau(x)} \right) - \frac{AL(t)D^2(t)}{2\xi L'(t)}.$$

Applying (75) and (22) here gives:

$$\begin{aligned} \frac{1+2\xi}{2\xi} A = \lim_{\tilde{x}\to 1} \frac{D(t)}{2\xi(1-\tilde{x})^{\xi-1/2}} \left(\frac{1}{1-\tilde{x}} - \frac{L(t)\tau'(x)}{\sqrt{t-\tau(x)}} \frac{D(t)}{\sqrt{1-\tilde{x}}} \right) \\ - \frac{AD(t)L(t)}{2\xi} \lim_{\tilde{x}\to 1} \frac{\tau'(x)\sqrt{1-\tilde{x}}}{\sqrt{t-\tau(x)}}. \end{aligned}$$

By repeating the same process one more time we have:

$$(2+2\xi)A = \lim_{\tilde{x}\to 1} \frac{D(t)}{(1-\tilde{x})^{\xi}} \left(\frac{1}{\sqrt{1-\tilde{x}}} - \frac{L(t)\tau'(x)D(t)}{\sqrt{t-\tau(x)}} \right).$$

Finally by eliminating the square root with use of (77) we obtain (after some algebra)

$$(3+2\xi)A = D(t)\lim_{\tilde{x}\to 1}\frac{1-L(t)\tau'(x)D^2(t)}{(1-\tilde{x})^{\xi+1/2}}.$$

This relationship gives a finite value of A if and only if $\xi = 1/2$ and, as a result, we find:

$$A = \frac{1}{4}D^3(t)L^2(t)\tau''(L(t)) = -\frac{1}{4}\frac{L''(t)}{L'(t)}\sqrt{\frac{L(t)}{L'(t)}}.$$

B Asymptotics of the solutions for different leak-off functions

Asymptotic expansion for the crack opening and the fluid velocity near the crack tip in the normalised variables (18) can be written in the following general forms:

$$w(t,x)=\sum_{j=0}^{N} w_j(t)(1-x)^{\alpha_j}+O((1-x)^{\varrho_w}),\quad x\to 1, \tag{78}$$

and

$$V(t,x)=\sum_{j=0}^{N} V_j(t)(1-x)^{\beta_j}+O((1-x)^{\varrho_V}),\quad x\to 1, \tag{79}$$

with $\varrho_w > \alpha_n$, $\varrho_V > \beta_n$, $\alpha_0 = 1/3$, $\beta_0 = 0$ and some increasing sequences $\alpha_0, \alpha_1, \ldots, \alpha_n$ and $\beta_0, \beta_1, \ldots, \beta_n$. Note that the asymptotics are related to each other by the speed Eq. (19) and thus, regardless of the chosen leak-off function, we can write

$$\sum_{j=0}^{N} V_j(t)(1-x)^{\beta_j}+\ldots = \frac{1}{3L(t)}\sum_{k=0}^{N}\sum_{m=0}^{N}\sum_{j=0}^{N} \times\alpha_j w_j(t)w_m(t)w_k(t)(1-x)^{\alpha_j+\alpha_m+\alpha_k-1}. \tag{80}$$

In line with the discussion after Eq. (16), we are interested only in the terms such that $\beta_j \le 1$, restricting ourselves to the smallest $\varrho_V > 1$, since the values of β_j are combinations of a sum of three consequent components of the exponents α_j. However, since α_0 is known ($\alpha_0 = 1/3$), one can write (compare with (17)):

$$V_0(t) = \frac{1}{3L(t)}w_0^3(t), \tag{81}$$

$$V_1(t)=\frac{1}{L(t)}\left(\alpha_1+\frac{2}{3}\right)w_0^2(t)w_1(t),\quad \beta_1=\alpha_1-\frac{1}{3}. \tag{82}$$

To continue the process one now needs to compute the value of the exponent α_1 as it is not clear a priori which value determining the next exponent $\beta_2 = \min\{2/3+\alpha_2, 1/3+2\alpha_1\}$ is larger. To do so let us rewrite the continuity Eq. (20) in the form:

$$\frac{\partial w}{\partial t}+\frac{V_0(t)}{L(t)}(1-x)\frac{\partial w}{\partial x}=\frac{1}{L(t)}\frac{\partial\big(w(V_0-V)\big)}{\partial x}-q_l(t,x). \tag{83}$$

Here, the terms on the left-hand side of the equation are always bounded near the crack tip, while those on the right-hand side can behave differently depending on the chosen leak-off function.

Consider the following three cases of q_l behaviour.

(i) Assume first that

$$q_l(t,x) = o\big(w(t,x)\big),\quad x\to 1.$$

This case naturally includes the impermeable rock formation. Analysing the leading order terms in the Eq. (83), it is clear that $w(V_0-V) = O((1-x)^{4/3})$, as $x\to 1$. This, in turn, is only possible for $\beta_1 = 1$ and, therefore, $\alpha_1 = 4/3$. Finally, comparing the left-hand side and the right-hand side of the equation we obtain:

$$w_0'(t) = \frac{w_0(t)}{3L(t)}\big(V_0(t)+4V_1(t)\big),$$
$$V_1(t) = \frac{2}{L(t)}w_0^2(t)w_1(t). \tag{84}$$

This case has been considered in Linkov (2011d) and Mishuris et al. (2012).

(ii) If we assume that the leak-off function is estimated by the solution as $O\big(w(t,x)\big)$, or equivalently;

$$q_l(t,x)\sim\Upsilon(t)w_0(t)(1-x)^{1/3},\quad x\to 1,$$

then the previous results related to the values of α_1 and β_1 and, therefore, the Eq. $(84)_2$ remain the same, while the first one changes to

$$w_0'(t) = \frac{1}{3L(t)}w_0(t)\big(V_0(t)+4V_1(t)\big)-\Upsilon(t)w_0(t). \tag{85}$$

This case corresponds to $(21)_3$ when $C_{32} = 0$ and $\Upsilon(t) = kC_{31}(t)$.

(iii) The leak-off function in a general form:

$$q_l(t,x)=\Phi(t)(1-x)^{\theta}+o((1-x)^{1/3}),\quad x\to 1,$$

where $-1/2 \le \theta < 1/3$. Here, one can conclude that $w(V_0-V) = O((1-x)^{1+\theta})$, as $x\to 1$

or equivalently, $\beta_1 = \theta + 2/3$, and $\alpha_1 = 1 + \theta$. Moreover, in this case:

$$(1+\theta)w_0 V_1 = L(t)\Phi(t),$$
$$V_1(t) = \frac{1}{L(t)}\left(\theta + \frac{4}{3}\right) w_0^2(t) w_1(t), \tag{86}$$

and, thus

$$w_1(t) = \frac{3L^2(t)\Phi(t)}{(4+3\theta)(1+\theta)w_0^3(t)}. \tag{87}$$

Note, that as one would expect, the particle velocity function is not smooth in this case near the crack tip, its derivative is unbounded and exhibits the following behaviour:

$$\frac{\partial V}{\partial x} = O\big((1-x)^{\theta-1/3}\big), \quad x \to 1.$$

To formulate the equation similar to $(84)_1$ or (85), one needs to continue asymptotic analysis of the Eq. (83) incorporating the available information. Apart from the fact that the analysis can be done in the general case, we restrict ourselves only to three variants used from the beginning (compare (4)), respectively: $\theta = 0$, $\theta = 1/3 - 1/2 = -1/6$ and $\theta = -1/2$.

When $\theta = 0$, $\alpha_1 = 1$ and $\beta_1 = 2/3$, returning to the Eq. (80), one concludes that $\beta_2 > 1$ and, therefore,

$$w_0'(t) = \frac{1}{3L(t)} w_0(t) V_0(t). \tag{88}$$

This case corresponds to $(21)_3$ when $\Phi(t) = C_3^{(2)}(t)w_0(t)$ and $C_3^{(1)} = 0$.

If $\theta = -1/6$, then $\alpha_1 = 5/6$ and $\beta_1 = 1/2$. In this case the function $\Phi(t)$ can be written as $\Phi(t) = C_2 D(t) w_0(t)$ [compare to $(21)_2$] and again Eq. (80) gives $\beta_2 > 1$, while Eq. (83) leads to

$$w_0'(t) = \frac{1}{3L(t)}\big(w_0(t)V_0(t) + 4w_1(t)V_1(t)\big). \tag{89}$$

Summarizing, in both mentioned above cases, there exists a single term in asymptotics of the particle velocity which has singular derivative near the crack tip. Moreover, those terms (w_1 and V_1, respectively) are fully defined by the leak-off function $\Phi(t)$ and the coefficient w_0 in front of the leading term for the crack opening in (87) and $(86)_1$.

The situation changes dramatically when $\theta = -1/2$ (Carter law). We now have $\alpha_1 = 1/2$ and $\beta_1 = 1/6$ and $\Phi(t) = C_1 D(t)$. In this case, however, $\beta_2 < 1$ and we need to continue the asymptotic analysis further to evaluate all terms of the particle velocity which exhibit non-smooth behaviour near the crack tip. We omit the details of the derivation, presenting only the final result in a compact form. The first six exponents in the asymptotic expansions (78) and (79), that introduce the singularity of w_x, are:

$$\alpha_j = \frac{1}{2} + \frac{j}{6}, \quad \beta_j = \frac{j}{6}, \quad j = 1, 2, \ldots, 6.$$
$$w_j(t) = \kappa_j \frac{\Phi^j(t) L^{2j}(t)}{w_0^{4j-1}(t)}, \quad V_j(t) = \psi_j \frac{\Phi^j(t) L^{2j-1}(t)}{w_0^{4j-3}(t)},$$

where $j = 1, 2, \ldots, 5$ and

$$\kappa_1 = \tfrac{12}{7}, \quad \psi_1 = 2, \quad \kappa_2 = -\tfrac{270}{49}, \quad \psi_2 = -\tfrac{24}{7},$$
$$\kappa_3 = \tfrac{9768}{343}, \quad \psi_3 = \tfrac{828}{49}, \quad \kappa_4 = -\tfrac{2097252}{12005}, \quad \psi_4 = -\tfrac{5136}{49},$$
$$\kappa_5 = \tfrac{1081254096}{924385}, \quad \psi_5 = \tfrac{1234512}{1715}.$$

C Benchmark solutions

There are several benchmarks in the literature to be utilized for investigation of the numerical algorithms. Benchmark solutions for impermeable rock have been constructed in Kemp (1989); Linkov (2011d), while that corresponding to the non-zero leak-off model with q_l vanishing at a crack tip has been analyzed in Mishuris et al. (2012).

In this paper, we introduce three different analytical benchmark solutions corresponding to the representations (21). Moreover, for each of the leak-off functions under consideration we take two different relationships between the injection flux rate q_0 and the leak-off to formation q_l. In this way six different benchmark solutions are analyzed.

In order to formulate the benchmark solutions let us assume the following form of the crack opening function:

$$w(t,x) = W_0(1+t)^\gamma h(x), \quad W_0 = \sqrt[3]{\frac{3}{2}(3\gamma+1)}, \tag{90}$$

where γ is an arbitrary parameter, and the function $h(x)$ $(0 < x < 1)$ is given by:

$$h(x) = (1-x)^{\frac{1}{3}} + b_1(1-x)^{\lambda_1} + b_2(1-x)^{\lambda_2}. \tag{91}$$

The choice of the next powers $1/3 < \lambda_1 < \lambda_2$ will depend on the leak-off variant from (4). On consecutive substitutions of (90)–(91) into the relations (19), (24),

Table 9 The values of parameters b_1 and b_2 for different benchmark solutions modelling desired leak-off to fluid injection ratios

	$Q_l/q_0 = 0.9$			$Q_l/q_0 = 0.5$		
	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$	$q_l^{(1)}$	$q_l^{(2)}$	$q_l^{(3)}$
b_1	0.1	0.19	0.74	0.02	0.03	0.15
b_2	0.5	0.41	−0.13	0.1	0.08	−0.02
γ_v	1.69	1.70	1.65	0.55	0.56	0.55

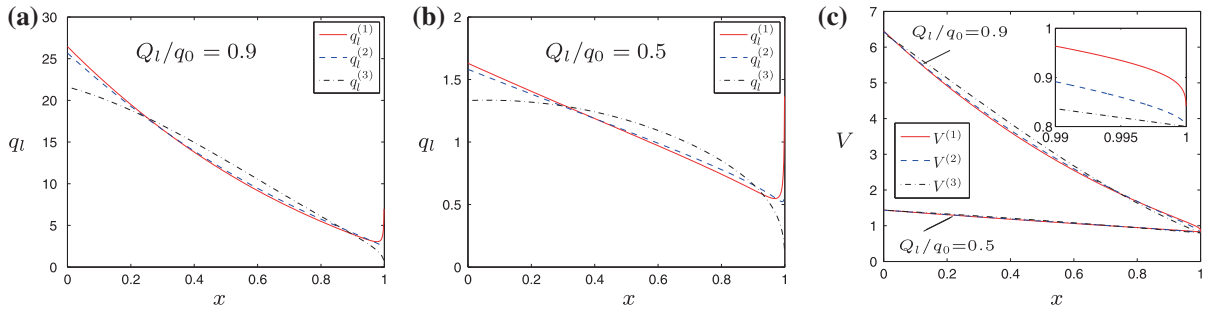

Fig. 14 Distributions of the leak-off functions $q_l(t, x)$ and the respective particle velocity $V(t, x)$ over $\tilde{x} \in (0, 1)$ at initial time $t = 0$

(29) and (31) one obtains the remaining benchmark quantities:

$$L(t) = (1+t)^{\frac{3\gamma+1}{2}},$$

$$V(t,x) = -W_0^3(1+t)^{\frac{3\gamma-1}{2}}h^2(x)\frac{\partial h}{\partial x}. \tag{92}$$

$$q_0(t) = -W_0^4(1+t)^{\frac{5\gamma-1}{2}}\left(h^3\frac{\partial h}{\partial x}\right)|_{x=0}. \tag{93}$$

$$q_l(t,x) = W_0(1+t)^{\gamma-1}\times$$
$$\left(\frac{3}{2}(3\gamma+1)\left[\frac{1}{3}x\frac{\partial h}{\partial x}+3h^2\left(\frac{\partial h}{\partial x}\right)^2+h^3\frac{\partial^2 h}{\partial x^2}\right]-\gamma h\right). \tag{94}$$

It can be easily checked that for $\lambda_1 = 1/2$ and $\lambda_2 = 4/3$ the leak-off function incorporates a square root singular term of type $(21)_1$. By setting $\lambda_1 = 5/6$ and $\lambda_2 = 4/3$ we comply with representation $(21)_2$. Although in both of these cases $q^*_{1(2)}$ exhibits a singular behaviour at the crack tip, it does not detract from the applicability of our benchmarks. Finally, when using $\lambda_1 = 4/3$ and $\lambda_2 = 7/3$, the benchmark gives a non-singular leak-off function in the form $(21)_3$.

Note also, that by manipulating with the value of γ one can simulate some very specific regimes of crack propagation. For example $\gamma = 1/5$ corresponds to the constant injection flux rate, while $\gamma = 1/3$ gives a constant crack propagation speed. For our computations we always set the value of $\gamma = 1/5$.

Choosing appropriate values b_1 and b_2 one can change the relation between the amount of fluid loss to formation and the injection rate. This ratio can be defined by the measure, Q_l/q_0, where Q_l is the total volume of leak-off $\int_0^1 q_l dx$. It is important to note that this measure decreases in time, from its maximum value to zero, for all chosen benchmarks. Thus, taking the maximal value high enough and tracing the solution accuracy in time, one can analyse performance of the algorithm for any possible value of the parameter. We consider two variants of Q_l/q_0, one where fluid injection doubles the size of total fluid loss, and a second where the total fluid loss is close to injection rate. The values of the corresponding parameters b_1, b_2 are presented in Table 9.

Additionally one can compute a parameter γ_v defined in Mishuris et al. (2012) as a measure of the uniformity of fluid velocity distribution:

$$\gamma_v = [\max(V(t,x)) - \min(V(t,x))]\left[\int_0^1 V(t,\xi)d\xi\right]^{-1}. \tag{95}$$

Interestingly, this measure is directly correlated with the leak-off ratio Q_l/q_0.

In Fig. 14 the distributions of the leak-off functions and the corresponding particle velocities for the respective benchmarks are presented. It shows that the veloc-

ity near the crack tip depends strongly on the benchmark variant. To highlight this fact, a zoom picture is placed in the Fig. 14b.

Note that the benchmark $q_l^{(1)}$ is worse, in a sense, than the original Carter's model as it contains additional singular terms of the leak-off function. These terms are absent in the normalised Carter's law as it follows from "Appendix B".

References

Adachi J, Detournay E (2002) Self-similar solution of a plane-strain fracture driven by a power-law fluid. Int J Numer Anal Methods Geomech 26:579–604

Adachi JI, Peirce AP (2007) Asymptotic analysis of an elasticity equation for a finger-like hydraulic fracture. J Elast 90(1):43–69

Adachi J, Siebrits E, Peirce A, Desroches J (2007) Computer Simulation of Hydraulic Fractures. Int J Rock Mech Min Sci 44:739–757

Aiken RC (ed) (1985) Stiff computation. Oxford University Press, Oxford

Carter E (1957) Optimum fluid characteristics for fracture extension. In: Howard G, Fast C (eds) Drilling and production practice. American Petroleum Institute, New York, pp 261–270

Crittendon BC (1959) The mechanics of design and interpertation of hydraulic fracture treatments. J Pet Tech 21:21–29

Clifton RJ, Wang JJ (1988) Multiple fluids, proppant transport, and thermal effects in threedimensional simulation of hydraulic fracturing. SPE 18198

Desroches J, Thiercelin M (1993) Modeling the propagation and closure of micro-hydraulic fracturing. Int J Rock Mech Min Sci 30:1231–1234

Desroches J, Detournay E, Lenoach B, Papanastasiou P, Pearson J, Thiercelin M, Cheng A-D (1994) The crack tip region in hydraulic fracturing. Proc R Soc Lond Ser A 447:39–48

Detournay E (2004) Propagation regimes of fluid-driven fractures in impermeable rocks. Int J Geom 4:1–11

Economides M, Nolte K (eds) (2000) Reservoir Stimulation, 3rd edn. Wiley, Chichester, UK

Garagash D, Detournay E, Adachi J (2011) Multiscale tip asymptotics in hydraulic fracture with leak-off. J Fluid Mech 669:260–297

Geertsma J, de Klerk F (1969) A rapid method of predicting width and extent of hydraulically induced fractures. J Pet Tech 21:1571–1581 [SPE 2458]

Harrison E, Kieschnick WF, McGuire WJ (1954) The mechanics of fracture induction and extension. Petroleum Trans AIME 201:252–263

Hubbert MK, Willis DG (1957) Mechanics of hydraulic fracturing. J Pet Tech 9(6):153–168

Khristianovic SA, Zheltov YP (1955) Formation of vertical fractures by means of highly viscous liquid. In: Proceedings of the fourth world petroleum congress. Rome, pp 579–586

Kemp LF (1989) Study of Nordgren's equation of hydraulic fracturing. SPE Prod Eng 5:311–314

Kovalyshen Y, Detournay E (2009) A reexamination of the classical PKN model of hydraulic fracture. Transp Porous Med 81:317–339

Kovalyshen Y (2010) Fluid-driven fracture in poroelastic medium. Ph.D. thesis, The University of Minnesota.

Lenoach B (1995) The crack tip solution for hydraulic fracturing in rock of arbitrary permeability. J Mech Phys Solids 43:1025–1043

Linkov AM (2011a) Speed equation and its application for solving Ill-posed problems of hydraulic fracturing. ISSM 1028–3358, Doklady Phys 56(8):436–438. Pleiades Publishing, Ltd. 2011

Linkov AM (2011b) Use of a speed equation for numerical simulation of hydraulic fractures. arXiv:1108.6146

Linkov AM (2011c) On numerical simulation of hydraulic fracturing. In: Proceedins of XXXVIII summer school-conference 'Advanced Problems in Mechanics-2011', Repino, St. Petersburg, 1–5 July 2011, pp 291–296

Linkov AM (2011) On efficient simulation of hydraulic fracturing in terms of particle velocity. Int J Eng Sci 52:77–88

Lister JR (1990) Buoyancy-driven fluid fracture: the effects of material toughness and of low-viscosity precursors. J Fluid Mech 210:263–280

Mack MG, Warpinski NR (2000) Mechanics of hydraulic fracturing. In: Ecomides N (ed) Reservoire stimulation, 3rd edn. Wiley, Chicester [Chapter 6]

Mathias SA, van Reeuwijk M (2009) Hydraulic fracture propagation with 3-D leak-off. Transp Porous Media 80: 499–518

Mishuris G, Wrobel M, Linkov A (2012) On modeling hydraulic fracture in proper variables: stiffness, accuracy, sensitivity. Int J Eng Sci 61:10–23

Mitchell SL, Kuske R, Peirce AP (2007) An asymptotic framework for finite hydraulic fractures including leak-off. SIAM J Appl Math 67(2):364–386

Moschovidis ZA, Steiger RP (2000) The Mounds drill-cuttings injection experiment:final results and conclusions. In: Proceedings of the IADC/SPE drilling conference, New Orleans, February 23–25. Society of Petroleum Engineers, Richardson [SPE 59115]

Nordgren RP (1972) Propagation of a vertical hydraulic fracture. J Pet Tech 253:306–314

Perkins TK, Kern LR (1961) Widths of hydraulic fractures. J Pet Tech 13(9):37–49 [SPE 89]

Pine RJ, Cundall PA (1985) Applications of the Fluid-Rock Interaction Program (FRIP) to the modelling of hot dry rock geothermal energy systems. In: Proceedings of the international symposium on fundamentals of rock joints, Bjorkliden, Sweden, September 1985, pp 293–302

Rubin AM (1995) Propagation of magma filled cracks. Ann Rev Earth Planet Sci 23:287–336

Savitski A, Detournay E (2002) Propagation of a fluid-driven penny-shaped fracture in an impermeable rock: asymptotic solutions. Int J Solids Struct 39(26):6311–6337

Sneddon IN, Elliot HA (1946) The opening of a Griffith crack under internal pressure. Q Appl Math 4:262–267

Sneddon IN (1946) The distribution of stress in the neighbourhood of a crack in an elastic solid. Proc R Soc Lond A 187:229–260

Spence DA, Sharp P (1985) Self-similar solutions for elastohydrodynamic cavity flow. Proc R Soc Lond A 400:289–313

Tsai VC, Rice JR (2010) A model for turbulent hydraulic hydraulic fracture and application to crack propagation at glacier beds. J Geophys Res 115:1–18

Wrobel M, Mishuris G (2013) Efficient pseudo-spectral solvers for the PKN model of hydrofracturing. Int J Fract.doi:10.1007/s10704-013-9847-y

DOI 10.1007/978-3-319-04397-5_15
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9878-4

ORIGINAL PAPER

Prediction of grain boundary stress fields and microcrack initiation induced by slip band impingement

Maxime Sauzay · Mohamed Ould Moussa

Received: 22 February 2013 / Accepted: 27 August 2013 / Published online: 27 November 2013

Abstract Slip localization is widely observed in metallic polycrystals undergoing cyclic deformation or post-irradiation tensile deformation, whatever their crystallographic structure. Hence, strong strain localization occurs in thin slip bands (SBs) inducing by the way local stress concentrations at their intersections with grain boundaries (GBs). Many GB stress field formulae based on the dislocation pile-up theory have been proposed since the pionnering work of Stroh and others. These allow the use of the Griffith criterion for prediction GB fracture initiation. However, recent observations show that assuming that slip is localized on a single atomic plane leads to unrealistic results. In fact, a large number of slip planes are plastically activated and then finite slip band thickness should be accounted for. Numerous crystalline finite element (FE) computations have been carried out using considering a slip bands with low critical resolved shear stress embedded in an elastic matrix. The computed GB normal and shear stress fields:

– are considerable lower than the pile-up ones and exhibit strong dependency on the slip band thickness close to the SB corner
– but are in fair agreement with the solution predicted by the pile-up theory far away.

Since the pile-up theory leads to the overestimation of the local GB stress fields, the main goal of the current paper is to perform analytical model of GB stress components based upon FE calculations. The effect of various parameters can be understood in the framework of matching asymptotic expansions which is usually applied to cracks with V notches of finite thickness. Finally, the predicted remote stresses to GB fracture in pre-irradiated austenitic stainless steels subjected to tensile loading in various environment are compared to experimental data and the pile-up based predictions.

Keywords Micro-cracks · Slip bands · Pile-up theory · Linear fracture mechanics · FE method · Crystalline plasticity

M. Sauzay (✉) · M. O. Moussa
CEA, DEN, DMN, SRMA, 91191 Gif-sur-Yvette, France
e-mail: maxime.sauzay@cea.fr

1 Introduction

Slip localization occurring at the grain scale has been extensively observed, particularly in Faced-Centred Cubic (FCC) metals and alloys subjected to either post-irradiation tensile tests (proton or neutron irradiation at high dose) (Sharp 1967; Victoria et al. 2000; Lee et al. 2001; Edwards et al. 2005; Jiao et al. 2005; Byun et al. 2006), cyclic loadings (Lukas and Knesnil 1968; Finney and Laird 1975; Winter et al. 1981; Blochwitz and Veit 1982; Lim and Raj 1984a; Mughrabi and Wang 1988; Man et al. 2002), or even simply tensile loading (Perrin et al. 2010). Plastic slip is localized in thin slip bands. Their thickness is lower than 1 μm but higher than a few ten nm. Usually, slip bands cross all the grains from one grain boundary to the

opposite one. Therefore the slip band length is approximately equal to the grain size which usually varies from a few microns to a few hundred microns. Following the Atomic Force Microscopy (AFM) measurements of Jiao et al. (2005), the local SB plastic deformation is ten times higher than the remote one in austenitic steels subjected to post-irradiation tensile loading. Transmission Electronic Microscopy (TEM) observations lead to similar evaluations (Sharp 1967; Edwards et al. 2005; Sauzay et al. 2010). Such thin SBs are called channels or clear bands because they are almost empty of irradiation defects whereas the surrounding matrix is still full of defects. Their thickness is about 50 nm. Plastic strain is highly localized in slip bands induced by cyclic loadings as well. Such SBs are often called persistent slip bands (PSBs). Their local plastic strains are about hundred times higher than the remote ones (Winter et al. 1981; Wejdemann and Pedersen 2004; Weidner et al. 2006, 2010). PSB thickness is about 0.5μm in Face Centred Cubic (FCC) polycrystals (Mughrabi and Wang 1988; Man et al. 2002). Slip localization is observed during tensile loading as well (Tong and Hector 1997; Perrin et al. 2010).

At least two main mechanisms explain the slip band formation. The first mechanism holds for mechanical tests carried out on pre-irradiated FCC metals/alloys, as well as alloys containing shearable precipitates (Gerold and Steiner 1982). It is based on interactions between mobile dislocations and irradiation-induced defects/precipitates (Lee et al. 2001). The second mechanism corresponds to the case of FCC (ductile) metals and alloys subjected to cyclic loadings. Slip localization seems to be essentially due to dislocation glide, cross-slip and interactions. Cross-slip seems to be required for PSB formation (Essmann and Differt 1996). Similar slip localization has often been reported in body centred cubic (BCC) or hexagonal compact (HCP) metals and alloys.

Several computations were carried out for evaluating the plastic slips inside slip bands. Authors first modelled slip bands as elongated inclusions embedded in a matrix which mimics the whole polycrystal (Rasmussen and Pedersen 1980). This permitted them to use the analytical solution given by Eshelby (1957) for a bulk inclusion. Polycrystalline homogeneisation taking into account slip localization at the grain size allowed Collard et al. to take into account the ratios between SB spacing and grain size as well as SB thickness and grain size (Collard et al. 2010). Then, Finite Element computations using crystalline plasticity permitted the investigation of surface effects (Repetto and Ortiz 1997; Sauzay and Gilormini 2000, 2002). Analytical modelling has been proposed as well (Li et al. 2007; Sauzay and Gilormini 2000).

Clear bands and slip bands impinge to grain boundaries. This induces stress or plastic strain concentrations as shown in a copper polycrystal deformed after neutron irradiation by Edwards et al. (2005) who observed indeed either local lattice rotations corresponding to high local strain concentrations or a considerable amount of (plastic) shearing at the grain boundary as another channel has been nucleated on the opposite side of the grain boundary and spreads in the neighbour grain. Such propagation of a channel in the neighbouring grain was more often observed in the case of singular grain boundaries such as twin boundaries (Jiao et al. 2005; Liu et al. 1992). If no transmission through GBs occurs, then large stress concentrations are induced by the impingement of SBs towards GBs. Recently, high resolution EBSD allowed the measurement of elastic strains at the submicron scale showing the strong stress concentrations induced by slip localization (Ben Britton and Wilkinson 2012).

Because of these interactions with GBs, clear bands or slip bands are often considered as triggering grain boundary crack initiation and propagation. Such a mechanism has been investigated experimentally for copper (Liu et al. 1992) and nickel (Lim and Raj 1984b) polycrystals subjected to cyclic loadings. Similarly, slip localization in clear bands has been considered as promoting intergranular crack initiation in the case of irradiation assisted stress corrosion cracking (IASCC) (Simonen and Bruemmer 1995; Toivonen et al. 2005; Edwards et al. 2005). Impinging deformation-induced twin bands have been found to trigger intergranular crack initiation as well (Toivonen et al. 2005; Chung et al. 2001; Onchi et al. 2003). Twinning is another mechanism of slip localization inducing in addition crystallographic rotation (Kadiri et al. 2013). Finally, previous cold-work deformation of austenitic steels leads to earlier GB crack initiation during SCC tests, which is considered to be due to slip localization during cold-work (Couvant 2001). Many experimental studies concerning the influence of grain boundary characteristics on intergranular crack initiation have been published (McMurtrey and Was 2011). Concerning grain boundaries, two extreme cases can be considered (Sutton and

Balluffi 1995; Priester 2001). On the one hand, general GBs display almost no crystal periodicity along the grain boundary plane. Their energy as well as their diffusion coefficients are very high (Sutton and Balluffi 1995; Priester 2001). On the other hand, special boundaries present a crystal periodicity along the grain boundary plane. Their grain boundary energy as well as their diffusion coefficients are low (Sutton and Balluffi 1995; Priester 2001). All authors concluded that special boundaries, and particularly $\Sigma 3$ twin boundaries, are the less prone to stress corrosion cracking (SCC) initiation even if some of them may crack (Alexandreanu and Was 2003; Tan et al. 2005). The same result was obtained in copper (Liu et al. 1992) or nickel (Lim and Raj 1984b) subjected to cyclic deformation carried out in either air or inert environment. Their low GB energy values lead to high fracture energy and the observed slip band transmission through special GBs decreases GB stress concentrations which is not the case for general GBs.

Several modelling approaches focused on the evaluation of grain boundary stress concentrations. Neumann showed that crystalline elasticity induces stress concentration at grain boundaries (Neumann 1992) in agreement with experimental analysis of slip traces (Chuang and Margolin 1973; Margolin and Stanescu 1975). Recently, Diard et al. (2005) used large-scale Finite Element computations for evaluating stress gradients in the vicinity of grain boundaries, induced by plastic deformation incompatibilities between neighbour grains. All these studies highlighted stress concentrations which may promote intergranular crack initiation. Concerning the influence of slip band impingement, GB stress fields have been evaluated analytically using the theory of discrete or continuous dislocation pile-ups. This approach is based on the well-known Stroh model (1957). The stress singularity induced by an edge or screw pile-up of length $L^{\text{pile-up}}$ is the same as the one of a crack in the framework of linear elastic fracture mechanics (LEFM) (Stroh 1957; Smith and Barnby 1967). This length is usually assumed to be close to one-half of the grain size, L. Thanks to the similarity with the LEFM crack problem, the energy release rate, G, may be computed in a straightforward way. Following the pioneering work of Griffith, an energy balance criterion has been applied by Smith and Barnby for predicting GB microcrack initiation (Smith and Barnby 1967). As the stress singularity exponent is 1/2, the application of the Griffith criterion leads to possible microcrack initiation, which is not be true for lower stress exponent values (Leguillon 2002; Taylor et al. 2005). The Griffith criterion is based on the equality between the energy release rate, G, and the GB fracture energy, γ_{fract}. This means that only an energy criterion is required and the crack increment is assumed to be infinitesimal. This modelling has been applied to the prediction of GB microcrack initiation, either in copper polycrystals subjected to cyclic loading (Liu et al. 1992) or pre-irradiated austenitic stainless steels subjected to tensile loading (Evrard and Sauzay 2010). These authors predicted critical remote tensile stresses much lower than the observed ones, whatever the environment (Evrard and Sauzay 2010). Therefore, the pile-up theory seems to lead to underestimations of the critical remote stresses when compared to experimental data.

Pile-up theories assume that slip is localized on one atomic plane only. But, many experiments and observations show that for many materials and loading conditions, a non-negligible fraction of the slip occurs inside the fatigue slip bands (interferometry measurements (Finney and Laird 1974), TEM observations (Sauzay et al. 2010) and AFM measurements (Jiao et al. 2005; Wejdemann and Pedersen 2004; Tabata et al. 1983; Weidner et al. 2006 and Weidner et al. 2010). Concerning 316L austenitic stainless steel deformed after pre-irradiation, Byun et al. (2006) concluded that shear strain is uniformly distributed through the thickness of channels (clear bands). Similar conclusions were drawn by Jiao et al. (2005) and Sauzay et al. (2010). As plastic slip is indeed much more homogeneously distributed than assumed by pile-up theories, these last ones may overestimate the local GB normal stress fields as well as energy release rate values which may lead to the underestimation of the critical remote stress mentioned previously. Taking into account not only the slip band length, L, but also its thickness, t, may lead to more realistic GB stress fields and improve the microcrack initiation predictions. The finite element (FE) method has been used recently in the framework of crystalline elastoplasticity because of the non-linear behaviour of slip bands (Sauzay and Vor 2013). Slip bands of various thickness and lengths were embedded at the free surface of an elastic matrix. The effect of slip band thickness and length as well as remote tensile stress was studied based on the results of numerous FE computations. Analytical formulae describing the GB normal stress singularities induced by slip bands

were deduced but only for steels and only particular microstructure geometry.

The present study aims to compute both GB normal and shear stress fields in various FCC metals/alloys and for various microstructure geometries. For instance, various GB orientations should be considered (McMurtrey and Was 2011) and slip bands appear not only in well-oriented grains (Schmid factor equal to 0.5), but also in grains of various crystallographic orientations and therefore various orientations may characterize slip planes (Mughrabi and Wang 1988; McMurtrey and Was 2011). In order to deduce analytical formulae predicting GB stress fields and GB fracture, numerous FE computations considering a much larger range of material parameters and microstructure geometries have been carried out. The study focuses on:

- both normal and shear stress fields;
- the effect of isotropic elasticity parameters as well as crystalline elastic anisotropy;
- the influence of grain boundary orientation and crystal orientation;
- use of finite fracture mechanics for predicting the remote stress to GB microcrack initiation. The GB stress fields and critical remote tensile stress predicted by pile-up theories are compared to the ones taking into account slip bands of finite thickness, $t > 0$;
- finally, the remote stress required for grain boundary fracture of pre-irradiated austenitic stainless steels subjected to tensile loading in either inert or oxidizing environment is computed using fracture energy data published in the literature and avoiding any adjusted parameter. The values are quantitatively compared to experimental data and once more to the predictions of the classical theory of pile-ups.

2 Computation hypothesis

2.1 Modeled microstructures

To study the influence on GB stress fields of SB thickness and length as well as crystallographic and GB orientation, various microstructures (Fig. 1) have been generated using the FE software Cast3m (Cast3m 2012). Individual SBs, defined by their thickness, t, length, L, inclined by an angle α_{SB} with respect the tensile axis, and surrounded by a grain, have been generated. The considered grain and its SB are embedded at the free surface of a homogeneous matrix which mimics the other grains of the whole polycrystal. The vector perpendicular to the GB is inclined by an angle α_{GB} with respect the tensile axis. In order to maximize strain localization, the SB emerges at the free surface. To study the influence of the two angles, α_{SB} and α_{GB}, various microstructures, have been generated, using various values of both angles, belonging to the 40°–70° and 20°–50° ranges respectively. For instance, the mesh of the microstructure built for $\alpha_{SB} = 45°$ and $\alpha_{GB} = 0°$, is plotted in Fig. 2a (zoom on the slip band). The SB length, L, is at least ten times lower than the whole mesh size in order to avoid any influence of the whole matrix size. Therefore, using a large matrix does not affect the computed GB stress fields.

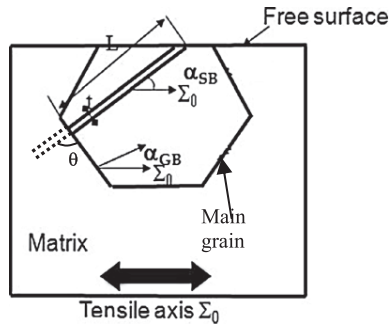

Fig. 1 Microstructure with a slip band and its grain embedded at the free surface of the matrix (zoom). Microstructure angles: $\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$ (for instance). The angle between the SB and the GB planes, θ, may be easily deduced: $\theta = 90_{-\alpha_{SB}+\alpha_{GB}} = 78°$

Following several observations from the literature (Mughrabi and Wang 1988; Man et al. 2002), slip bands are mostly well-oriented (maximum value of the Schmid factor: 0.5). For the FCC metals or alloys, the slip systems are defined by the {111} normal directions and ⟨110⟩ slip directions (12 slip systems). The well-oriented grains present one slip system (n,m) with a Schmid factor, f, equal to its maximal value, 0.5. The (unit) normal vector to the slip system is denoted as n whereas the (unit) slip vector is denoted as m. For such a crystal orientation, both the n and m vectors are inclined of 45° with respect to the tensile direction with $\alpha_{SB} = 45°$, (Fig. 2a). The following directions of the well-oriented slip system have been chosen: $n = 1/\sqrt{3}(1, 1, 1)$ and $m = 1/\sqrt{2}(1, 0, -1)$ for FCC metals. Slip bands are parallel to the easy slip plane

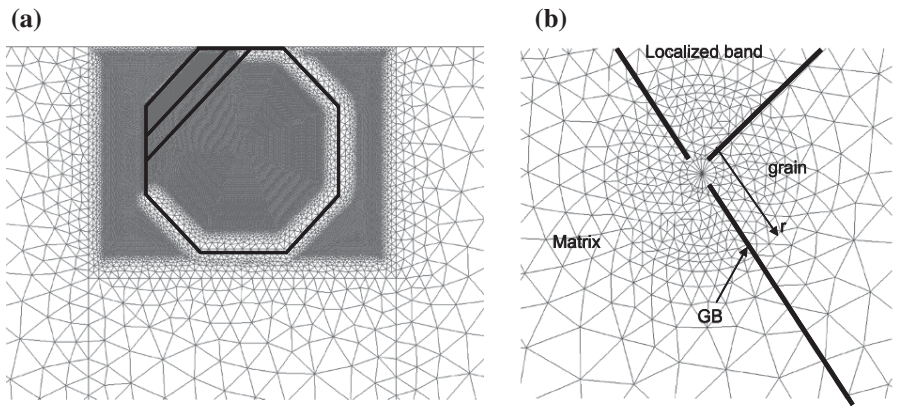

Fig. 2 **a** Mesh of the microstructure corresponding to: $\alpha_{SB} = 45°$ and $\alpha_{GB} = 0°$ (Fig. 1) and built using the Cast3M FE software. **b** Radial mesh generated at the intersection between a slip band and a GB. Mesh built using the Cast3M FE software

and therefore perpendicular to the n vector. (Byun et al. 2006). This conclusion holds for BCC and HCP metals and alloys in which slip localization is observed as well (Onimus et al. 2004).

The resolved shear stress is defined as the maximum value among the shear stress amplitudes computed on the 12 slip systems and the Schmid factor is the ratio between the resolved shear stress and the macroscopic tensile stress or stress amplitude, Σ_0. Assuming a homogeneous stress state, the Schmid factor, f, depends on the α_{SB} angle only: $f = \cos(\alpha_{SB})\sin(\alpha_{SB})$. Because of the scatter induced by the neighbour grain effect (Sauzay and Jourdan 2006) or because of an increasing remote stress (Jiao et al. 2005), grains of various crystallographic orientations and lower Schmid factors may contain slip bands. That is why we make the α_{SB} angle vary in the 35°–70° range. This corresponds to the following range of Schmid factor values: 0.32–0.5.

As random microstructures usually contain GBs of various orientations, that is why we make the angle α_{GB} vary in the 20°–50° range. The remote GB normal stress depends on the α_{GB} angle and remote tensile stress only: $\Sigma_n^\infty = \cos^2(\alpha_{GB})\Sigma_0$. The same holds for the remote GB shear stress: $T_{nm}^\infty = \cos(\alpha_{GB})\sin(\alpha_{GB})\Sigma_0$.

Following the pile-up model (Stroh 1957), the GB stress field in the vicinity of the head of a pile-up may be written as $\sigma_n^{\text{pile-up}}(r,\theta) = f(r)h(\theta)$. The SB would be replaced by a single plane in Fig. 1 and (r, θ) are the polar coordinates (Figs. 1, 2b). Then, because of the similarity between a pile-up and a slip band, a radial mesh has been generated at the intersection Channel-GB (Fig. 2b). The angle between the SB and GB plane, θ, depends on the SB and GB orientation as follows: $\theta = 90_{-\alpha_{SB}+\alpha_{GB}}$. This angle will be used in Sect. 4 for plotting the dependence of on the singularity parameters with respect to both α_{SB} and α_{GB} angles. Numerous meshes including a SB have been built considering not only values of the length L varying from 3 to 100 μm, but also values of the thickness t, varying from 40 to 500 nm. The meshes have been extruded in the third direction and 3D meshes of the microstructures are finally built. The influence of mesh size has been carefully studied and it is low enough to let us avoid any influence on the computed GB stress fields. The relative numerical error concerning the GB stress values in the vicinity of the SB-GB intersection is lower than 5 %. In the following, the GB stress fields will be plotted as a function of the distance, r, such as defined in Fig. 2b. Finally, the time increment has been checked to be low enough to avoid any effect on the computed stress fields

The microstructures are subjected to a uniform tensile stress, Σ_0, on the two right and left matrix vertical boundaries, such as shown in Fig. 1. Some displacements have been set to zero in order to avoid any rigid body motion and impose plane strain state.

2.2 Constitutive laws

2.2.1 Elasticity

As a first approach, slip band, grains and matrix are assumed to obey isotropic elasticity which may be characterized for instance by the values of the Young's modulus, Y, and the Poisson ratio, v. Using isotropic elasticity allows an easy comparison between the FE results and the predictions of the pile-up theory which

generally assumes homogeneous isotropic elasticity. In order to provide general results, several materials have been considered: aluminium, copper and austenitic stainless steels. The values of the parameters of isotropic elasticity have been found in (Matweb.com 2012). For instance, the Young's modulus and Poisson ratio used for austenitic stainless steel are equal to 200 GPa and 0.3, respectively.

In reality, grains obey crystalline elasticity rather than isotropic elasticity. In the case of cubic symmetry, only three parameters are required, C_{11}, C_{12} and C_{44} (Huntington 1958). Even if the polycrystal is untextured, the effect of the anisotropy of crystalline elasticity may be noticeable at the grain scale as shown by multicrystal FE computations (Sauzay and Jourdan 2006; Sauzay 2007; Sauzay and Man 2008). In the case of cubic symmetry, the anisotropy level may be evaluated using the anisotropy coefficient, $a = 2C_{44}/(C_{11} - C_{12})$. It is equal to the ratio between the maximum and minimum elastic shear moduli considering all slip systems in a 3D continuum framework. In copper and austenitic stainless steels, its value is as high as 3.3, in nickel and ferrite, it reaches 2.5 and is only 1.1 in aluminium crystals which behave almost on an isotropic manner. As expected, the anisotropy coefficient, a, is equal to one in an isotropic material. For these metals and alloys, the maximum of the Young's modulus values is reached along the $\langle 111 \rangle$ direction, Y_{111}, and its minimum along the $\langle 100 \rangle$ direction, Y_{100}. The ratio of both, Y_{111}/Y_{100}, is very close to the value of the anisotropy coefficient, a.

Computations have been carried out for austenitic stainless steels using:

- either crystalline elasticity in the surface grain ($C_{11} = 197$ GPa, $C_{12} = 125$ GPa and $C_{44} = 122$ GPa leading to a strongly anisotropic behaviour with $a = 3.3$);
- or isotropic elasticity (same parameters as the macroscopic ones which leads to $a = 1$). This allows us to study the effect of crystalline elasticity on the local GB stress fields.

2.2.2 Crystalline (visco)plasticity

For a Face-Centred Cubic (FCC) metal or alloy, twelve easy slip systems, (n_i, m_i), are defined in each crystal. The unit normal vector of the ith slip system is denoted as n_i whereas its unit slip vector is denoted as m_i. The shear stress on each slip system is defined by $|\tau_i| = |m_i^T \boldsymbol{\sigma} n_i|$ ($i = 1, \ldots, 12$). The local stress tensor is denoted as $\boldsymbol{\sigma}$.

Basic crystalline (visco)plasticity laws are applied to each of the twelve slip systems through a slip criterion and a linear hardening evolution law of slope H_0. In the case of increasing load, the Schmid criterion (Eq. 1) characterizes the activation of the i^{th} slip system, using the shear stress, $|\tau_i|$, and its critical value, $\tau_{c,i}$, which initial value is denoted as τ_0. Equation 2 describes the evolution law of the critical shear stress $\tau_{c,i}$ with $\mathbf{d}$ the interaction matrix between the twelve slip systems. The plastic slip on the ith slip system is denoted as γ_i^p. The three parameters of the crystalline (visco)plasticity laws are: τ_0 (the initial critical shear stress), H_0 (the linear hardening slope) and q (the latent hardening coefficient).

$$|\tau_i| = \tau_{c,i} \tag{1}$$

$$\dot{\tau}_{c,i} = H_0 \sum_{j=1}^{12} d_{ij} \left| \dot{\gamma}_j^p \right| \quad \text{with } d_{ij} = q \text{ for } i \neq j \text{ and } d_{ij} = 1 \text{ for } i = j \tag{2}$$

These laws have been implemented in the software Cast3m using a subroutine UMAT which has already used in (Sauzay et al. 2010; Evrard and Sauzay 2010) for studying the effect of slip localization in irradiated metals and alloys. As explained below, viscosity effects may be neglected for the considered materials and loading conditions. Viscoplasticity laws are nevertheless used because they improve numerical convergence. The exponent of the power law linking shear stress to (visco)plastic slip rate is higher than 100. Practically, no difference in the GB stress fields is observed for applied strain rate between 10^{-8} and 10^{-3}/s. That is why we may considere that the SB behaviour is rate-independent and SBs obey quasi-elastoplasticity.

The finite strain theory as well as crystallographic rotations are accounted for. But, it should be noticed that our FE computations show that their effect rotations on GB normal and shear stress fields is rather weak.

Enhanced (visco)plasticity laws may be used in case of Portevin-Le Châtellier effect (Zavattieri et al. 2009) or yield point elongation followed by hardening (Xie et al. 2004).

2.2.3 Adjustment of slip band plasticity parameters

The typical values of the initial critical shear stresses, τ_0, are about a few ten MPa for either channels (Sharp 1967; Yao et al. 2002, 2004) or PSBs (Lukas and Knesnil 1968; Finney and Laird 1975; Winter et al. 1981; Blochwitz and Veit 1982).

The hardening slope in channels in pre-irradiated metals is usually considered to be weak (Sharp 1967; Yao et al. 2002, 2004), similar to the ones measured during tensile loading of well-oriented single crystals ($H_0 \sim \mu/2000$). Most often, only single slip is observed in slip bands (Sharp 1967) which means that very high values of the latent hardening coefficient, $q \gg 1$, should be used in order to hinder secondary slip as experimentally observed. And viscosity is considered to be negligible in agreement with macroscopic and microscopic observations (Sharp 1967). Typical tensile curves of pre-irradiated materials with single crystal microstructure are plotted in Fig. 3a.

Similar observations have been made concerning PSB hardening (Lukas and Knesnil 1968; Finney and Laird 1975; Winter et al. 1981; Blochwitz and Veit 1982), predominant single slip (Lukas and Knesnil 1968; Finney and Laird 1975) and very weak viscosity effect at room temperature (see Sauzay and Kubin 2011 for a recent review).

The effect of SB plasticity parameters, τ_0, H_0, and q, has been studied in detail in Sauzay and Vor (2013). As expected, the higher the SB hardening, the lower the stress fields. Activation of several slip systems instead of the primary one only leads to slightly higher GB stresses (+10 %). The closed-form expressions proposed in Sect. 3.2 and valid for quasi perfect plasticity may be easily extended to quasi plasticity including hardening using the formulae given in (Sauzay and Vor 2013).

3 Influence of slip band thickness on grain boundary stress fields: numerical and theoretical results

3.1 Finite element computation results

First of all, the dependence of grain boundary normal stress field with respect to the slip band crystalline length, L, and thickness, t, is studied. The microstructure is defined by: is $\alpha_{SB} = 45°$ (well-oriented grain) and is $\alpha_{GB} = 33°$ (slightly inclined GB with respect to the tensile direction). Stress concentrations at the bottom of the slip band plotted in Fig. 3b is clearly visible. As we intend to predict GB microcrack fracture initiation, the study of the GB normal stress fields, $\sigma_n(r)$, is of main interest, as well as the GB shear stress field,

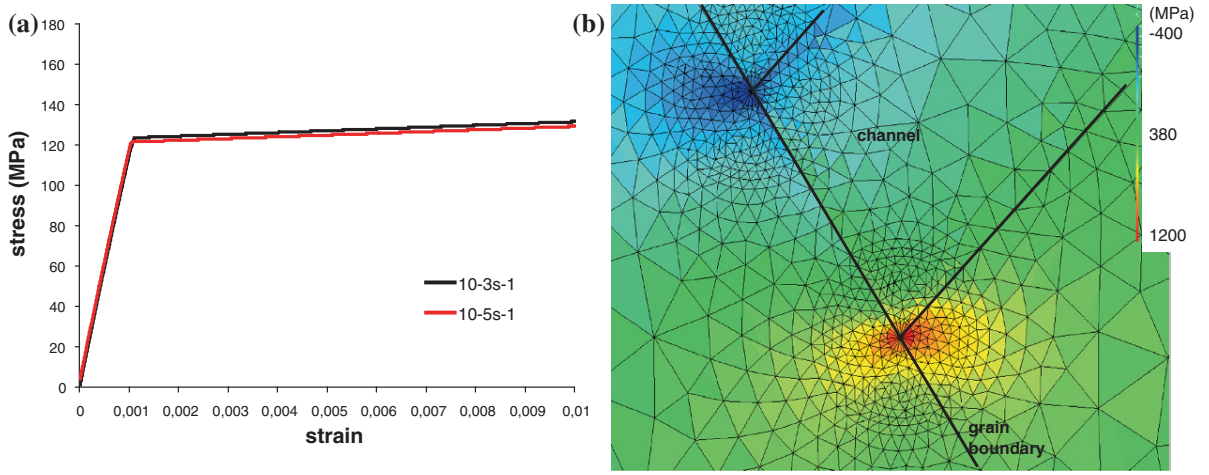

Fig. 3 **a** Tensile curve characterizing the plasticity behaviour of slip bands. Low initial critical shear stress: $\tau_0 = 1 - 100$ MPa (here 60 MPa), latent hardening coefficient: $q \sim 10^5$ (single slip), low hardening slope: $H \sim 0 - 250$ MPa, no strain rate effect (quasi-perfect plasticity). **b** Isovalues of the local tensile stress, σ, plotted in the vicinity of the intersection between the slip band and the grain boundary. Parameters: $L = 10\,\mu m$, $t = 0.1\,\mu m$, $\tau_0 = 60$ MPa, $H_0 = 250$ MPa, $q = 1.4$, $\Sigma_0 = 300$ MPa ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$)

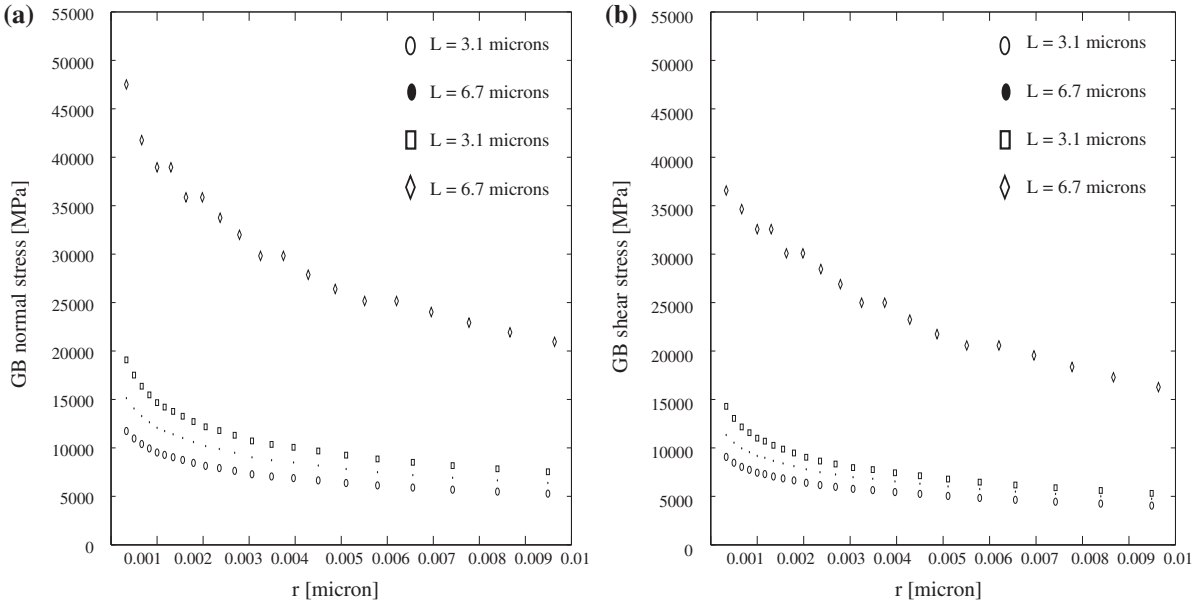

Fig. 4 **a** Plots of the GB normal stress versus r for various slip band length values, L ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $t = 0.09\,\mu m$ $\tau_0 = 60\,MPa$, $H_0 = 1\,MPa$, $q = 10^5$, $\Sigma_0 = 878\,MPa$); **b** plots of the GB shear stress in similar conditions

$\tau_n(r)$. Large ranges of variation of L have been taken into account: 3–200 μm as well as the thickness range 40–500 nm.

The higher the slip band length, the higher the GB normal stress as shown in Fig. 4a. The same result holds for the GB shear stress (Fig. 4b). As shown later, the dependence to the SB length, L, is similar to the one predicted by the pile-up theory, that is the square root one. It should be noticed that GB normal and shear stress are rather close, which means that both GB brittle fracture and GB sliding may be triggered by slip localization depending on temperature and GB properties (Lim and Raj 1984a and Lim and Raj 1984b). As shown in Sect. 4, the ratio between the GB shear and normal stress depends strongly on the GB orientation.

But, contrary to the pile-up theory, the slip band thickness affects the GB normal stress field (Fig. 5a, b). The higher the thickness, the lower the GB normal stress computed at the same distance from the slip band corner. Its influence is not negligible which shows that the pile-up theory is not valid any more close to the slip band corner. Indeed, the pile-up based modeling overestimates local GB normal stress fields near the slip band corner (Fig. 6). GB stress fields are more sensitive to the slip band length than to its thickness, as may be guessed by comparing Figs. 4 and 5. As shown in Sect. 3.2, the power-law dependence of the stress fields with respect to the SB thickness, t, is in fact lower than the one with respect to the SB length, L.

These curves are plotted considering the close-field domain only. As expected from the theory of matching asymptotic expansions (Leguillon and Sanchez-Palencia 1987; Murer and Leguillon 2010) applied to cracks with a V-notch tip, two domains should indeed be distinguished (Fig. 6a, b):

- the close-field domain (Fig. 7a) in which the local geometry of the crack tip affects the stress field ($r \ll t$). The slip band behaves similarly (Fig. 5a, b);
- the far-field domain (Fig. 7b) in which the tip geometry does not affect the stress field any more and the crack can be considered as sharp crack ($t \ll r \ll L$). Once more, the same result holds for a slip band (Fig. 5a, b).

3.2 Analytical description of the GB stress fields

As expected, the stress fields computed by the FE method are not affected by the thickness t far away from the slip band (Fig. 5a, b). They are in fact very close to the one predicted by the pile-up theory. This means

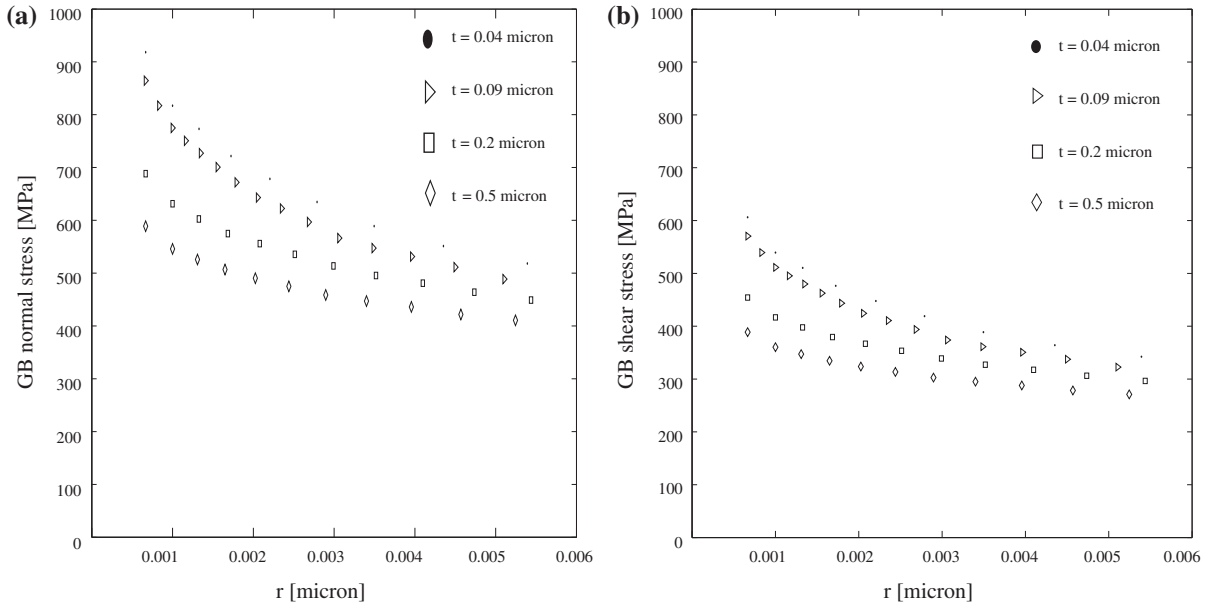

Fig. 5 **a** Plots of the GB normal stress versus r for various slip band thickness values, t ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $L = 10.7\,\mu m$, $\tau_0 = 60\,MPa$, $H_0 = 1\,MPa$, $q = 10^5$, $\Sigma_0 = 156\,MPa$); **b** plots of the GB shear stress in similar conditions

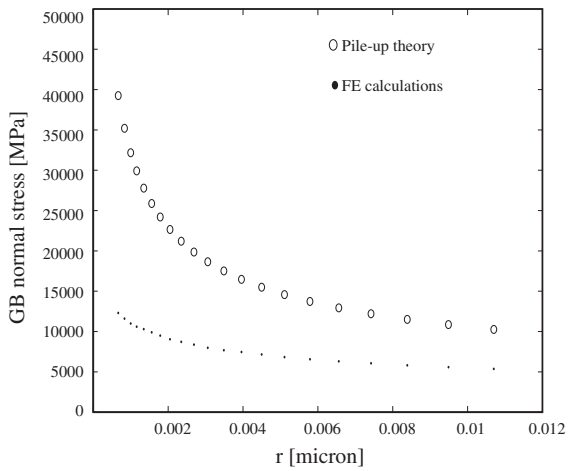

Fig. 6 Plots of the GB normal stress field computed either using the FE method (finite SB thickness, $t = 90\,nm$) or the pile-up theory (edge dislocations, $L_{pile\text{-}up} \approx L/2$, see Sect. 2.1) ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $L = 10.7\,\mu m$, $\tau_0 = 60\,MPa$, $H_0 = 1\,MPa$, $q = 10^5$, $\Sigma_0 = 878\,MPa$)

that the GB stress field is well predicted by the pile-up theory, provided $t \ll r \ll L$ which means that the considered GB segment is far away enough from the slip band corner. On the contrary, close to the slip band corner (Fig. 6a), the pile-up theory overestimates largely the GB stress fields obtained in case of slip bands of finite thickness, t. For $r/t = 0.001$, the GB normal stress is overestimated by a factor three by the pile-up theory. As it may guessed from the curves plotted in Fig. 5a, b, the stress singularity exponent in the case of finite thickness is lower than the classical pile-up or LEFM crack one which is about 1/2. This is now discussed in details in order to provide analytical formulae allowing an easy computation of the GB normal stress fields in case of slip bands of finite thickness, t.

The theory of matching expansions (Leguillon and Sanchez-Palencia 1987; Leguillon 2002) applied to the case of a crack with a V-noch tip in an elastic matrix predicts the dependence of the stress close field with respect to the crack length, L, the notch thickness, t, and the stress exponent characterizing the V notch singularity, $1 - \lambda$:

$$\sigma_{ij}\left(r,\theta'\right) = A'\Sigma_0\sqrt{\frac{L}{t}}\left(\frac{t}{r}\right)^{1-\lambda} p_{ij}\left(\theta'\right) + \Sigma_{0,ij} \qquad (3)$$

with A' a geometry factor which is usually computed by the FE method, $p_{ij}(\theta')$ describing the dependence with respect to the polar angle, θ', and Σ_{ij} the remote stress tensor components (Leguillon 2002). Polar coordinates are indeed used as plane strain problems are considered here. This theory is based on the inner expansion (close-field for $r \ll t$) and the outer expansion (far-field for

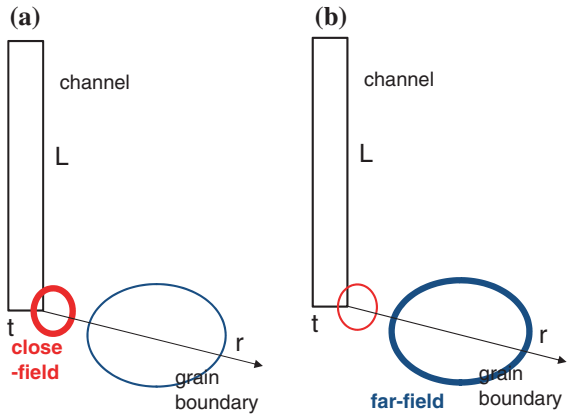

Fig. 7 **a** Schematic view of the close-field area ($0 < r \ll t$) (*red circle*) with t and L the thickness and length of the channel/slip band ($t \ll L$). **b** Schematic view of the far-field area ($t < r \ll L$) (*blue circle*) with t and L the thickness and length of the channel/slip band ($t \ll L$)

$t < r \ll L$) (Leguillon and Sanchez-Palencia 1987; Leguillon 2002). In an intermediate area, it is assumed that the inner and outer expansions hold and are then close. This corresponds to the so-called 'matching conditions'.

The problem we are interested in differs from this one by several ways

- the slip band corner is not a V-notch but a flat notch. Nevertheless, this does not change the previous general expansion except that the geometry coefficient and the stress singularity exponent, $1 - \lambda$, would be different from the ones computed for a V-notch;
- the slip band is inclined with respect to the loading axis. This only means that not only mode I load is of interest but also mode II;
- more importantly, slip bands obey non-linear behaviour whereas the theory of matching expansions is usually applied in the LEFM framework. Nevertheless the close-form asymptotic fields predicted by the linear theory of matching expansions may be used as guidelines for proposing analytical formulae which should then be validated with respect to the GB stress fields plotted using the FE method. It should be added that if slip band hardening is negligible and the latent hardening coefficient, q, is very high, the long slip band boundaries of length, L, are loaded in mode II by a shear stress of almost τ_0. Moreover, a condition is applied on the stress vector acting on the short slip band boundary which depends on τ_0 (shear component of the stress tensor in plane strain). In the case of a GB perpendicular to the SB plane ($\theta = 90° + \alpha_{GB} - \alpha_{SB} = 90°$, Fig. 1a) then the short bottom boundary is subjected to an almost uniform shear stress, τ_0, in the case of negligible hardening. This may looks like to an elastic problem involving stress boundary conditions on the boundaries of a blunted crack with a square tip at its bottom (Fig. 1) and reinforces the analogy with the problem of a LEFM crack with a V-notch solved by the theory of matching expansions. But this analogy does not allow the direct comparison of an analytical approach and FE computations carried out independtly as it can be made in case of elasticity (Zavattieri et al. 2007a).

Therefore, the quantitative dependence of the GB normal and shear stress fields with respect to the driving force, $T = f\Sigma_0 - \tau_0$, the slip band length, L and t and finally the distance to the slip band corner, r, has been studied in details based on the numerous results of FE computations:

- the stress fields are found to be proportional to the driving force, $T = f\Sigma_0 - \tau_0$ which is expected following the analogy with a crack subjected to a remote mode II shear stress, $f\Sigma_0$, and a crack lip friction stress, τ_0 (Fig. 8a);
- the dependence of the stress close-fields with respect to r/t is shown to be correctly described by a power law of exponent $(1 - \lambda)$. The stress fields are in addition inversely proportional to the square root of the aspect ratio, L/t as expected from the theory of matching expansions (Eq. 3);
- the exponent of the stress singularity about $1 - \lambda \approx 0.27$ whatever the GB an slip band orientations whereas the one of the far-field is about 0.5. The close-field stress singularity is indeed weaker than the pile-up one as shown in Fig. 6.

The GB normal stress field may therefore be described by an equation similar to the one provided by the theory of matching expansions, that is, for a negligible hardening slope:

$$\sigma_n(r) = A_{nn}(f\Sigma_0 - \tau_0)\sqrt{\frac{L}{t}}\left(\frac{t}{r}\right)^{1-\lambda} + \Sigma_n^{\infty} \quad (4)$$

Similarly, the GB shear stress field may be described by:

$$\tau_{nm}(r) = A_{nm}(f\Sigma_0 - \tau_0)\sqrt{\frac{L}{t}}\left(\frac{t}{r}\right)^{1-\lambda} + T_{nm}^{\infty} \quad (5)$$

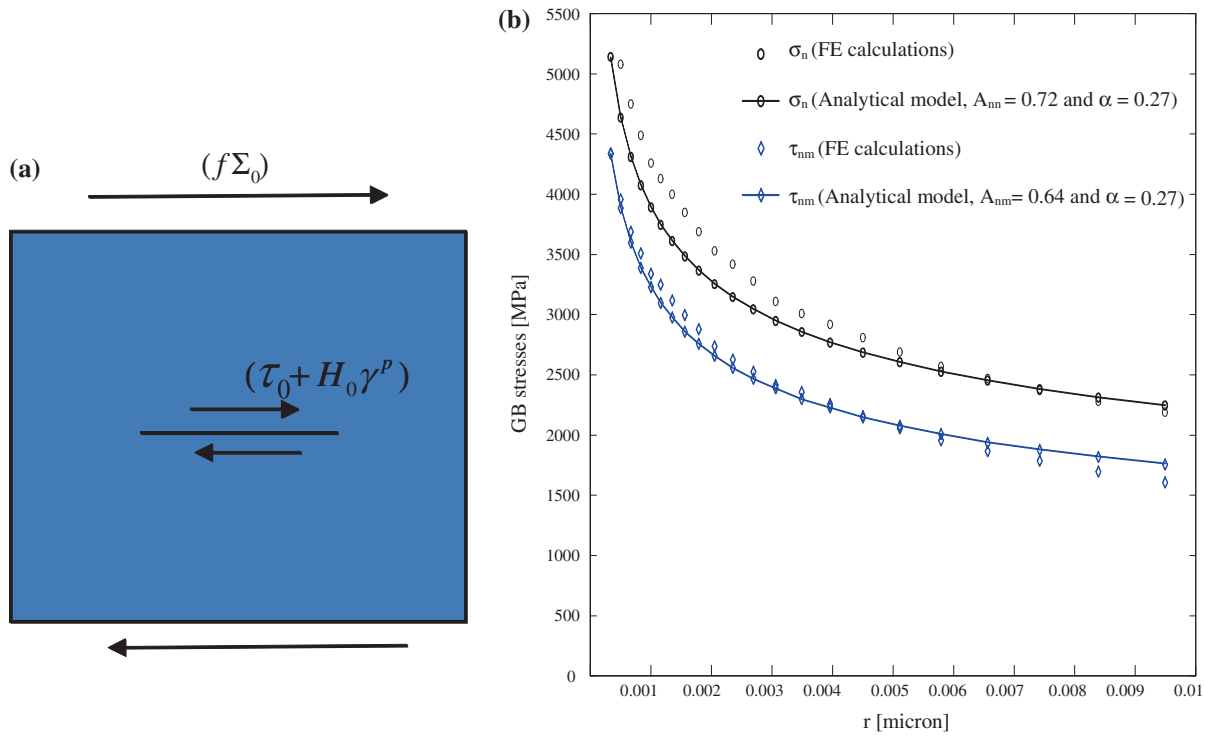

Fig. 8 **a** Crack or slip band embedded in an infinite matrix loaded by pure shear. Remote shear stress: $f\Sigma_0$ with the Schmid factor, f = 0.5. Shear stress applied on the crack lips or slip band shear stress: $\tau_c = \tau_0 + H_0\,\gamma^p$; **b** Numerical evaluation of the stress singularity parameters: $A_{nn} = 0.72$, $A_{nm} = 0.64$ and $1 - \lambda = 0.27$ (Eqs. 4 and 5) ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, L = 10.7 μm, t = 0.09 μm, τ_0 = 60 MPa, H_0 = 1 MPa, q = 10^5, Σ_0 = 393 MPa)

The geometry coefficients, A_{nn} and A_{nm}, as well as the stress exponent, $1 - \lambda$, depend on the GB–SB angle, $\theta = 90_{-\alpha_{SB}+\alpha_{GB}}$ (Fig. 1). The Schmid factor, f, depends on the SB angle, α_{SB}, as mentioned in Sect. 2.1. And the remote GB normal and shear stresses, denoted as Σ_n^∞ and T_{nm}^∞ depend on the GB angle, α_{GB}, as shown in Sect. 2.1. For each set of GB / SB angles, the stress singularity parameters, A_{nn}, A_{nm} and $(1 - \lambda)$ are evaluated using one set of GB normal/shear stress curves computed by the FE method for one value of the remote stress and one single set of SB sizes, L and t (Fig. 7b). For instance, for $\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°(\theta = 53°)$, the adjusted values are: $A_{nn} = 0.72$ $A_{nm} = 0.64$ and $(1 - \lambda) = 0.27$ (Fig. 8b). Then, the analytical formulae are validated using numerous curves computed by the FE method for various applied stresses, Σ_0, and slip band characteristic lengths, L and t, but for the same value of the GB–SB angle. As shown in Figs. 9 and 10, the agreement between analytical modelling and FE computations is satisfying for rather large ranges of SB characteristic sizes. The remote stress values differ as well, which shows the validity of the closed-form expressions (Eqs. 4 and 5), whatever the remote stress, close or much higher than 2 τ_0. The considered SB sizes, L and t, vary by factors of 15 and 5, respectively. The stress values computed either by the FE method or the analytical formulae differ by $<7\,\%$.

In case of non-negligible hardening slope, the driving shear stress, $T = f\Sigma_0 - \tau_0$, should be replaced by: $T = f\Sigma_0 - \tau_c$ with $\tau_c = H_0\,\gamma^p$ and γ^p the mean slip band plastic slip which may be esaily computed depending on the SB aspect ratio, L/t, SB orientation and remote stress (Sauzay and Gilormini). Equations (4) and (5) may be modified slightly to take into account non negligible SB hardening. More details about the effect of SB plasticity parameters may be found in (Sauzay and Vor 2013).

It should be noticed that the inverse approach may be used as FE computations are not suitable: analytical modelling may be the direct way of solving such mechanical problems (Zavattieri et al. 2007a,b).

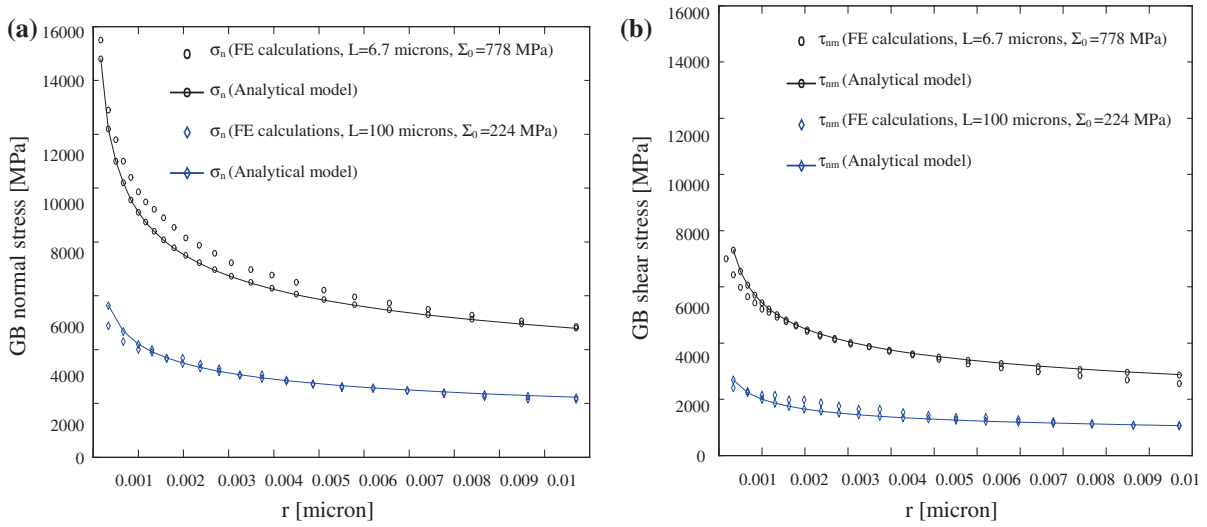

Fig. 9 Validation of the closed-form expressions (Eqs. 4 and 5) by comparing the predicted stress fields with the ones computed by the FE method. **a** GB normal stress fields; **b** GB shear stress fields ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $t = 0.09\,\mu m$ $\tau_0 = 60\,MPa$, $H_0 = 1\,MPa$, $q = 10^5$). SB lengths, L: 6.7 and 100 microns, and various remote tensile stresses, Σ_0

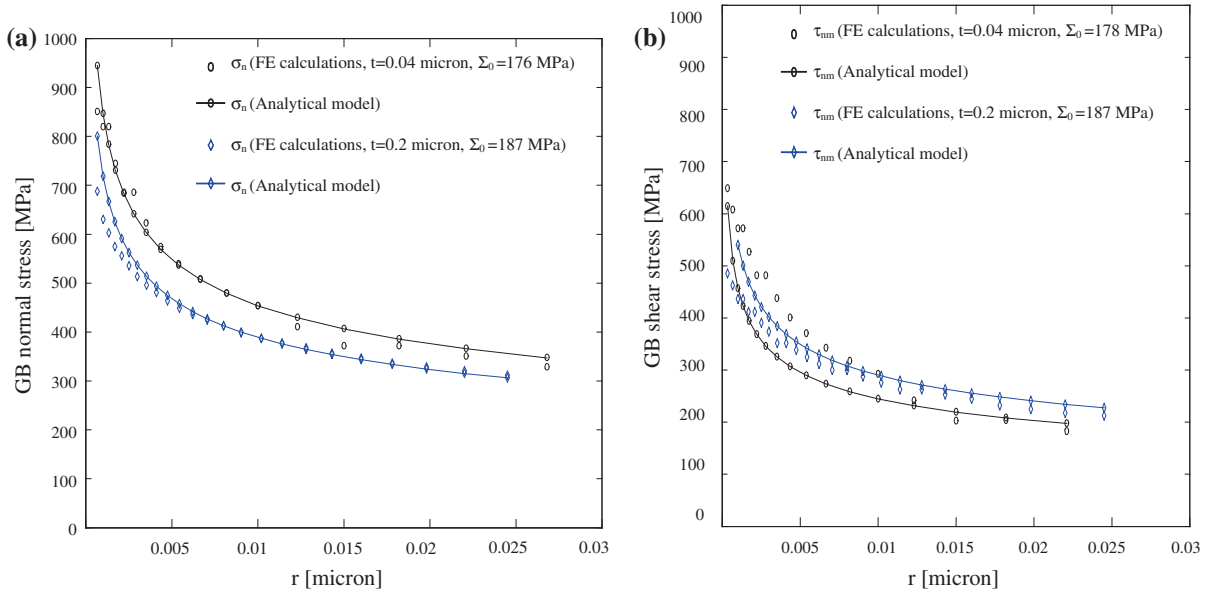

Fig. 10 Validation of the closed-form expressions (Eqs. 4 and 5) by comparing the predicted stress fields with the ones computed by the FE method. **a** GB normal stress fields; **b** GB shear stress fields ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $L = 10.7\,\mu m$, $\tau_0 = 60\,MPa$, $H_0 = 1\,MPa$, $q = 10^5$). SB thickness: 0.04 and 0.2 micron, and various remote tensile stresses, Σ_0

4 Influence of elasticity and plasticity parameters

4.1 Influence of isotropic elasticity parameters

First, various FCC metals and alloys have been considered: aluminium, copper and austenitic stainless steels. As shown in Fig. 11a, the GB normal and shear stress fields are not affected by the elasticity parameter values. As mentioned earlier, both GB normal and shear stress values are close even if the GB normal stress is a bit higher than the shear stress for the considered GB and SB orientations. The influence of the Poisson

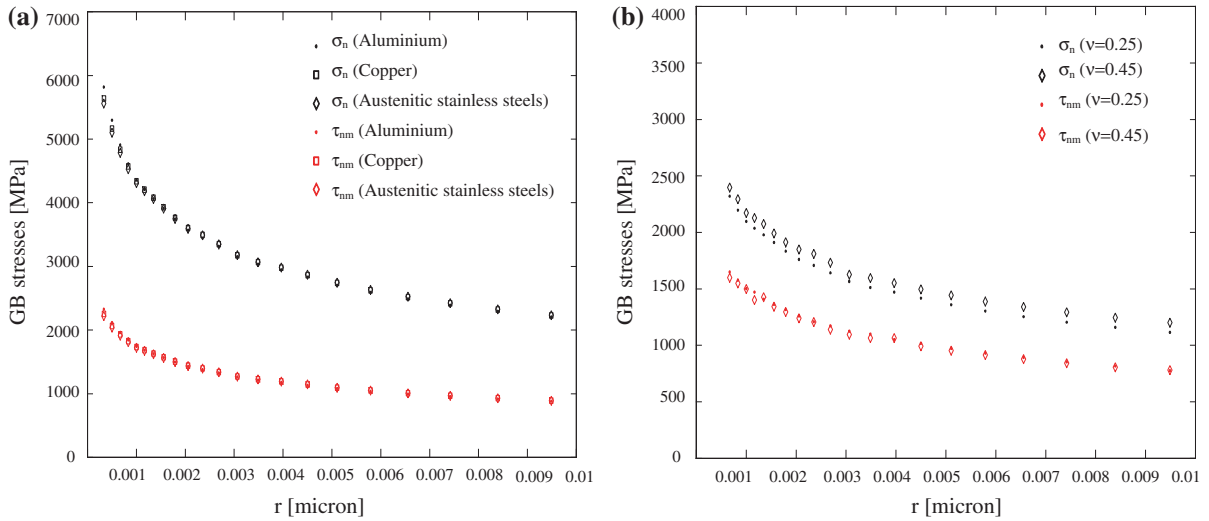

Fig. 11 Effect of isotropic elasticity parameters on the GB stress fields: **a** various Young's moduli but very close Poisson ratios (austenitic stainless steel, copper, aluminium); **b** Various Poisson ratio values ($Y = 200$ GPa) ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $L = 10.7\,\mu m$, $t = 0.09\,\mu m$, $\tau_0 = 60$ MPa, $H_0 = 1$ MPa, $q = 10^5$, $\Sigma_0 = 393$ MPa)

ratio has been studied as well. The same value of the Young's modulus was used but various values of the Poisson ratio have been used between 0.25 and 0.45. Once more, it can be noticed that the value of the Poisson ratio does affect only slightly the GB stress fields (Fig. 11b). Only the GB normal stress is modified by about 3 %. On the whole, these results are similar to the ones obtained by the LEFM and pile-up theories. Therefore, the analytical formulae proposed previously (Eqs. 4 and 5) are valid whatever the values of isotropic elasticity parameters.

4.2 Influence of cubic elasticity parameters

On the contrary, assuming that the considered grain obeys crystalline elasticity rather than isotropic elasticity may lead to non negligible differences as shown in Fig. 11a, b. For copper and austenite ($a \approx 3.3$), taking into account cubic elasticity rather than isotropic one, leads to GB normal and resolved shear stresses reduced by -33 and -23 %. This may be explained by the low Young's and shear stress moduli of well-oriented grains if $a > 1$ (Sauzay and Jourdan 2006; Sauzay 2007). Such elasticity anisotropy leads to a decrease of the shear stress magnitude in well-oriented grains. Therefore the driving shear stress is no more directly linked to the remote shear stress, but rather to the mean grain shear stress which is much lower (Sauzay and Jourdan 2006; Sauzay 2007). The Schmid factor $f = 0.5$ should be replaced by the effective Schmid factor, f_{eff}, which is equal to 0.43 in the case of austenite or copper ($a \approx 3.3$) and a well-oriented grain obeying crystalline elasticity embedded in a homogeneous matrix obeying isotropic elasticity (Sauzay and Jourdan 2006; Sauzay 2007). The effective Schmid factor value is defined by the ratio between the mean grain resolved shear stress computed by the FE method, and the remote tensile stress, Σ_0. The effective driving shear stress may be defined by: $T_{eff} = f_{eff}\,\Sigma_{0-\tau_0}$. For $\Sigma_0 = 393$ MPa and $\tau_0 = 60$ MPa (Fig. 12a, b), this effective driving force is 20 % lower than the usual driving shear stress, $T = f\Sigma_0 - \tau_c$, computed for isotropic elasticity. Therefore, this simple analysis explains qualitatively the effect of crystalline elasticity on GB stress field. In fact, the slip band and grain sizes are similar and therefore the slip band may not be considered as small with respect to the grain, which shows that the previous analysis is not strictly valid. The effect of crystalline elasticity should depend on the anisotropy coefficient, a, grain orientation, SB aspect ratio Schmid factor, f, remote tensile stress, Σ_0, and critical shear stress, τ_c. And the use of the effective Schmid factor shows qualitatively what the effect of crystalline elasticity is, depending on the previous parameters.

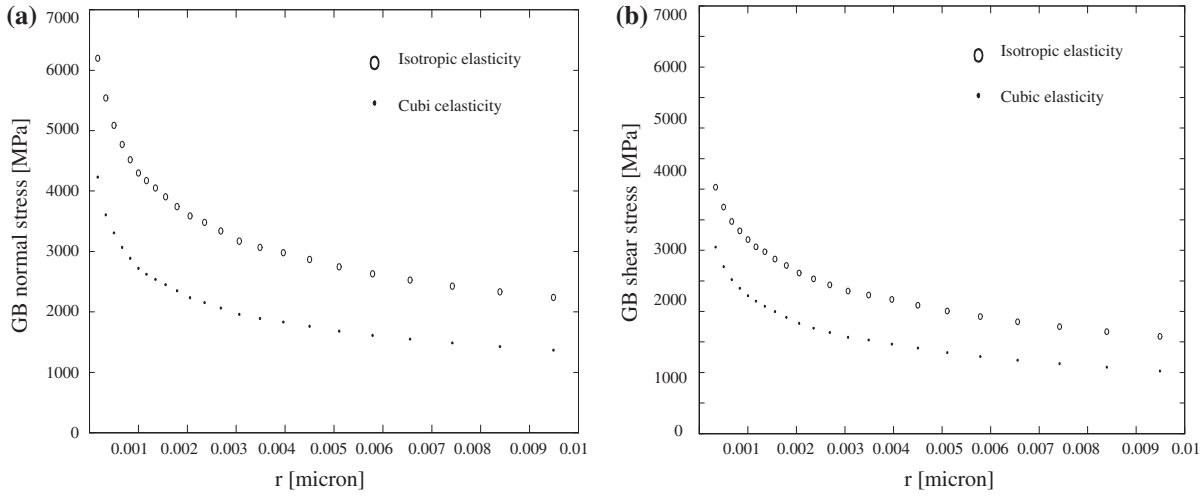

Fig. 12 Effect of crystalline elasticity on the GB stress fields: **a** normal stress; **b** shear stress ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $\tau_0 = 60\,\text{MPa}$, $H_0 = 1\,\text{MPa}$, $q = 10^5$, $L = 10.7\,\mu\text{m}$, $t = 0.09\,\mu\text{m}$, $\Sigma_0 = 393\,\text{MPa}$). The material is austenitic stainless steel ($a = 3.3$)

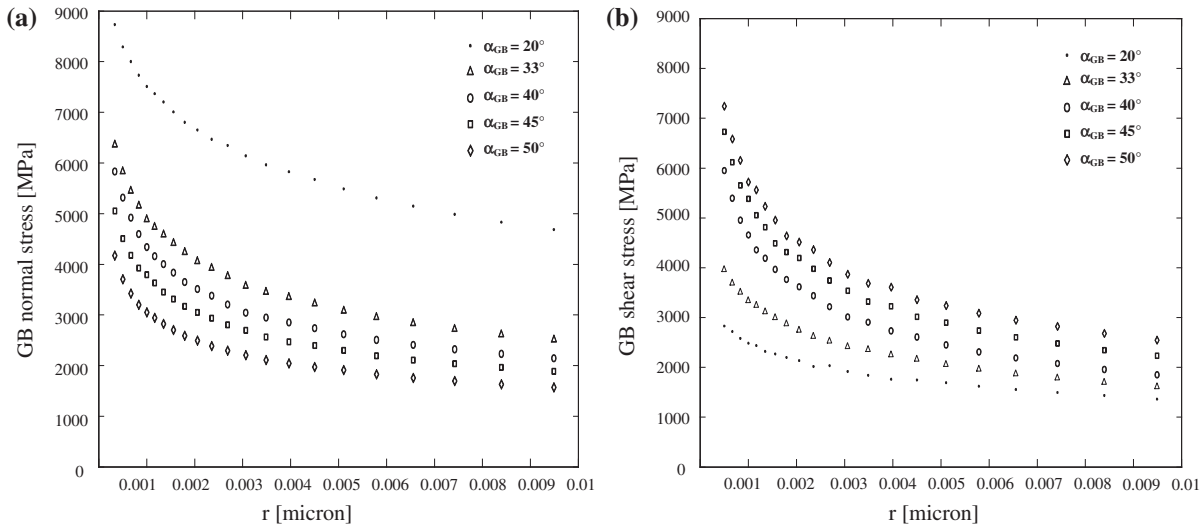

Fig. 13 Effect of the GB orientation on the GB stress fields (Fig. 1): **a** normal stress; **b** shear stress ($\alpha_{SB} = 45°, L = 10.7\,\mu\text{m}, t = 0.09\,\mu\text{m}, \tau_0 = 60\,\text{MPa}, H_0 = 1\,\text{MPa}, q = 10^5, \Sigma_0 = 393\,\text{MPa}$)

5 Effect of random microstructure

5.1 Influence of grain boundary orientation

Grain boundary angles, α_{GB}, between 20° and 50° have been considered while the SB angle is kept constant: $\alpha_{SB} = 45°$ (Fig. 1). The GB normal and shear stress fields are plotted in Fig. 13a, b, respectively. The lower the GB angle, the higher the GB normal stress. The shear stress behaves in the opposite way. The effect is rather strong as the local stress is affected by more than a factor 2 for the angle α_{GB} varying between 20° and 50°. For each value of the α_{SGB} angle, the corresponding stress singularity parameters, A_{nn}, A_{nm}, and $(1 - \lambda)$ have been deduced from one GB normal stress curve and the corresponding shear stress one (Eqs. 4 and 5, see Fig. 7b). The singularity parameters are plotted with respect to the SB–GB angle, $\theta = 90_{-\alpha_{SB}+\alpha_{GB}}$,

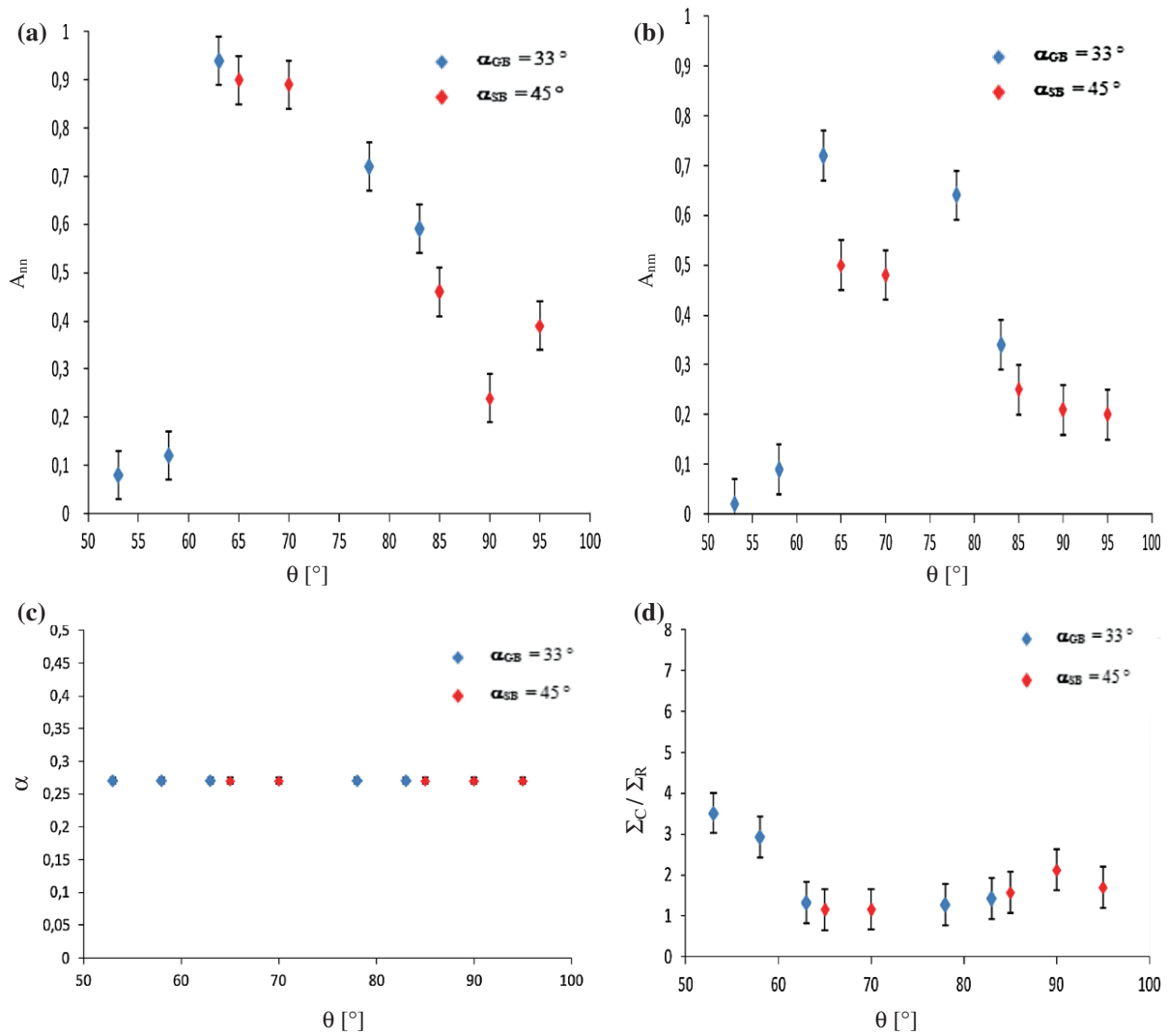

Fig. 14 Plot of the singularity parameters (Eqs. 4 and 5) depending on the GB angle, α_{GB}, and SB angle, α_{GB} (Fig. 1): **a** A_{nn}; **b** A_{nm}; **c** $(1-\lambda)$ ($H_0 = 1$ MPa, $q = 10^5$, $L = 10.7\,\mu$m, $t = 0.09\,\mu$m, $\Sigma_0 = 393$ MPa) and **d** plot of the normalized remote tensile stress to GB microcrack initiation ($\tau_0 = 0$ MPa, see Sect. 6.4). The SB–GB angle, $\theta = 90_{-\alpha_{SB}+\alpha_{GB}}$ is used for the plotting (Fig. 1)

in Fig. 14a–c. First of all, the exponent, $(1-\lambda)$, is not affected by the value of the α_{SGB} angle, at least in the studied range of variation (Fig. 14c). It is about 0.27 for the whole range of variation of α_{GB}, that is only about one-half of the pile-up exponent, 0.5. The SB singularity is therefore much weaker than the pile-up one. This means that the same singularity exponent holds for the whole considered range of GB orientations whereas the singularity pre-factors are much more affected as shown in Fig. 14a, b. It should be noticed that in the pile-up theory, the pre-factor of the normal stress singularity is much less affected by change in α_{GB} as it changes by less than $\pm 13\,\%$, provided θ belongs to a rather broad range of variation: 25°–100° (Smith and Barnby 1967; Evrard and Sauzay 2010).

5.2 Influence of crystallographic orientation

The SB angle, α_{SB}, defined in Fig. 1, varies between 35° and 70° for a constant GB angle of $\alpha_{GB} = 33°$. The GB normal and shear stress fields are plotted in Fig. 15a, b, respectively. As expected, the higher

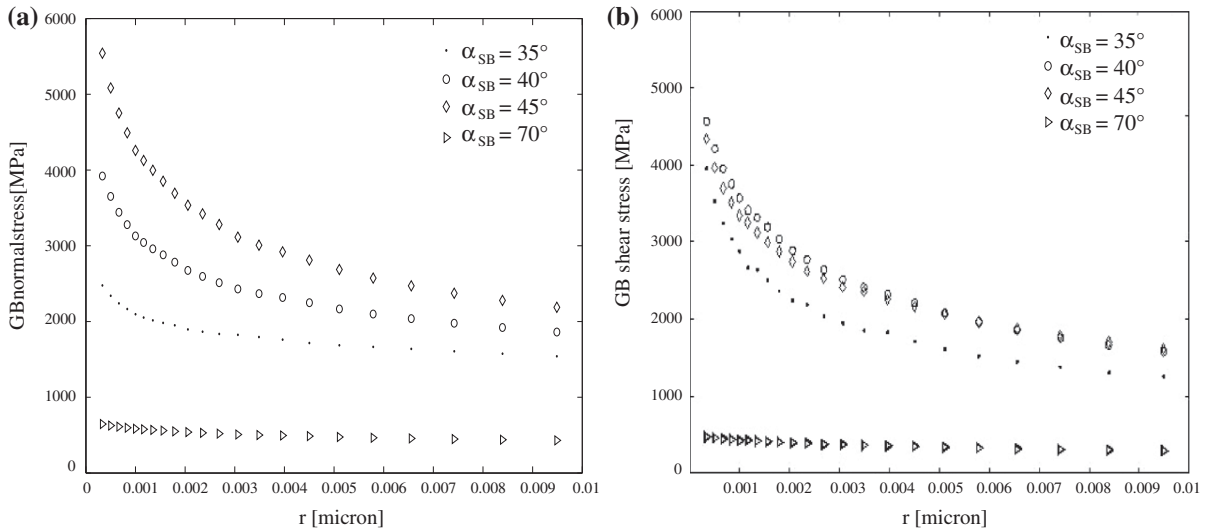

Fig. 15 Effect of the SB orientation on the GB stress fields (Fig. 1): **a** normal stress; **b** shear stress ($\alpha_{GB} = 33°$, $\tau_0 = 60$ MPa, $H_0 = 1$ MPa, $q = 10^5$, $L = 10.7\,\mu$m, $t = 0.04\,\mu$m, $\Sigma_0 = 393$ MPa)

the Schmid factor, the higher the GB stress fields. The Schmid factor value can be computed as follows: $f(\alpha_{SB}) = \cos(\alpha_{SB})\sin(\alpha_{SB})$. It reaches its maximum for $\alpha_{SB} = 45°$, which leads to the maximum remote shear stress, $T(\alpha_{SB}) = f(\alpha_{SB})\Sigma_0 - \tau_0$. But, the SB angle affects not only the driving shear stress, $T(\alpha_{SB})$, but also the geometry coefficients, A_{nn} and A_{nm}, as shown in Fig. 14a, b. More precisely, as the SB angle becomes far from 45°, the geometry coefficients, A_{nn} and A_{nm} (Fig. 14a, b), drop down whereas the stress singularity exponent is still stable with respect to α_{SB} (Fig. 14c). As the angle becomes low, the ratio between the GB shear and normal stresses increases (Fig. 15a, b). For the considered GB and crystal orientation, the ratio varies between 1/2 and 2 which shows that both stress components have the same order of magnitude.

As expected, the SB–GB angle, θ, allows the plot of a kind of master curve for each singularity parameter. The maximum prefactor is reached for θ values between 60° and 80°, which is very close to the angle value predicted by the pile-up theory, 70.5° (Smith and Barnby 1967; Evrard and Sauzay 2010). It should be noticed that the shear stress singularity pre-factor, A_{nm}, does not vanish as θ is close to 70.5°, contrary to the prediction of the pile-up theory (Smith and Barnby 1967). Finally, the dependence of the pre-factor, A_{nn}, with respect to the θ angle, predicted by the pile-up theory differs quantitatively with the one found considering SBs of finite thickness, $t > 0$. A much smoother dependence evolution is indeed predicted by the pile-up theory (Smith and Barnby 1967; Evrard and Sauzay 2010).

6 Prediction of grain boundary microcrack initiation

6.1 Finite fracture mechanics

As mentioned in Sect. 1, the weak stress singularity ($1 - \lambda \approx 0.27 < 0.5$) does not allow us to use the Griffith criterion which corresponds to an energy balance equation for an infinitesimal crack nucleus. On the contrary, the use of finite fracture mechanics (Leguillon 2002; Taylor et al. 2005) (or quantized fracture mechanics (Pugno and Ruoff 2004)) involving a finite crack nucleus, $a_c > 0$, allows us to predict GB microcrack initiation. Nevertheless an additional equation is needed because two unknown values should be computed, on one hand the critical remote tensile stress for GB microcrack initiation, $\Sigma_c > 0$, and on the other hand the microcrack nucleus length, $a_c > 0$. Following the physics of brittle fracture, a stress criterion seems to be suitable. A local critical stress, σ_c, should indeed be reached following atomistic analytical computations (Averbach 1968; Rose et al. 1981; Rice and Wang 1993) or more recent molecular dynamics ones (Yamakov et al. 2006). Cohesive zone modelling

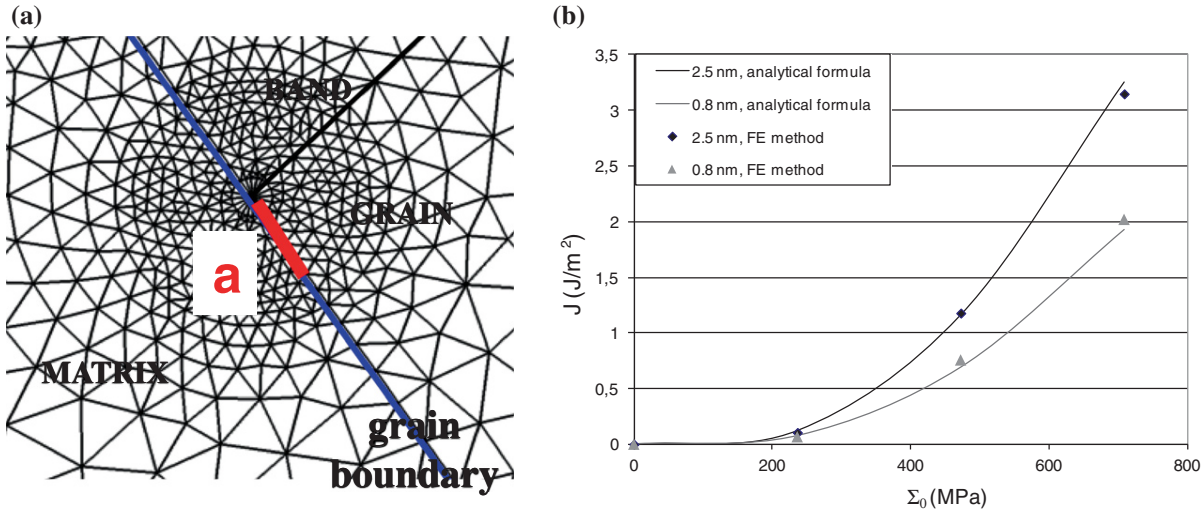

Fig. 16 **a** Mesh of the area around the grain boundary close to the slip band corner. Cracks of various lengths, a, have been introduced in the mesh in order to compute the J integral using the G θ method (Cast3m 2012). Geometry parameters: L = 10 μm, t = 0.1 μm and a = 0.75, 1, 1.5, 2.5, 3 nm; **b** dependence of the J integral value with respect to the remote stress, Σ_0, and crack length, a

(CZM) of brittle fracture may be directly based on the atomic force—separation curves (Yamakov et al. 2006). The same two main parameters are used by CZM as well: local critical stress, σ_c, and fracture energy, γ_{fract}. And the comparison between the crack initiation predictions based on either CZM or finite fracture mechanics using both stress and energy criteria leads to similar results (Murer and Leguillon 2010; Henninger et al. 2007) provided the fracture parameters are the same.

Some directs observations of the fracture of oxidized GBs during in-situ testing show that fracture energies as low as a few J m^{-2} may be required (Dugdale et al. 2013). This seems to correspond to brittle fracture. In case of slip localization, as the matrix is almost elastically deformed because of dislocation source exhaustion and/or strong pinning effect induced by irradiation defects, precipitates,… at least some of the GBs are fractured without plastic blunting, at least during the initiation stage. And these correspond to the preferential GB fracture sites because of their low fracture energies. In some cases plasticity may occur in the vicinity of GBs or small GBs cracks because of stress concentrations. Then crack tip blunting and much higher dissipated energy should be taken into account (Zavattieri et al. 2007b).

6.2 Prediction of grain boundary microcrack initiation

The stress criterion can be written simply as follows:

$$| \sigma_n (r) | = \sigma_c \quad \text{if} \quad r < a_c \tag{6}$$

As the slip band obeys elastoplasticity, the J integral is computed instead of the energy release rate, G. The J integral is computed using the FE method after introducing GB crack of different lengths in some of the meshes used previously plotted (Fig. 2b). For instance, in the mesh built using L = 10 μm and t = 0.1 μm, short cracks of lengths lengths of a = 0.75/1/1.5/2.5 and 3 nm have been introduced in five different meshes (Fig. 16a). The J integral has been computed depending on both crack length, a, and driving shear stress, T. The computation of J is based on the G θ method (Cast3m 2012). The dependence of J with respect to the various geometry and loading parameters is the following:

– it is proportional to the square of the remote driving shear stress, T, as may be guessed from Fig. 16b and from the theory of matching expansions (Leguillon 2002);
– a power law dependence with respect to the crack length, a, is found with an exponent of (2λ − 1)

as expected once more from the theory of matching expansions (Leguillon 2002).

– it is finally expected from the asymptotic expansion of the GB stress fields close to the slip band corner (Eqs. 4 and 5) that the dependence of the J integral with respect to the slip band thickness and length, t and L should be the following:

$$J\,(a,\, f\,\Sigma_0 - \tau_0) = C\,(f\,\Sigma_0 - \tau_0)^2\,L\left(\frac{a}{t}\right)^{2\lambda-1} \quad (7)$$

with C depending on the isotropic elasticity coefficients only. It is found to be about: $2.3\times10^{-11}\ \mathrm{Pa}^{-1}$. Using the Young's modulus value, Y = 180 GPa, an adimensional parameter is obtained: YC ≈ 3.69 for $\alpha_{GB} = 33°$ and $\alpha_{SB} = 45°$.

The opposite of the variation in mechanical energy is computed by integration of J between a = 0 and a = a_c (Pugno and Ruoff 2004). Then the energy balance equation is applied, using the GB fracture energy. The combination of the stress and energy criteria allows us to evaluate both critical remote stress, Σ_c, and crack nucleus length, a_c:

$$f\,\Sigma_c - \tau_0 = \left(\frac{2\lambda}{C}\right)^{1-\lambda} A_{nn}^{1-2\lambda} \left(\frac{t}{L}\right)^{1/2} \left(\frac{\gamma_{fract}}{t}\right)^{1-\lambda} \sigma_c^{2\lambda-1} \quad (8)$$

and then:

$$\Sigma_c = (1/f)\left(\tau_0 + \left(\frac{2\lambda}{C}\right)^{1-\lambda} A_{nn}^{1-2\lambda} \left(\frac{t}{L}\right)^{1/2} \times \left(\frac{\gamma_{fract}}{t}\right)^{1-\lambda} \sigma_c^{2\lambda-1}\right) \quad (9)$$

Whatever the hardening slope value, the critical crack length is given by

$$a_c = 2\lambda\left(\frac{A_{nn}^2}{C}\right)\left(\frac{\gamma_{fract}}{\sigma_c^2}\right) \quad (10)$$

Physical theories of brittle fracture of metals allow us to link the local critical stress, σ_c, to the fracture energy, γ_{fract}. For instance, Stroh assumed that the Hooke elasticity law is valid up to fracture (Stroh 1957; Averbach 1968). He computed the mechanical work up to fracture and deduced the following equation (Averbach 1968):

$$\sigma_c = \sqrt{\frac{Y\gamma_{fract}}{d_0}} \quad (11)$$

With Y the Young's modulus and d_0 the distance between two neighbour close-packed atomic planes. Some more accurate formulae have been proposed (Averbach 1968; Rose et al. 1981; Rice and Wang 1993) but they differ from the previous one (Eq. 11) only by less than a factor two. The Orowan theory assumes that the stress dependence with respect to the inter-plane distance may be described by a sinusoïdal law which leads to a microscopic brittle fracture stress, σ_c, $\sqrt{2}$ lower than the Stroh one (Eq. 11). The universal bounding theory of Rose et al. is the most physically-based and leads to a microscopic fracture stress, σ_c, $e/\sqrt{2}$ lower than the one given by Eq. (11). Following Eq. (10), and using the universal bounding theory as well as the highest values of A_{nn}, evaluated in Sect. 4 leads to: $a_c \sim 1.5d_0$. This length is low but nevertheless meaningful. In addition, it is much smaller than t which is required for using the above stress close-field expansion (Eq. 4).

Using Eq. (9) and for instance the Orowan equation, the dependence of the critical remote stress with respect to the GB fracture energy, γ_{fract}, is deduced:

$$\Sigma_c = (1/f)\left(\tau_0 + \left(\frac{2\lambda}{YC}\right)^{1-\lambda} \times A_{nn}^{1-2\lambda}\left(\frac{t}{L}\right)^{1/2}\left(\frac{d_0}{t}\right)^{1-\lambda}\sqrt{\frac{Y\gamma_{fract}}{2d_0}}\right) \quad (12)$$

still in case of negligible slip band hardening.

Or equivalently with respect to the local critical stress, σ_c:

$$\Sigma_c = (1/f)\left(\tau_0 + \left(\frac{2\lambda}{YC}\right)^{1-\lambda} A_{nn}^{1-2\lambda}\left(\frac{t}{L}\right)^{1/2}\left(\frac{d_0}{t}\right)^{1-\lambda}\sigma_c\right) \quad (13)$$

The final expression of the length of the critical crack nucleus is rather simple:

$$a_c = 4\lambda\left(\frac{A_{nn}^2}{YC}\right)d_0 \quad (14)$$

The effect of the hardening slope, H_0, may be taken into account as well as shown by Sauzay and Vor (2013).

6.3 Comparison with the predictions of the pile-up theory

As mentioned earlier, the Griffith criterion can be applied directly because of the stress singularity exponent of 1/2 induced by dislocation pile-ups. The energy balance equation may be written as follows:

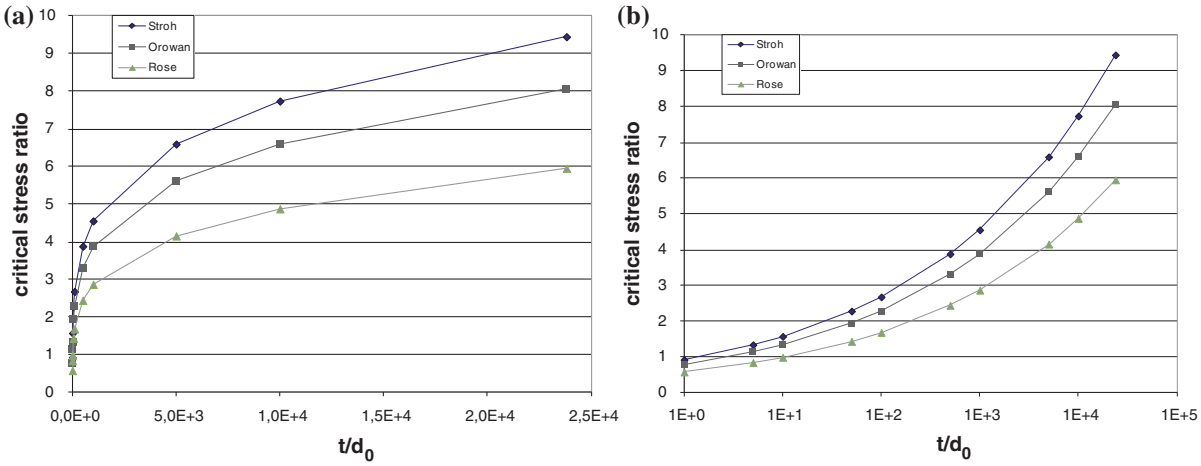

Fig. 17 Ratio between the remote tensile stress to GB microcrack initiation computed using either Eq. (12) (SBs of finite thickness, $t > 0$) or the pile-up theory (Eqs. 16, 17, see Sect. 6.3). Dependence of microscopic tensile, σ_c, with respect to the fracture surface energy, γ_{fract}: Stroh theory, Orowan theory or Rose et al. universal bounding correlation; **a** linear scale and **b** logarithm ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $\tau_0 = 0$ MPa, $\upsilon = 0.3$, $d_0 = 0.21$ nm)

$$G = \gamma_{fract} = 2\gamma_{s-\gamma_{GB}} \tag{15}$$

The energy release rate, G, can be computed using the mode I and mode II stress intensity factors and the elasticity constants (Smith and Barnby 1967). The critical remote tensile stress leading to GB microcrack initiation is easily deduced (Smith and Barnby 1967):

$$\Sigma_c^{\text{pile-up}} = \frac{1}{f}\left[\sqrt{\frac{\pi\mu\gamma_{fract}}{4(1-\upsilon)L^{\text{pile-up}}}}\frac{1}{(F(\theta))^{1/2}} + \tau_0\right] \tag{16}$$

with:

$F(\theta) = (1/4)\left(5 + 2\cos(\theta) - 3\cos^2(\theta)\right)$,

$\theta = 45° + \alpha_{GB} - \alpha_{SB}$ (Fig. 1) and $L_{\text{pile-up}} \approx L/2$.

It should be noticed that the application of both energy and stress criteria to the pile-up stress singularity leads to the same expression, as already noticed by Leguillon (2002) in case of a sharp crack which singularity exponent is 1/2. Using the Orowan computation of the microscopic brittle fracture stress, the ratio between the critical remote stresses computed either assuming a finite thickness, $t > 0$, and assuming the formation of a pile-up is easily deduced from Eqs. (12) and (16) by neglecting τ_0:

$$\frac{\Sigma_c}{\Sigma_c^{\text{pile-up}}} = \left(\frac{2\lambda}{YC}\right)^{1-\lambda} A_{nn}^{1-2\lambda}\left(\frac{t}{d_0}\right)^{\lambda-\frac{1}{2}} \times \left(\frac{2\left(1-\upsilon^2\right)}{\pi}\right)^{\frac{1}{2}} (F(\theta))^{\frac{1}{2}} \approx \left(\frac{t}{d_0}\right)^{\lambda-\frac{1}{2}} \tag{17}$$

It is proportional to $(t/d_0)^{\lambda-1/2}$ with $\lambda - 1/2 \approx 0.23$. The higher the slip band thickness, the larger the difference between the critical remote tensile stresses predicted by the FE method (finite thickness) and the pile-up theory. The ratio is plotted with respect to the normalized SB thickness using the three previous atomistic theories of brittle fracture (Fig. 17a, b). The trends are very similar. The exponent value, $\lambda - 1/2$, is rather weak but the effect of slip band thickness is nevertheless large provided it belongs to its physically-based range of variation (from a few ten nm to a few hundred nm) as shown in Fig. 16a, b. Considering the universal bonding correlation (Rose et al. 1981), and using the previous numerical evaluations of λ, A_{nn} and YC ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$), as well as $b = 0.25$ nm ($d_0 = 0.21$nm) and $\upsilon = 0.3$, a ratio equal to 1.6 is found for $t = 50$ nm (channels in FCC metals and alloys subjected to tensile loading) and about 2.9 for $t = 500$ nm (FCC metals and alloys subjected to cyclic loading). Therefore, the finite thickness of slip bands affects strongly the values of the critical remote tensile stresses and the pile-up theory leads to large underestimations whatever the material and loading condition considered here. It should be noticed that the relation ship between the microscopic critical stress, σ_c, and the fracture energy, γ_{fract}, affects the results: the critical remote stress, Σ_c, computed using the universal bounding correlation is about 30 % lower than the one computed using the crude Stroh theory (Fig. 17a, b).

As it is more physically-based, the universal bounding correlation is used in the following.

6.4 Influence of GB and crystal orientations on microcrack initiation

Grain boundary and crystal orientations affects the remote stress to GB microcrack initiation by at least two ways (Eq. 12):

- the Schmid factor, f, depends on the SB angle, α_{SB}, following: $f = \cos(\alpha_{SB})\sin(\alpha_{SB})$ in the case of uniaxial tension;
- the singularity parameters, A_{nn} and YC, depend on the angle $\theta = 90_{-\alpha_{SB}+\alpha_{GB}}$ as shown in Sect. 5, but the singularity exponent, $(1-\lambda,)$, does not (Fig. 14c).

It has been checked that the YC value varies only slightly with respect to the GB and SB angles in the considered ranges of variation. The dependence of the critical remote stress, Σ_c, with respect to YC is a power law one. Its exponent, $(1-\lambda) = 0.27$, is low, then the influence of this geometry parameter, YC, is weak. For the sake of clarity, the normalized critical remote stress, Σ_c/Σ_R is plotted with: $\Sigma_R = (t/L)^{0.5}(d_0/t)^{0.5}\ \sigma_c$ and τ_0 is assumed to be negligible with respect to Σ_c (Eq. 13).

The dependence of Σ_c with respect to the θ angle has been plotted in Fig. 14d for the whole range of GB and SB orientations considered in this work. It reaches its minimum for θ belonging to the 60°–80° range, which agrees with pile-up theory predictions. Nevertheless, the pile-up theory predicts a very smooth variation of the remote shear stress to GB microcrack initiation in the range of variation $\theta = 0° - 180°$ [see Fig. 4 in (Smith and Barnby 1967)]. But the modelling including the effect of SB finite thickness predicts a stronger variation, even without taking into account the dependence of the Schmid factor with respect to the SB angle. The remote critical stress increases strongly for high α_{SB} values (65°–70°) because of the low Schmid factor values as well as the low A_{nn} prefactor values (Fig. 14a).

It should be noticed that GBs vary in terms of orientation, as well as SBs. Grain sizes varies rather strongly from one grain to another [see in Fig. 1a (Sauzay and Man 2008)] and SB thickness as well but not so much (Byun et al. 2006). All these variations in microstructure features affect the value of the predicted remote stress to GB fracture (Eq. 16). This explains at least qualitatively numerous experimental observations showing the random character of GB cracking as well as continuous microcrack initiation during tensile loading (McMurtrey and Was 2011). Intrinsic characteristics of GBs play a role as well (Eq. 15) with general GBs of high GB energy whereas the ones of special GBs are much lower (Sutton and Balluffi 1995; Caul et al. 1997; McMurtrey and Was 2011).

Finally, the modeling of interface fracture and induced crack propagation by CZM (Cordisco et al. 2012; Antico et al. 2012) does not take into account microstructure details such as slip bands and the mesh size may not be too low for saving computation time and memory. Therefore, the parameters of the CZM used in such larger scale computations should be modified for taking into account only implicitly both stress concentrations induced by slip localization and physically-based fracture energy values such the ones obtained from DFT computations (see Sect. 7). It may be suitable to adjust the CZM parameters of such large scale computations by inverse identification. They would be adjusted for allowing the prediction of GB fracture initiation at remote stresses comparable to the ones predicted using the micromechanical modeling including slip bands. Of course the obtained fracture stress and energy would be lower than the one used in the micromechanical modeling because of the lack of stress concentrators in the large scale CZM computation.

7 Application to GB microcrack initiation in pre-irradiated austenitic stainless steels

7.1 Tensile deformation in inert environment

Nishioka et al. (2008) and Fukuya et al. (2008) tested in argon environment austenitic stainless steels previously pre-irradiated up to a neutron dose equivalent to 35 displacements per atom (dpa) at a temperature of 320 °C. The tensile tests have been carried out at very slow strain rate ($\approx 10^{-8}s^{-1}$) and at a temperature of about 320 °C. In these experimental conditions, dislocation channels are observed at the surface of the specimens as well as close to GBs. Both authors measured the macroscopic tensile stress to GB microcrack initiation which is found to be close to the conventional yield stress: i.e $\Sigma_c/\Sigma_y \approx 1$ (Fig. 18). Furthermore, Nishioka

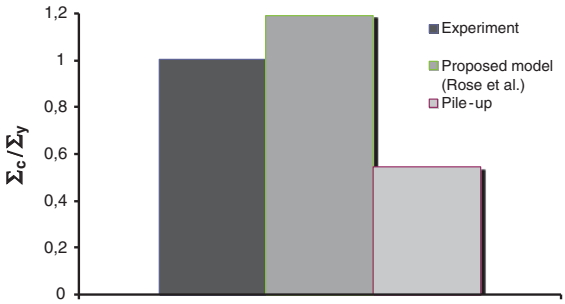

Fig. 18 Comparison between the predictions of the normalized remote tensile stress leading to GB microcrack nucleation in pre-irradiated austenitic stainless steels and experimental data (Nishioka et al. 2008; Fukuya et al. 2008). The predictions based on either Eq. (12) (slip band of finite thickness, t) assuming universal bounding correlation or the pile-up theory (Eq. 16) are shown. Inert environment ($\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $\tau_0 = 60$ MPa, $q = 10^5$, $H_0 = 1$ MPa, $Y = 180$ GPa, $\upsilon = 0.3$, $L = 50\,\mu$m, $t = 50$ nm). Surface energy: $\gamma_s = 2.5\,\mathrm{J m^{-2}}$ and grain boundary energy $\gamma_{GB} = 1.2\,\mathrm{J\,m^{-2}}$ (general grain boundary) (see Sect. 7.1)

et al. observed no influence of the irradiation dose on the macroscopic microcrack initiation stress for doses higher than 4dpa (Nishioka et al. 2008).

In the following, the predictions of Smith and Barnby (Eq. 16) and the proposed criterion (Eq. 12) are compared to the previous experimental results. The grain size and the channel thickness of austenitic stainless steels studied by Nishioka et al. and Fukuya et al. are about $L = 50\,\mu$m and $t = 50$ nm, respectively. Nogaret et al. (2008) simulated by discrete dislocation dynamics the formation of a channel during tensile loading of a pre-irradiated austenite crystal. They found a channel shear stress, τ_0, of a few ten MPa which is indeed negligible with respect to the remote stress, Σ_0. Channels in irradiated FCC metals and alloys usually display almost no hardening (Sharp 1967; Nogaret et al. 2008), that is why $H_0 \sim 0$. The free surface energy for an austenitic stainless steel belongs to the range: $\gamma_s = 2{-}3\,\mathrm{J\,m^{-2}}$ (Vitos et al. 1998) and this one depends only slightly on the considered crystallographic plane. The grain boundary energy, γ_{GB}, for an austenitic stainless steel, without strong segregation at GBs, has been reported by Caul et al. (1996) to belong to a rather broad range: $\gamma_{GB} = 0.3 - 1.2\,\mathrm{J\,m^{-2}}$. These values depend strongly of the GB characteristics (LABs, special GBs or general GBs, see Sect. 1). In order to apply both criteria to predict microcrack initiation in argon environment, an average free surface energy is used, $2.5\,\mathrm{J\,m^{-2}}$ (Vitos et al. 1998). Concerning, the GB energy value, general GBs should be the most sensitive to microcrack initiation following the definition of the fracture energy: $\gamma_{fract} = 2\gamma_s - \gamma_{GB}$, which is confirmed by experiments (Dropek et al. 2004). That is why the highest GB energy has been considered, $\gamma_{GB} = 1.2\,\mathrm{J\,m^{-2}}$ (Caul et al. 1996). The yield stress is considered to be equal to $\Sigma_y = 1{,}000$ MPa (Jiao et al. 2005). The results plotted in Fig. 18 show that the Smith and Barnby criterion, based on the Stroh pile-up model, underestimates strongly the microcrack initiation stress. This result confirms that pile-up models are not suitable for predicting intergranular microcrack initiation in pre-irradiated metallic polycrystals. On the contrary, the proposed criterion, using the universal bounding correlation, is able to correctly predict the remote stress to GB microcrack initiation. It should be noticed that no adjustable parameter was used as applying the proposed criterion (Eq. 12, modified to take into account the universal bounding correlation).

7.2 Tensile deformation in water environment

Recently, Nishioka et al. (2008) and Takakura et al. (2007) carried out uniaxial constant load tests on austenitic stainless steels previously pre-irradiated up to 38dpa. The tests have been carried out in pressurized water reactor (PWR) environment. They measured the stress at which failure occurred (named irradiation assisted stress corrosion cracking (IASCC) failure stress). In the experimental conditions (temperature of 300 °C and constant load), it is reasonable to assume that channelling occurred and contributed to intergranular microcrack initiation and failure. The experimental results (Fig. 19a, b) show that at 38dpa, the ratio Σ_{crit}/Σ_y is equal to about 0.5, while this ratio evaluated for the same irradiation dose is equal to about 1 in argon environment as mentioned previously. The role of oxidation on the lower value of the ratio Σ_{crit}/Σ_y for pre-irradiated material tested in PWR environment is now investigated theoretically.

Many authors measured the GB chromium concentration in pre-irradiated austenitic stainless steels (Asano et al. 1992; Bruemmer et al. 1999). They observed that the chromium concentration decreases when the irradiation dose increases. The minimum concentration reaches about 11 % at 10dpa. Furthermore, an enrichment in nickel is observed at GBs. The GB chemical composition variations are due to the Kirkendall inverse effect (Bruemmer et al. 1999; Was 2007).

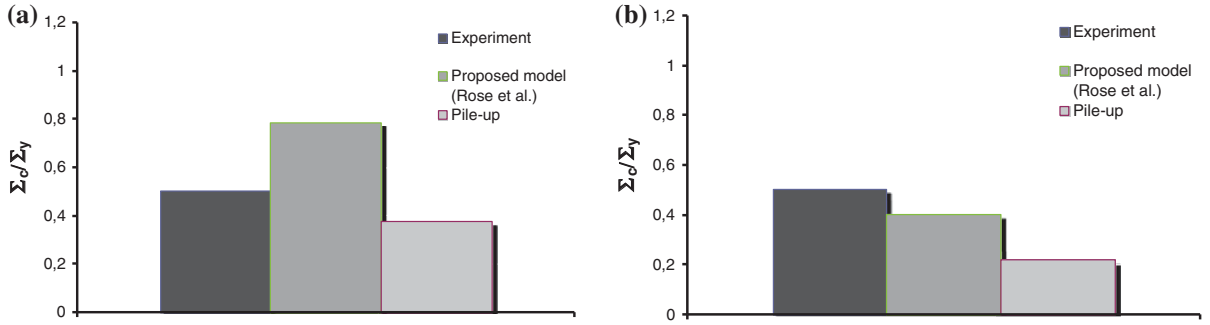

Fig. 19 Comparison between the predictions of the normalized remote tensile stress leading to GB microcrack nucleation in pre-irradiated austenitic stainless steels and experimental data (Nishioka et al. 2008; Takakura et al. 2007). The predictions based on either Eq. (12) (slip band of finite thickness, t) assuming universal bounding correlation or the pile-up theory (Eq. 16) are shown. Water environment: **a** hydrated surfaces and **b** hydroxylated (see Sect. 7.2). Parameters: $\alpha_{SB} = 45°$ and $\alpha_{GB} = 33°$, $\tau_0 = 60\,\text{MPa}$, $q = 10^5$, $H_0 = 1\,\text{MPa}$, $Y = 180\,\text{GPa}$, $\upsilon = 0.3$, $L = 50\,\mu\text{m}$, $t = 50\,\text{nm}$

It is probable that at the intersection between the free surface and the GB, the oxide layer can be broken due to both channelling and the depletion of chromium which does not allow the repassivation of the oxide layer. Oxygen can then diffuse inside the material. Oxide such as Fe_2O_3, Fe_3O_4, Cr_2O_3, NiO can be formed and embrittle GBs (Bruemmer and Thomas 2005). The fracture surface energy is also modified by the presence of oxide. At 38dpa, GB chromium content is too low for continuous repassivation and GBs can be oxidized. Free surface energies per unit area have been calculated for Fe_2O_3, Fe_3O_4, Cr_2O_3, or NiO oxides using ab-initio computations. The obtained values are about $2\gamma_s = 3-5\,\text{J m}^{-2}$ (Fisher 2004; Sun et al. 2006). De Leeuw and Cooper (2007) computed free surface energy, γ_s, for pure, hydrated and hydroxylated Fe_2O_3. They obtained respectively, $2\gamma_s = 4.2\,\text{J m}^{-2}$, $2\gamma_s = 2.9\,\text{J m}^{-2}$ and $2\gamma_s = 1.7\,\text{J m}^{-2}$. Only the two last values should be close to the one of iron oxide surfaces newly formed in water environment. Very quickly, hydratation and/or hydroxylation should indeed occur in water at the surface of the newly formed oxide. Unfortunately, we are not aware of any study allowing the computation or measurement of GB energies for these oxides. Nevertheless, ab-initio computations showed that in rutile TiO_2, the value of the energy of the $\Sigma = 1536.9°$ (210)[001] tilt boundary is $1.7\,\text{J m}^{-2}$ (Dawson et al. 2006). Molecular Dynamics computations showed that the $\Sigma = 5$ symmetrical tilt boundary has an energy of about $1\,\text{J m}^{-2}$ (Chartier et al. 2010). As mentioned in the introduction, general GBs are more prone to microcrack initiation (Alexandreanu and Was 2003; Tan et al. 2005). Consequently, a GB energy of $\gamma_{GB} = 1.2\,\text{J m}^{-2}$ is used which gives the right order of magnitude for general GBs for various materials (Sutton and Balluffi 1995). Using the values computed by De Leeuw and Cooper (2007), two applications of the proposed criterion as well as the Smith and Barnby one are performed using: (i) fracture surface energy of $2\gamma_s = 2.9\,\text{J m}^{-2}$ minus a GB energy of $\gamma_{GB} = 1.2\,\text{J m}^{-2}$ (i.e $2\gamma_s - \gamma_{GB} = 1.7\,\text{J m}^{-2}$) (hydrated oxide) and (ii) fracture surface energy of $2\gamma_s = 1.7\,\text{J m}^{-2}$ minus a GB energy of $\gamma_{GB} = 1.2\,\text{J m}^{-2}$ (i.e $2\gamma_s - \gamma_{GB} = 0.5\,\text{J m}^{-2}$) (hydroxylated oxide). As before, the grain size, the dislocation channel thickness and the yield stress are chosen to be: $L = 50\,\mu\text{m}$, $t = 50$ nm and $\Sigma_y = 1{,}000$ MPa respectively, in agreement with experimental observations. The results plotted in Fig. 18a, b shows that the Smith and Barnby criterion underestimates strongly the IASCC stress whatever the used fracture surface energy. Using the fracture surface energy of $2\gamma_s - \gamma_{GB} = 1.7\text{J m}^{-2}$ (hydrated Fe_2O_3) (Fig. 19a), the proposed criterion underestimates slightly the IASCC stress. However, based on a fracture surface energy of $2\gamma_s - \gamma_{GB} = 0.5\,\text{J m}^{-2}$ (hydroxylated Fe_2O_3), the proposed criterion underestimates slightly the IASCC stress (Fig. 19b). Therefore, taking into account the real channel thickness improves the prediction of GB microcrack initiation in oxidizing as well as in inert environment.

7.3 Application to GB crack initiation during cyclic tests

Applying the pile-up theory to the prediction of intergranular crack initiation at the free surface of copper

polycrystals subjected to cyclic loading, Liu et al. (1992) found that the predicted critical remote stress was generally reached when GB microcracks were observed. Their modelling is based on the Smith and Barnby one. A similar approach was followed by Tanaka and Mura (1981). But Kim and Laird (1977) showed that GB microcracks appear only when the steps between adjacent grains increasing cycle by cycle are high enough. Generally many cycles are required, which contradicts the conclusions of Liu et al. of instantaneous microcrack initiation provided slip bands exist and stress saturation is reached. The number of cycles should indeed be explicitely introduced in such modelling. Either cumulation of slip irreversibilities along cycling (Mughrabi 2009; Sauzay and Gilormini 2002) or continuous production of point defects may allow the introduction of the number of cycles in the modelling. SEM pictures published in the literature show often kinds of extrusions in the surface plane, through the GBs where PSBs impinge (Weidner and Skrotzki 2010). These extrusions seem to be similar to the ones classically observed at the specimen surface showing a rather round smooth shape. This make us think that this profiles are due to point defects accumulation rather than slip irreversibility with leads to finer and sharper roughness (Mughrabi 2009). For explaining the continuous extrusion growth observed experimentally at room temperature, Polak proposed an extension of the well-known EGM theory (Mughrabi 2009), highlighting the effective role of point defect production by cyclic plasticity in PSBs and diffusion towards the surrounding matrix (Polak 1970, 1987). The analytical modelling of point defect production, annihilation and diffusion has been proposed recently (Polak and Sauzay 2009) following the numerical one of Repetto and Ortiz (1997). Work is in progress for computing both GB stress fields and crack initiation depending on the production and annihilation rates of point defect production in PSBs as well as diffusion towards the matrix.

8 Conclusions

Slip localization is widely observed in metallic polycrystals after tensile deformation, cyclic deformation or pre-irradiation followed by tensile deformation. Such strong deformation localized in thin slip bands induces local stress concentrations in the quasi-elastic matrix around, at the intersections between slip bands (SBs) and grain boundaries (GBs) where microcracks are then often observed.

Since the work of Stroh, such stress fields have been mostly modeled using the dislocation pile-up theory which leads to stress singularities similar to the LEFM ones. The Griffith criterion has then been widely applied, leading usually to strong underestimations of the macroscopic stress to GB crack initiation.

In fact, slip band thickness is finite. It varies from a few 10 nm to a few μm, depending on material, temperature and loading conditions. Then, many slip planes are plastically activated through the thickness, and not only one single atomic plane. To evaluate more realistic stress fields, numerous crystalline finite element (FE) computations are carried out using microstructure inputs (slip band aspect ratio, crystal and GB orientation...). Slip bands of low critical resolved shear stress are embedded in an elastic matrix. The following results are obtained concerning GB normal and shear stress fields:

- strong influence of slip band thickness close to the slip band corner, which is not accounted for by the pile-up theory. But far away, the thickness has a negligible effect and the predicted stress fields are close to the one predicted by the pile-up theory provided the pile-up length is assumed to be equal to the grain size;
- closed-form expressions are deduced from the numerous FE computation results allowing an easy prediction of GB stress fields. Slip band plasticity parameters, such as length and thickness, as well as crystal orientation, GB plane and remote stress are taken into account. The dependence with respect to the various parameters can be understood in the framework of matching expansions usually applied to cracks with V notches of finite thickness;

As the exponent of the GB stress close-field is about two times lower than the pile-up or LEFM crack one, the Griffith criterion may not be used for GB microcrack prediction in case of finite thickness. That is why finite crack fracture mechanics is used together with both energy and stress criteria leading to the following results:

- the effect of SB finite thickness, t, is found to affect strongly the predicted value of the remote stress to GB microcrack initiation with respect to pile-up predictions;

– the predicted dependence of this critical remote stress with respect to SB orientation as well as GB plane is discussed with respect to the one predicted by the pile-up theory;
– the predicted macroscopic tensile stresses to GB microcrack initiation are close to the experimental values measured in the case of pre-irradiated austenitic stainless steels subjected to slow tensile loading in either inert or water environment. But, the pile-up theory leads to large underestimation of the critical remote stress because it overestimates the local GB stress fields.

Acknowledgments The authors are grateful to CEA (DEN-RSTB, projects RACOC and MASOL) and the European project PERFORM60 (FP7-232612) for financial and scientific support.

References

Alexandreanu B, Was GS (2003) Grain boundary deformation-induced intergranular stress corrosion cracking of Ni-16Cr-9Fe in 360°C water. Corrosion 59:705

Antico FC, Zavattieri PD, Mance A, Rodgers WR, Okonski DA (2012) Adhesion of Nickel–Titanium shape memory alloy wires to polymeric materials: theory and experiment. Smart Mater Struct 21:1–17

Asano K, Fukuya K, Nakata K, Kodoma M (1992) Changes in grain boundary composition induced by neutron irradiation on austenitic stainless steels. In: Cubicciotti D, Simonen EP, Gold RE (eds) Proceedings of the fifth international symposium on environmental degradation of materials in nuclear power systems—water reactors, American Nuclear Society, LaGrange Park, IL, 1992, p 838

Averbach BL (1968) Some physical aspects of fracture. In: Leibowitz H (ed) Fracture chap. 7, 441. Academic Press, New York

Ben Britton TB, Wilkinson AJ (2012) Stress fields and geometrically necessary dislocation density distributions near the head of a blocked slip band. Acta Mater (Online)

Blochwitz C, Veit U (1982) Cryst Res Technol 17:529

Bruemmer SM, Simonen EP, Scott PM, Andresen PL, Was GS, Nelson JL (1999) Radiation-induced material changes and susceptibility to intergranular failure of light-water-reactor core internals. J Nucl Mater 274:299

Bruemmer SM, Thomas LE (2005) High-resolution characterizations of stress-corrosion cracks in austenitic stainless steel from crack growth tests in BWR-simulated environments. In: 12th International conference on environmental degradation of materials in nuclear power system

Byun TS, Hashimoto N, Farrell K, Lee EH (2006) Characteristics of microscopic strain localization in irradiated 316 stainless steels and pure vanadium. J Nucl Mater 351:303

Cast3m (2012) http://www-cast3m.cea.fr

Caul M, Fiedler J, Randle V (1996) Grain-boundary plane crystallography and energy in austenitic steel. Scripta Mater 35:831

Chartier A, van Brutzel L, Freyss M (2010) Atomistic study of stability of xenon nanoclusters in uranium oxide. Phys. Rev. B 81:174111

Chuang Y, Margolin H (1973) Beta-brass bicrystal stress-strain relations. Metall Trans 4:1905

Chung HM, Strain RV, Shack WJ (2001) Tensile and stress corrosion cracking properties of type 304 stainless steel irradiated to a very high dose. Nucl Eng Des 208:221

Collard C, Favier V, Berbenni S, Berveiller M (2010) Role of discrete intra-granular slip bands on the strain-hardening of polycrystals. Int J Plast 26:310

Cordisco F, Zavattieri PD, Bower AF (2012) Toughness of a sinusoidal interface between two elastically dissimilar solids. Eng Fract Mech 96:192–208

Couvant Th (2001) Private communication

Dawson I, Bristowe PD, Lee M-H, Payne MC, Segall MD, White JA (2006) First-principles study of a tilt grain boundary in rutile. Phys Rev B 54:13727

de Leeuw NH, Cooper TG (2007) Surface simulation studies of the hydration of white rust Fe(OH)(2), goethite alpha-FeO(OH) and hematite alpha-Fe2O3. Geochim Cosmochim Acta 71:1655

Diard O, Leclercq S, Rousselier G, Cailletaud G (2005) Evaluation of finite element based analysis of 3D multicrystalline aggregates plasticity. Application to crystal plasticity model identification and the study of stress and strain fields near grain boundaries. Int J Plast 21:691

Dropek RB, Was GS, Gan J, Cole I, Allen TR, Kenik EA (2004) 11th International conference on American Nuclear Society, La Grange Park, IL, p 1132

Dugdale H, Armstrong D EJ, Tarleton E, Roberts S G, Lozano-Perez S (2013) How oxidized grain boundaries fail. Acta Mater (online)

Edwards DJ, Singh BN, Bikde-Sorensen JB (2005) Initiation and propagation of cleared channels in neutron-irradiated pure copper and a precipitation hardened CuCrZr alloy. J Nucl Mater 342:164

El Kadiri H, Kapil J, Oppedal AL, Agnew SR, Cherkaoui M, Vogel SC (2013) The effect of twin–twin interactions on the nucleation and propagation of 10–12 twinning in Magnesium. Acta Mater 61:3549–3563

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc R Soc Lond A 241:376

Essmann U, Differt K (1996) Dynamic model of the wall structure in persistent slip bands of fatigued metals II. The wall spacing and the temperature dependence of the yield stress in saturation. Mater Sci Eng A 208:56

Evrard P, Sauzay M (2010) Modelling of the effect of dislocation channel on intergranular microcrack nucleation in pre-irradiated austenitic stainless steels during low strain rate tensile loading. J Nucl Mater 405:83

Finney JM, Laird C (1975) Strain localization in cyclic deformation of copper single crystals. Philos Mag 31:339

Fisher CAJ (2004) Molecular dynamics simulations of reconstructed NiO surfaces. Scripta Mater 50:1045

Fukuya K, Nishioka H, Fujii K, Torimaru TJ (2008) Effects of dissolved hydrogen and strain rate on IASCC behavior in highly irradiated stainless steels. J Nucl Sci Technol 45:452

Gerold V, Steiner D (1982) Fatigue softening in precipitation hardened copper-cobalt. Scripta Metall 16:405

Henninger C, Leguillon D, Martin E (2007) Crack initiation at a V-notch-comparison between a brittle fracture criterion and the Dugdale cohesive model. C.R. Mécanique 335:388

Huntington HB (1958) The elastic constants of crystals. In: Seitz F, Turnbull D (eds) Solid state physics, vol 7. Academic Press Incorporation Publishers, New York, pp 214–351

Jiao Z, Busby JT, Obata R, Was GS (2005) Influence of localized deformation on irradiation-assisted stress corrosion cracking of proton-irradiated austenitic alloys. In: 12th International conference on environmental degradation of materials in nuclear pPower systems—water reactors. Salt Lake City, Utah, p 379

Kim WH, Laird C (1977) Crack nucleation and stage I propagation in high strain fatigue, I and II. Acta Metall 26:777

Lee EH, Yoo MH, Byun TS, Hunn JD, Farrell K, Mansur LK (2001) On the origin of deformation microstructures in austenitic stainless steel: part II—mechanisms. Acta Mater 49:3277

Lee EH, Byun TS, Hunn JD, Farrell K, Mansur LK (2001) Origin of hardening and deformation mechanisms in irradiated 316 LN austenitic stainless steel. J Nucl Mater 296:183

Leguillon D, Sanchez-Palencia E (1987) Computation of singular solutions in elliptic problems and elasticity. Wiley/Masson, New York

Leguillon D (2002) Strength or toughness? A criterion for crack onset at a notch. Eur J Mech A/Solids 21:61

Li S, Sauer RA, Wang G (2007) The Eshelby tensors in a finite spherical domain—part I: theoretical formulations. J Appl Mech 74:770

Lim LC, Raj R (1984a) On the distribution of sigma for grain-boundaries in polycrystalline nickel prepared by strain-annealing technique. Acta Metall 32:1177

Lim LC, Raj R (1984b) Effect of boundary structure on slip-induced cavitation in polycrystalline nickel. Acta Metall 32:1183

Liu W, Bayerlein M, Mughrabi H, Day A, Quested PN (1992) Crystallographic features of intergranular crack initiation in fatigued copper polycrystals. Acta Metall Mater 40:1763

McMurtrey MD and Was GS (2011) Role of slip behavior in the irradiation assisted stress corrosion cracking in austenitic stainless steels. In: Busby JT, Ilevbare G, Andresen L (eds) 15th International Conference on Environmental Degradation. The Minearlas, Metals & Materials Society, p 1383

Man J, Obrtlik K, Blochwitz C, Polak J (2002) Atomic force microscopy of surface relief in individual grains of fatigued 316L austenitic stainless steel. Acta Mater 50:3767

Margolin H, Stanescu MS (1975) Polycrystalline strengthening. Acta Metall 23:1411

Matweb.com (2012) http://www.matweb.com

Mughrabi H, Wang R (1988) Cyclic stress-strain response and high-cycle fatigue behaviour of copper polycrystals. In: Lukas P, Polak J (eds) Basic mechanisms in fatigue of metals. Elsevier, Amsterdam, p 1

Lukas P, Knesnil M, Krejci J (1968) Dislocations and persistent slip bands in copper single crystals fatigued at low stress amplitude. J Phys Stat Sol 27:545

Mughrabi H (2009) Cyclic slip irreversibility and the evolution of fatigue damage. Metall Mater Trans A 40:1257–79

Murer S, Leguillon D (2010) Static and fatigue failure of quasi-brittle materials at a V-notch using a Dugdale model. Eur J Mech A/Solids 29:2010

Neumann P (1992) The effect of surface related grain-boundary stresses on fatigue. Scripta Metall Mater 26:1535

Nishioka H, Fukuya K, Fujii K, Torimaru TJ (2008) IASCC initiation in highly irradiated stainless steels under uniaxial constant load conditions. J Nucl Sci Technol 45:1072

Nogaret T, Rodney D, Fivel M, Robertson Ch (2008) Clear band formation simulated by dislocation dynamics: role of helical turns and pile-ups. J Nucl Mater 380:22

Onchi T, Dohi K, Soneda N, Cowan JR, Scowen RJ, Castano ML (2003) Fractographic and microstructural characterization of irradiated 304 stainless steel intergranularly fractured in inert gas. J Nucl Mater 320:194

Onimus F, Monnet I, Béchade J-L, Prioul C, Pilvin P (2004) A statistical TEM investigation of dislocation channeling mechanism in neutron irradiated zirconium alloys. J Nucl Mater 328:165

Perrin C, Berbenni S, Vehoff H, Berveiller M (2010) Role of discrete intragranular slip on lattice rotations in polycrystalline Ni: Experimental and micromechanical studies. Acta Mater 58:4649

Polak J (1970) The effect of intermediate annealing on the electrical resistivity and shear stress of fatigued copper. Scripta Metall 4:761–764

Polak J (1987) Resistivity of fatigued copper single crystals. Mater Sci Eng 89:35–43

Polak J, Sauzay M (2009) Growth of extrusions in localized cyclic plastic straining. Mater Sci Eng A 500:122

Priester L (2001) "Dislocation–interface" interaction—stress accommodation processes at interfaces. Mater Sci Eng A 309–310:430

Pugno N, Ruoff R (2004) Quantized fracture mechanics. Philos Mag 84:2829

Rasmussen KV, Pedersen OB (1980) Fatigue of copper polycrystals at low plastic strain amplitudes. Acta Metall 14: 1467

Repetto EA, Ortiz M (1997) A micromechanical model of cyclic deformation and fatigue-crack nucleation in fcc single crystals. Acta Mater 45:2577

Rice JR, Wang JS (1993) Embrittlement of interfaces by solute segregation. Mater Sci Eng A 107:23

Rose JH, Ferrante J, Smith JR (1981) Universal binding energy curves for metals and bimetallic interfaces. Phys Rev Lett 47:675

Sauzay M (2007) Cubic elasticity and stress distribution at the free surface of polycrystals. Acta Mater 55:1193–1202

Sauzay M, Gilormini P (2000) Surface and cyclic microplasticity. Fat Fract Eng Mater Struct 23:573

Sauzay M, Gilormini P (2002) Influence of surface effects on fatigue of microcracks nucleation. Theor Appl Fract Mech 38:53

Sauzay M, Caës C, Mottot M, Robertson C (2003) Une étude numérique du rôle de la couche d'oxydes dans l'amorçage de fissures de fatigue à moyenne température. J Phys IV France 106:99

Sauzay M, Jourdan Th (2006) Polycrystalline microstructure, cubic elasticity and initiation of high-cycle fatigue short cracks. Int J Fract 141:431–446

Sauzay M, Man J (2008) Influence of crystalline elasticity on the stress distribution at the free surface of an austenitic stainless steel polycrystal. Comparison with experiments. Mater Sci Forum 567–568:149–152

Sauzay M, Bavard K, Karlsen W (2010) TEM observations and finite element modeling of channel deformation in pre-irradiated austenitic stainless steels—interactions with free surfaces and grain boundaries. J Nucl Mater 406:152

Sauzay M, Kubin L (2011) Scaling laws for dislocation microstructures in monotonic and cyclic deformation off cc metals. Prog Mater Sci 56:725

Sauzay M, Vor K (2013) Influence of plastic slip localization on grain boundary stress fields and microcrack initiation. Eng Fract Mech (accepted for publication)

Sharp JV (1967) Deformation of neutron-irradiated copper single crystals. Philos Mag 16:77

Simonen EP, Bruemmer SM (1995) In: McIllree AR, Bruemmer SM (eds) Proceedings of seventh international symposium on environmental degradation of materials in nuclear power systems—water reactors. National Association of Corrosion Engineers, Houston, TX, vol 2, p 1

Smith E, Barnby JT (1967) Nucleation of grain-boundary cavities during high-temperature creep. Metall Sci J 1:56

Stroh AN (1957) A theory of the fracture of metals. Adv Phys 6:418

Sun J, Stirmer T, Matthews A (2006) Surf Coat Technol 201:4205

Sutton AP, Balluffi RW (1995) Interfaces in crystalline materials. Clarendon Press, Oxford

Tabata T, Fujita H, Hiraoka M-A, Onishi K (1983) Dislocation behaviour and the formation of persistent slip bands in fatigued copper single crystals observed by high-voltage electron microscopy. Philos Mag A 47:841

Takakura K, Nakata K, Sakima K, Fujimoto K, Kubo N (2007) Workshop. Beaune, France

Tan L, Sridharan K, Allen TR (2005) The effect of grain boundary engineering on the oxidation behaviour of INCOLOY alloy 800H in supercritical water. J Nucl Mater 348:263

Tanaka K, Mura T (1981) A dislocation model for fatigue crack initiation. J Appl Mech 103:97

Taylor D, Cornetti P, Pugno N (2005) The fracture mechanics of finite crack extension. Eng Fract Mech 72:1021

Toivonen A, Ehrnsten U, Karlsen W, Aaltonen P, Massoud JP, Boursier JM (2005) In: Allen TR, King PJ, Nelson L (eds) Proceedings of twelvesth international symposium on environmental degradation of materials in nuclear power systems—water reactors: TMS (The Minerals, Metals & Materials Society), p 327

Tong W, Hector LG Jr (1997) In situ characterization of a binary aluminum alloy during tensile deformation. Scripta Mater 36:1339–1344

Victoria M, Baluc N, Bailat C, Dai Y, Luppo MI, Schäublin R, Singh BN (2000) The microstructure and associated tensile properties of irradiated fcc and bcc metals. J Nucl Mater 276:114

Vitos L, Ruban AV, Skriver HL, Kollar J (1998) The surface energy of metals. Surf Sci 411:186

Was G (2007) Fundamentals of radiation materials science: metals and alloys. Springer, The Netherlands 235, chapter 15

Weidner A, Beyer R, Blochwitz C, Holste C, Schwab A, Tirschler W (2006) Slip activity of persistent slip bands in polycrystalline nickel. Mater Sci Eng A 435–436:540–546

Weidner A, Sauzay M, Skrotzki W (2010) Experimental observation of cyclic slip irreversibility factor. Key Eng Mater 465:223–226

Weidner A, Skrotzki W (2010) Proceedings of the Engineering

Winter AT, Pedersen OB, Rasmussen KV (1981) Dislocation microstructures in fatigued copper polycrystals. Acta Metall 29:735

Wejdemann C, Pedersen OB (2004) Atomic force microscopy of the intense slip localization causing fatigue crack initiation in polycrystalline brass. Mater Sci Eng A 387:556

Xie CL, Ghosh S, Groeber M (2004) Modeling cyclic deformation of HSLA steels using crystal plasticity. J Eng Mater Technol 126:339–352

Yamakov V, Saether E, Phillips DR, Glaessgen EH (2006) Molecular-dynamics simulation-based cohesive zone representation of intergranular fracture processes in aluminum. J Mech Phys Sol 54:1899

Yao Z, Schäublin R, Victoria M (2002) The microstructure and tensile properties of Ni single crystals irradiated with high energy protons. J Nucl Mater 329–333:1127–1132

Yao Z, Schäublin R, Victoria M (2004) Tensile properties of irradiated Cu single crystals and their temperature dependence. J Nucl Mater 329–333:1127–1132

Zavattieri PD, Hector LG Jr, Bower AF (2007a) Determination of the effective mode-I toughness of a sinusoidal interface between two elastic solids. Int J Fract 145:167–180

Zavattieri PD, Hector LG Jr, Bower AF (2007b) Cohesive zone simulations of crack growth along a rough interface between two elastic plastic solids. Eng Fract Mech 75:4309-4332

Zavattieri P, Savic V, Hector LG Jr, Fekete JR, Tong W, Xuan Y (2009) Spatio-temporal characteristics of the Portevin-Le Châtelier effect in austenitic steel with twinning induced plasticity. Int J Plast 25:2298–2330

DOI 10.1007/978-3-319-04397-5_16
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9879-3

ORIGINAL PAPER

Modeling the heterogeneous effects of retained austenite on the behavior of martensitic high strength steels

Q. Wu · P. Shanthraj · M. A. Zikry

Received: 20 December 2012 / Accepted: 28 August 2013 / Published online: 6 September 2013

Abstract The effects of pockets of retained austenite on the behavior of martensitic steels have been investigated. A dislocation-density based crystalline plasticity and specialized finite-element formulation were used to investigate how f.c.c. austenite pockets interact with b.c.c. martensitic laths. Quasi-static and dynamic analyses were undertaken to investigate how the effects of the orientations of parent austenite grains and different crystallographic interfaces affect shear strain localization, strength, and toughness. It is shown that the orientations of the parent austenite grain have a significant effect on the dominance of specific interfacial slip systems, and this subsequently affects whether the retained austenite has incompatible slip with martensitic laths, for low austenite Euler angles, or compatible slip with martensitic laths, for high values of austenite Euler angles.

Keywords Dislocation-density · Crystal plasticity · Martensitic steel · Retained austenite · Failure

1 Introduction

A small amount of retained austenite is frequently observed in martensitic steel depending on the carbon content and heat treatment conditions (Thomas 1978; Park et al. 2004). The morphology and volume fraction of retained austenite have been investigated by transmission electron microscopy (TEM) and scanning electron microscopy (SEM), and retained austenite is mainly located at martensitic inter-lath boundaries and within laths (Song et al. 2010; Ma et al. 2012). Experimental investigations indicate that retained austenite can improve ductility and toughness, but decrease the strength of martensitic steel (see, for example, Nakagawa and Miyazaki 1999; Moor et al. 2008), Using X-ray diffraction line profile analysis, Zhang et al. (2011) observed that the orientation relationship (OR) between martensite and retained austenite plays a significant role in how dislocation densities propagate from martensite to retained austenite pockets.

Lath martensitic steels due to their high strength, and toughness have myriad military and civilian applications. These properties are uniquely inherent to martensitic steels, as a result of its lath microstructure that has distinct orientations, distributions, and morphologies pertaining to martensitic transformations, (see, for example, Morito et al. 2003, 2006). Specifically, the effects of lath martensite (b.c.c.) morphology, parent austenite (f.c.c.) orientation, and retained austenite all have interrelated effects on deformation and failure.

One of the issues that need to be understood, however, is how dislocation density evolution at the interfaces of austenite–martensite affects failure modes

Q. Wu · P. Shanthraj · M. A. Zikry (✉)
Department of Mechanical and Aerospace Engineering,
North Carolina State University, Raleigh, NC
27695-7910, USA
e-mail:zikry@ncsu.edu

and overall behavior. The objective of the present work, therefore, is to obtain a detailed quasi-static and dynamic understanding of the interrelated physical mechanisms that can result in the interaction of martensite b.c.c. laths with f.c.c. retained austenite pockets. Specifically, we want to understand how dislocation-density accumulation and interaction along austenite (f.c.c.) and martensite (b.c.c.) interfaces affect plastic-slip accumulation and local stresses. Furthermore, a detailed understanding of how lath morphology and orientation, due to block and packet distributions, affects local behavior will be investigated. A multiple-slip dislocation-density based constitutive formulation is used to obtain a detailed understanding and accurate characterization of interrelated material mechanisms, which occur over different scales in crystalline materials due to different austenite grain orientations. The formulation is based on the framework recently developed by Shanthraj and Zikry (2011) for f.c.c. and b.c.c. crystalline structures. The dislocation-density evolution equations are coupled through the interaction of forest densities, which account for the formation and annihilation of junctions. The evolution equations are coupled to a multiple-slip crystal plasticity formulation, and specialized finite-element techniques (Zikry 1994) are used to characterize the dominant dislocation-density interaction mechanisms between the f.c.c. retained austenite pockets and the b.c.c. martensitic laths at different orientations that govern the competition between strength and toughness.

This paper is organized as follows: the dislocation-density crystalline plasticity formulation and the evolution equations are presented in Sect. 2, the numerical implementation is outlined in Sect. 3, the results are presented and discussed in Sect. 4, and a summary of the quasi-static and dynamic results and conclusions are given in Sect. 5.

2 Constitutive formulation

In this section, only a brief outline of the multiple-slip crystal plasticity rate-dependent constitutive formulation and the evolution equations for the mobile and immobile dislocation-densities, which are coupled to the constitutive formulation, are presented. A detailed presentation is given by Shanthraj and Zikry (2011).

2.1 Multiple-slip dislocation-density based crystal plasticity formulation

The dislocation-density based crystal plasticity constitutive framework used in this study is based on a formulation developed by Zikry (1994), Ashmawi and Zikry (2003), and Shanthraj and Zikry (2012), and a brief outline will be presented here. It is assumed that the velocity gradient is decomposed into a symmetric deformation rate tensor D_{ij} and an anti-symmetric spin tensor W_{ij} (Asaro and Rice 1977). The tensors D_{ij} and W_{ij} are then additively decomposed into elastic and inelastic components as

$$D_{ij} = D^*_{ij} + D^p_{ij}, \quad W_{ij} = W^*_{ij} + W^p_{ij}, \tag{1 a-b}$$

The superscript * denotes the elastic part, and the superscript p denotes the plastic part. W^*_{ij} includes the rigid body spin. The inelastic parts are defined in terms of the crystallographic slip-rates as

$$D^p_{ij} = \sum_{\alpha} P^{(\alpha)}_{ij} \dot{\gamma}^{(\alpha)}, \quad \text{and } W^p_{ij} = \sum_{\alpha} \omega^{(\alpha)}_{ij} \dot{\gamma}^{(\alpha)}, \tag{2 a-b}$$

where α is summed over all slip-systems, and $P^{(\alpha)}_{ij}$ and $\omega^{(\alpha)}_{ij}$ are the symmetric and anti-symmetric parts of the Schmid tensor in the current configuration respectively.

A power law relation can characterize the rate-dependent constitutive description on each slip system as

$$\dot{\gamma}^{(\alpha)} = \dot{\gamma}^{(\alpha)}_{ref} \left[\frac{\tau^{(\alpha)}}{\tau^{(\alpha)}_{ref}} \right] \left[\frac{\left|\tau^{(\alpha)}\right|}{\tau^{(\alpha)}_{ref}} \right]^{\frac{1}{m}-1}, \tag{3}$$

where $\dot{\gamma}^{(\alpha)}_{ref}$ is the reference shear strain-rate which corresponds to a reference shear stress $\tau^{(\alpha)}_{ref}$, and m is the rate sensitivity parameter. $\tau^{(\alpha)}$ is the resolved shear stress on slip system α. The reference stress used is a modification of widely used classical forms (Franciosi et al. 1980) that relate reference stress to immobile dislocation-density ρ_{im} as

$$\tau^{(\alpha)}_{ref} = \left(\tau^{(\alpha)}_{y} + G \sum_{\beta=1}^{nss} b^{(\beta)} \sqrt{a_{\alpha\beta} \rho^{(\beta)}_{im}} \right) \left(\frac{T}{T_0} \right)^{-\xi}, \tag{4}$$

where $\tau_y^{(\alpha)}$ is the static yield stress on slip system α, G is the shear modulus, *nss* is the number of slip systems, $b^{(\beta)}$ is the magnitude of the Burgers vector, and $a_{\alpha\beta}$ are Taylor coefficients which are related to the strength of interactions between slip-systems (Devincre et al. 2008; Kubin et al. 2008a,b). T is the temperature, T_0 is the reference temperature, and ξ is the thermal softening exponent, which is chosen as 0.3.

2.2 Mobile and immobile dislocation density evolution equations

Following the approach of Zikry and Kao (1996), it is assumed that, for a given deformed state of the material, the total dislocation-density, $\rho^{(\alpha)}$, can be additively decomposed into a mobile $\rho_m^{(\alpha)}$ and an immobile dislocation-density, $\rho_{im}^{(\alpha)}$. Furthermore, the mobile and immobile dislocation-density rates can be coupled through the formation and destruction of junctions as the stored immobile dislocations act as obstacles for evolving mobile dislocations. This is the basis for taking the evolution of mobile and immobile dislocation densities as

$$\frac{d\rho_m^\alpha}{dt} = |\dot{\gamma}^\alpha| \left(\frac{g_{sour}^\alpha}{b^2} \left(\frac{\rho_{im}^\alpha}{\rho_m^\alpha} \right) - g_{mnter-}^\alpha \rho_m^\alpha - \frac{g_{immob-}^\alpha}{b} \sqrt{\rho_{im}^\alpha} \right), \tag{5}$$

$$\frac{d\rho_{im}^\alpha}{dt} = |\dot{\gamma}^\alpha| \left(g_{mnter+}^\alpha \rho_m^\alpha + \frac{g_{immob+}^\alpha}{b} \sqrt{\rho_{im}^\alpha} - g_{recov}^\alpha \rho_{im}^\alpha \right), \tag{6}$$

where g_{sour} is a coefficient pertaining to an increase in the mobile dislocation-density due to dislocation sources, g_{mnter} are coefficients related to the trapping of mobile dislocations due to forest intersections, cross-slip around obstacles, or dislocation interactions, g_{recov} is a coefficient related to the rearrangement and annihilation of immobile dislocations, and g_{immob} are coefficients related to the immobilization of mobile dislocations.

2.3 Determination of dislocation density evolution coefficients

To couple the evolution equations for mobile and immobile dislocation densities to the crystal plasticity formulation, the non-dimensional coefficients in (5–6) were determined as functions of the crystallography and deformation mode of the material, by considering the generation, interaction and recovery of dislocation densities as discussed in Shanthraj and Zikry (2011). These expressions are summarized in Table 1, where f_0, and φ are geometric parameters. H_0 is the reference activation enthalpy, ρ_s is the saturation density and the average junction length, l_c, can be approximated as

Table 1 g coefficients in Eqs. (5–6)

g Coefficients	Expression
g_{sour}^α	$b^\alpha \varphi \sum_\beta \sqrt{\rho_{im}^\beta}$
g_{mnter-}^α	$l_c f_0 \sum_\beta \sqrt{a_{\alpha\beta}} \left[\frac{\rho_m^\beta}{\rho_m^\alpha b^\alpha} + \frac{\dot{\gamma}^\beta}{\dot{\gamma}^\alpha b^\beta} \right]$
g_{immob-}^α	$\frac{l_c f_0}{\sqrt{\rho_{im}^\alpha}} \sum_\beta \sqrt{a_{\alpha\beta}} \rho_{im}^\beta$
g_{mnter+}^α	$\frac{l_c f_0}{\dot{\gamma}^\alpha \rho_m^\alpha} \sum_{\beta,\gamma} n_\alpha^{\beta\gamma} \sqrt{a_{\beta\gamma}} \left[\frac{\rho_m^\gamma \dot{\gamma}^\beta}{b^\beta} + \frac{\rho_m^\beta \dot{\gamma}^\gamma}{b^\gamma} \right]$
g_{immob+}^α	$\frac{l_c f_0}{\dot{\gamma}^\alpha \sqrt{\rho_{im}^\alpha}} \sum_\beta n_\alpha^{\beta\gamma} \sqrt{a_{\beta\gamma}} \rho_{im}^\gamma \dot{\gamma}^\beta$
g_{recov}^α	$\frac{l_c f_0}{\dot{\gamma}^\alpha} \left(\sum_\beta \sqrt{a_{\beta\gamma}} \frac{\dot{\gamma}^\beta}{b^\beta} \right) e^{\left(\frac{-H_0 \left(1 - \sqrt{\frac{\rho_{im}^\alpha}{\rho_s}} \right)}{KT} \right)}$

$$l_c = \frac{1}{\sum_\beta \sqrt{\rho_{im}^{(\beta)}}}, \tag{7}$$

An interaction tensor, $n_\alpha^{\beta\gamma}$, is introduced and defined as having a value of 1 if dislocations on slip-systems β and γ interact to form an energetically favorable junction on slip system α, and a value of 0 if there are no interactions. This interaction tensor is used to map the dislocation-density interactions with the corresponding slip system, and the energy criterion based on Frank's rule is used to determine the formation of junctions as

$$n_\alpha^{\beta\gamma} = \begin{cases} 1 & \text{if } Gb^{(\alpha)2} < Gb^{(\beta)2} + Gb^{(\gamma)2} \text{ and } b^{(\alpha)} = b^{(\beta)} + b^{(\gamma)} \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

To obtain the interaction tensor, $n_\alpha^{\beta\gamma}$, the total number of interactions between slip systems has to be considered. In f.c.c. crystals, using the family of $<110>\{111\}$ slip systems, the total interactions can be reduced to six basic interaction types based on the symmetry of the crystal structure (Kubin et al. 2008b). These interactions are: the self interaction, between the

same slip system; the co-linear interaction, between slip systems with parallel Burgers vectors; the co-planar interaction, between co-planar slip systems; the interaction between slip systems forming Lomer locks; the glissile junction, between non-co-planar slip systems; and the interaction between slip systems forming Hirth locks. The interaction tensor can be obtained by considering the product of each interaction type. To explicitly account for the storage of locks, the mobile and immobile dislocation densities on the 12 slip systems in f.c.c. crystals are appended with six immobile dislocation densities pertaining to the storage of Lomer locks on the {100} planes. This leads to a total of 12×12 mobile–mobile interactions and 12×18 mobile-immobile interactions. Similarly, in b.c.c. crystals, using the $<111>\{110\}$ and $<111>\{112\}$ families of slip systems, the total interactions are reduced to three interaction types. These interaction types are: the self and co-linear interaction between the same slip system and slip systems with parallel Burger's vector; the interaction between slip systems to form binary junctions having $<100>$ Burger's vectors; and the interaction between slip systems and binary junctions to form ternary junctions having $<111>$ Burger's vectors. To explicitly account for the storage of junctions, the mobile and immobile dislocation densities on the 24 slip systems in b.c.c. crystals are appended with 19 immobile dislocation densities pertaining to the storage of $<100>$ binary junctions on various crystallographic planes. This leads to a total of 24×24 mobile–mobile interactions and 24×43 mobile-immobile interactions. Thus, the formation of ternary junctions (Bulatov et al. 2006; Madec and Kubin 2008), can be explicitly accounted for through the interaction of dislocation locks and binary junctions with mobile dislocation densities. It should be noted that the coefficients in Table 1 are functions of the immobile and mobile densities, and hence are updated as a function of the deformation mode (Shanthraj and Zikry 2011).

3 Numerical methods

The total deformation rate tensor, D_{ij}, and the plastic deformation rate tensor, D_{ij}^{p}, are needed to update the material stress state. The method used here is the one developed by Zikry (1994) for rate-dependent crystalline plasticity formulations, and only a brief outline will be presented here. For quasi-static deformations, an implicit FE method with BFGS iteration is used to obtain the total deformation rate tensor, D_{ij}. To overcome numerical instabilities associated with stiffness, a hybrid explicit-implicit method is used to obtain the plastic deformation rate tensor, D_{ij}^{p}. This hybrid numerical scheme is also used to update the evolutionary equations for the mobile and immobile densities. The hybrid method is based on using explicit Runge Kutta and it is switched to an implicit Euler method, when numerical stiffness is encountered. Numerical stiffness can be encountered when due to different rate changes along slip systems, slip-rates, resolved shear–stresses, and dislocation-densities can vary widely. For dynamic deformations, an explicit method is used with a lumped mass, one point integration, central-time-difference, and a stiffness based hourglass control. Details for this dynamic approach are given in Zikry (1994) and Shanthraj and Zikry (2011).

4 Results and discussion

The multiple-slip dislocation-density-based crystal plasticity formulation is coupled to the specialized FE method to investigate the large strain behavior of martensitic steel with distributions of retained austenite. To model the microstructure of the martensite, we used a combination of blocks and packets. This approach is based on the approach developed by Hatem and Zikry (2009). Blocks are collections of laths with low misorientation, and packets are collections of blocks that have the same habit plane. In this study, 40 martensitic blocks are distributed randomly with 14 packets from one parent austenite grain, and the variant of martensite blocks were obtained based on the orientation relationships between the parent austenite grain and martensite blocks. The variant arrangements were obtained from experimental EBSD observations by Morito et al. (2003). It is assumed that the retained austenite has a volume fraction of 5 % (Thomas 1978), and that the pockets of retained austenite are randomly distributed inside the martensite blocks, or between the blocks and the packets (Fig. 1). It is assumed that the retained austenite pockets has the same grain orientation as the parent austenite grain. The material properties (Table 2) that are used are representative of low-carbon martensitic steel and austenitic stainless steel (Byun et al. 2004).

The parent austenite grain is oriented based on the loading plane of $(0\,0\,1)_{\gamma}$ and a loading direction of

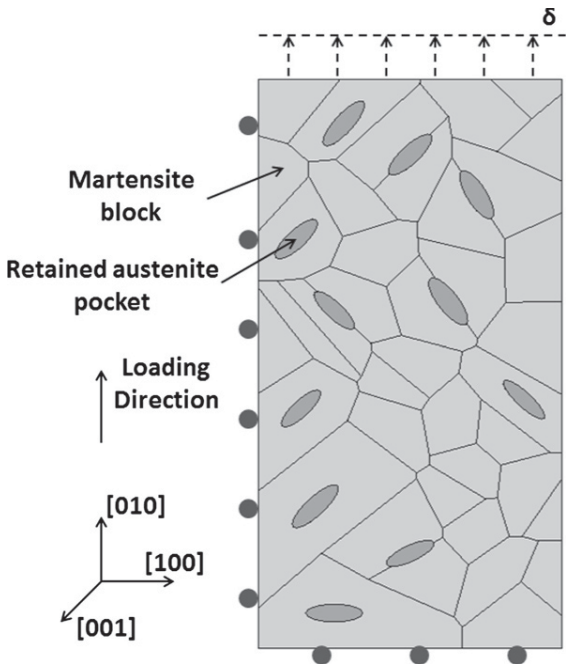

Fig. 1 Microstructural model and retained austenite pocket, block/packet arrangement

Table 2 Material properties

Properties	Retained austenite	Martensite
Young's modulus, E	100 GPa	228 GPa
Static yield stress, τ_y	100 MPa	517 MPa
Poisson's ratio, ν	0.3	0.3
Rate sensitivity parameters m	0.01	0.01
Reference strain rate, $\dot{\gamma}_{ref}$	0.001 s^{-1}	0.001 s^{-1}
Critical strain rate, $\dot{\gamma}_{critical}$	10^4 s^{-1}	10^4 s^{-1}
Burger vector, b	3.0×10^{-10} m	3.0×10^{-10} m
Saturation dislocation density, ρ_s	1.0×10^{14} m^{-2}	1.0×10^{16} m^{-2}
Thermal softening exponent, ξ	0.3	0.3
Fraction of plastic energy to heat, χ	1.0	1.0
Geometric parameter, φ	0.1	0.1
Geometric parameter, f_0	0.1	0.05

$[0\,1\,0]_\gamma$. The Kurdjumov–Sachs (K–S) OR is adopted as the martensite OR, and $\{111\}_\gamma$ is assumed as the habit plane. A convergent plane strain FE mesh of 4,893 elements was used with a specimen size of 3.2 mm × 6.4 mm, and a displacement load is applied

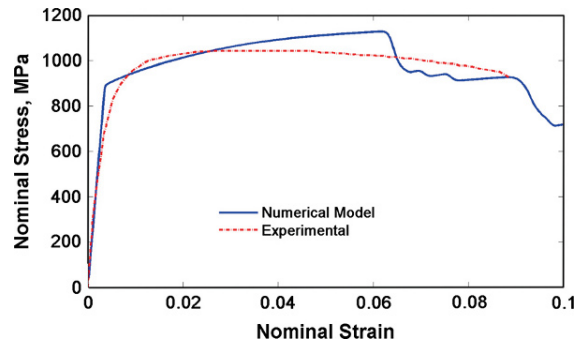

Fig. 2 Nominal stress–strain curve for experimental and numerical model

for a quasi-static nominal strain rate with symmetric boundary conditions applied on the left and bottom edges.

To validate the modeling approach, comparisons were made with experiments conducted by Shibata et al. (2012) on low carbon martensitic steel for a quasi-static tensile strain rate of 8.3×10^{-6} s^{-1}. For this comparison, it is assumed that the material is 100 % martensite. The nominal stress strain curves for the numerical and the experimental results are shown in Fig. 2. The maximum difference between the experimental and the numerical stress values is approximately 10 %. This small differences further validates the model. However, it should be noted the experimental results soften at approximately 4 %, and the model shows instability at approximately 6 %. This difference can be due to heterogeneities, such as second phase particles, which are not accounted for in the proposed model.

4.1 Retained austenite–martensite interaction

Different distributions of pockets of retained austenite were randomly distributed within the martensitic aggregate for a volume fraction of 5 % (Fig. 1). It is initially assumed that the austenite grain had a cube Euler orientation of (0°, 0°, 0°). The contours for the normalized (by the initial mobile dislocation density) mobile dislocation densities and the normalized (by the initial immobile dislocation density) immobile dislocation densities corresponding to the most active slip system of the martensite and retained austenite aggregate at a nominal strain of 20 % are shown in Fig. 3. For the most active slip system of martensite $(\bar{1}12)[1\bar{1}1]$, the maximum normalized mobile disloca-

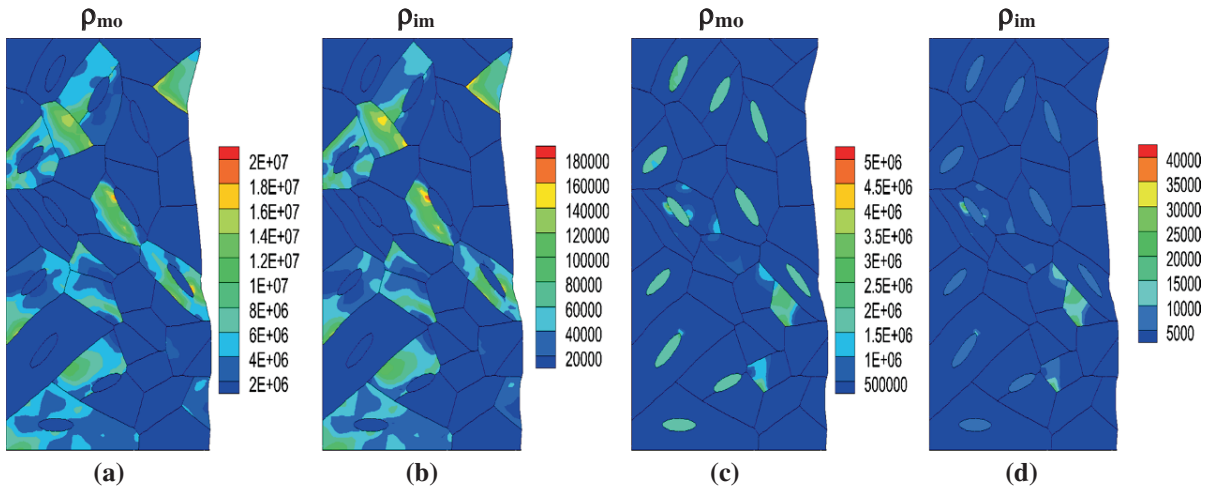

Fig. 3 Mobile and immobile dislocation densities at a nominal strain of 20 % for, **a**, **b** most active slip system in martensite $(\bar{1}12)[1\bar{1}1]$, **c**, **d** most active slip system in retained austenite $(1\bar{1}1)[011]$

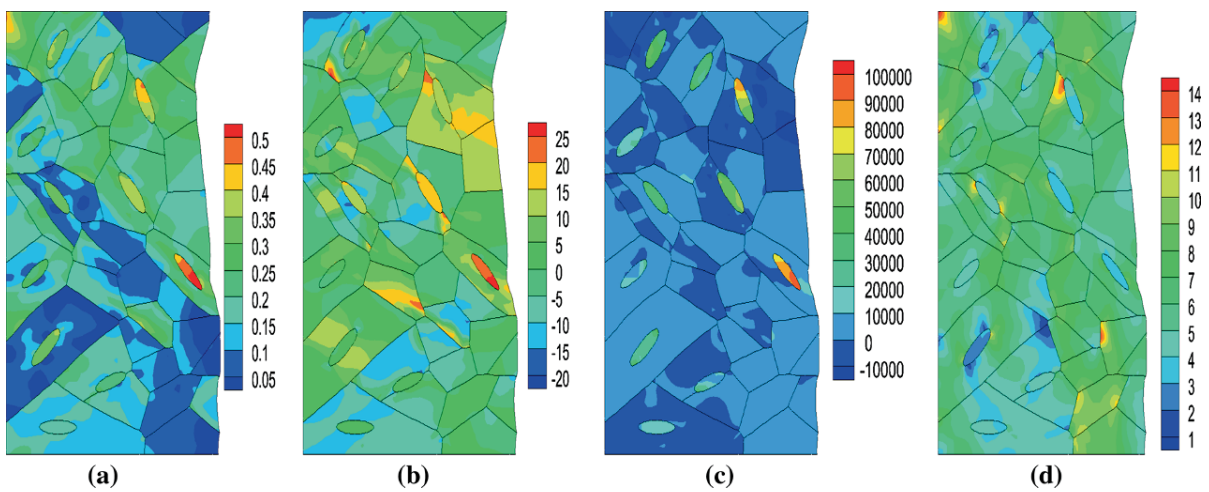

Fig. 4 Behavior at a nominal strain of 20 %, **a** shear slip, **b** lattice rotation, **c** normalized total interaction dislocation density, **d** normalized normal stress

tion density is 2.0×10^7, and the maximum normalized immobile dislocation density is 1.8×10^5. The evolution of dislocation densities along selected blocks results in the localization of plastic slip (Fig. 4a). The dislocation density at the interface between martensite blocks and retained austenite pockets is a maximum, and this is likely due to the interaction of the martensitic slip systems and the retained austenite slip systems. Based on the K–S orientation relationship, the slip systems {110}<111> in martensite should be aligned with the slip systems {111}<110> in austenite. However, the {112}<111> slip systems in martensite are incompatible with the austenite slip systems, which can impede dislocation density transmission between martensite blocks and retained austenite interfaces. For the most active slip system in austenite of $(1\bar{1}1)[011]$, the normalized mobile dislocation density is 2.0×10^6, and the normalized immobile dislocation density is 1.0×10^4, which are much lower than those in martensite. This can be due to the initial cube orientation of the austenite grain and the lower number of available f.c.c. slip-systems in comparison with the available b.c.c. slip-systems.

The accumulated plastic slip at a nominal strain of 20 % is shown in Fig. 4a. The maximum accumulated slip is 0.5, and it occurs within the retained austenite pockets. The tensile loading direction is aligned along the $[010]_\gamma$ direction, which can result in a maximum resolved shear stress along the $[011]_\gamma$ directions. The slip-direction $[011]_\gamma$ of the f.c.c. retained austenite is also parallel to the long direction of martensitic laths and blocks, and parallel to the slip direction $[111]_\alpha$ based on the K–S OR. The retained austenite grains are at the martensite inter-lath and block boundary, and the long direction should, therefore, be parallel to that of martensite. This would align the austenite slip systems with the maximum resolved shear stress along the long direction of the martensitic blocks and retained austenite grains, which would result in the shear–strain localization of plastic slip. Furthermore, due to the incompatibility of the b.c.c. slip system $(\bar{1}12)[1\bar{1}1]$ and the f.c.c. slip system $(1\bar{1}1)[011]$, the accumulation of plastic slip occurs at the b.c.c.-f.c.c. interface, which is also exacerbated by the geometrical softening associated with the lattice rotation of both slip-systems (Fig. 4b).

The interaction density on slip system α, which relates the increase in immobile dislocation density due to junction formation on the slip system relative to the decrease of mobile dislocation density, can be defined (cf. Eq. 6) as

$$\rho^{\alpha}_{\text{intr}} = \int \dot{\gamma}^{\alpha} \left(g^{\alpha}_{\text{mnter}+} \rho^{\alpha}_{m} + \frac{g^{\alpha}_{immob+}}{b} \sqrt{\rho^{\alpha}_{im}} - g^{\alpha}_{\text{mnter}-} \rho^{\alpha}_{m} - \frac{g^{\alpha}_{immob-}}{b} \sqrt{\rho^{\alpha}_{im}} \right) dt, \tag{9}$$

which can be used to characterize the dominant interaction mechanism on the active slip systems in the crystalline material. Values of $\rho^{\alpha}_{\text{intr}} < 0$ indicate that the annihilation of dislocation-density junctions is dominant, while values of $\rho^{\alpha}_{\text{intr}} > 0$ indicate that the formation of dislocation-density junctions is dominant (Shanthraj and Zikry 2012). The normalized (by the initial immobile dislocation density) total interaction dislocation density at a nominal strain of 20 % is shown in Fig. 4c. The maximum normalized total interaction dislocation density is 1.0×10^5, which occurs in the retained austenite. This indicates that the dominant interaction mechanism in retained austenite is the formation of dislocation junctions, which results in the localization of plastic slip and hardening of retained austenite. The normalized total interaction dislocation density around the periphery of retained austenite pockets in martensite is -1.0×10^4, which indicates that the dominant interaction mechanism is the annihilation of dislocation junctions. This can soften martensite, which can render it more susceptible to shear–strain localization.

The normalized (by the static yield stress of martensite) normal stress is shown in Fig. 4d. The maximum value is 14, which occurs at the interface of martensite and retained austenite. The incompatibility of slip systems in these regions can impede dislocation density transmission, and it would result in these high local stresses. These localized areas of retained austenite with such high stresses may transform to martensite, which can relax these stress accumulations, and inhibit crack nucleation (Jacques et al. 2001). The normal stresses in the retained austenite are much lower, almost by a factor of 5.0, which is an indication of the toughness (as opposed to its strength) of the retained austenite.

4.2 Effects of parent austenite orientation

The effects of the initial austenite orientation on the inelastic deformation of the martensite–austenite microstructures have also been investigated. The martensite blocks and retained austenite volume fractions and distributions are the same as before, but the parent austenite Euler angle orientations were varied. Two cases were investigated, one case with a low initial Euler angles of (2°, 4°, 8°), and a second case with a high initial Euler angles of (15°, 25°, 35°).

The contours for the normalized mobile dislocation densities and the normalized immobile dislocation densities corresponding to the most active slip system of martensite and retained austenite, for the Euler angles of (2°, 4°, 8°) at a nominal strain of 20 %, are shown in Fig. 5. The most active slip systems for the martensite blocks and retained austenite pockets are the same as those for the Euler angles (0°, 0°, 0°). For the most active slip system in martensite of $(\bar{1}12)[1\bar{1}1]$, the maximum normalized mobile dislocation density is 2.8×10^7, which is higher than the cube orientation case by approximately 40 %. The maximum normalized immobile dislocation density, for the low angle case, is 2.4×10^5, which is 33 % higher than the cube case. These larger immobile and mobile dislocation

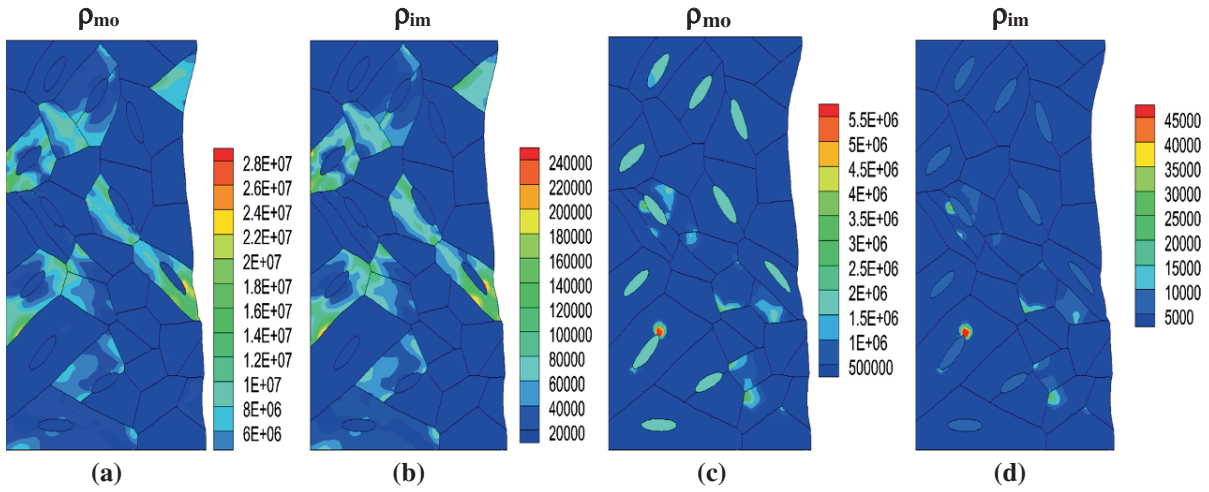

Fig. 5 Mobile and immobile dislocation densities at a nominal strain of 20 % for the low Euler angle case of (2°, 4°, 8°) for, **a**, **b** most active slip system in martensite $(\bar{1}12)[1\bar{1}1]$, **c**, **d** most active slip system in retained austenite $(1\bar{1}1)[011]$

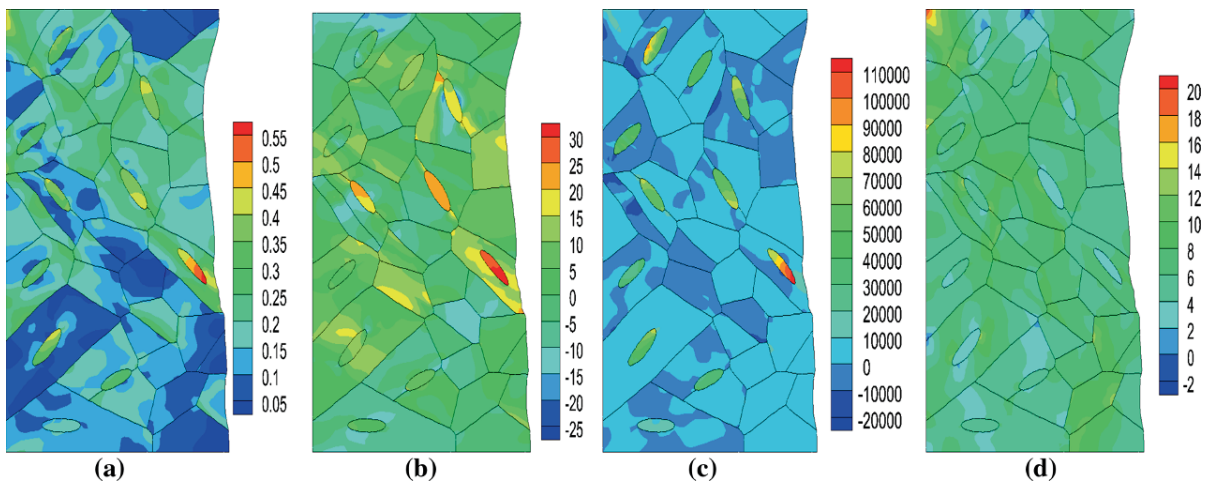

Fig. 6 Behavior at a nominal strain of 20 % for the low Euler angle case of (2°, 4°, 8°), **a** shear slip, **b** lattice rotation, **c** normalized total interaction dislocation density, **d** normalized normal stress

densities are obviously due to the higher incompatibilities of the slip-systems between the retained austenite and the martensite blocks.

The accumulated plastic slip, lattice rotation, and normalized total interaction dislocation density, for the low Euler angle case at a nominal strain of 20 %, are shown in Figs. 6a–c. The maximum plastic slip is 0.55, and the maximum normalized total interaction dislocation density is 1.1×10^5, which is 10 % higher than the cube orientation case. The maximum lattice rotation is 30°, which is higher than the cube orientation case by approximately 20 %. These maximum values occur in the retained austenite regions adjacent to the interface of martensite and retained austenite (Figs. 6a–c). The higher positive values of the total interaction dislocation density indicate that more dislocation density junctions can form, which results in the hardening of the retained austenite. The maximum normalized normal stress is 20 (Fig. 6d), and these high stresses are due to the slip system incompatibility between the retained austenite and martensite.

The contours for the normalized mobile dislocation densities and the normalized immobile dislocation densities corresponding to the most active slip system of

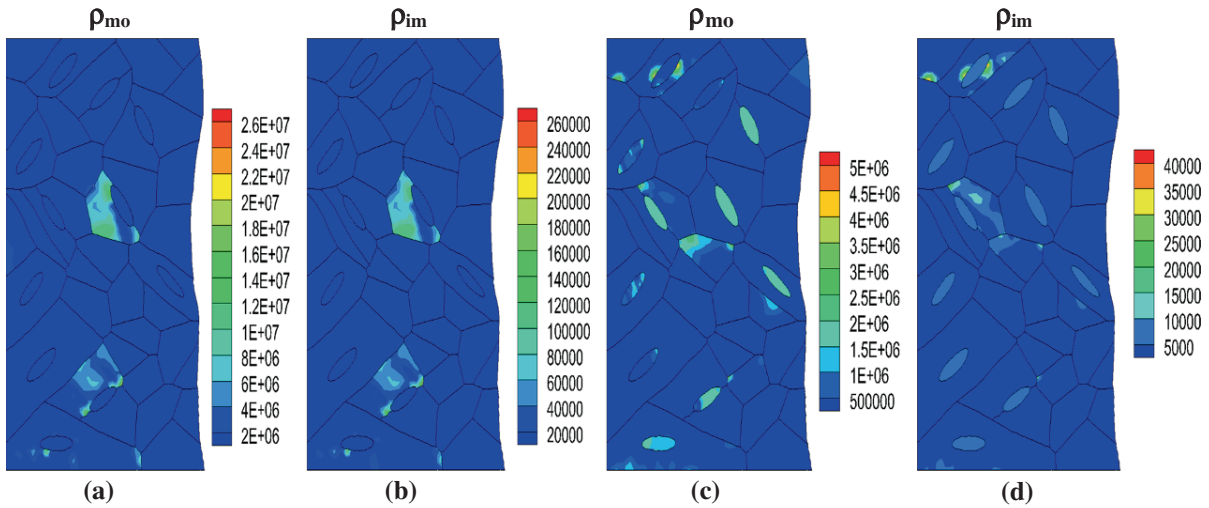

Fig. 7 Mobile and immobile dislocation densities at a nominal strain of 20 % for the high Euler angle case of (15°, 25°, 35°) for, **a**, **b** most active slip system in martensite $(2\bar{1}1)[11\bar{1}]$, **c**, **d** most active slip system in retained austenite $(1\bar{1}1)[\bar{1}01]$

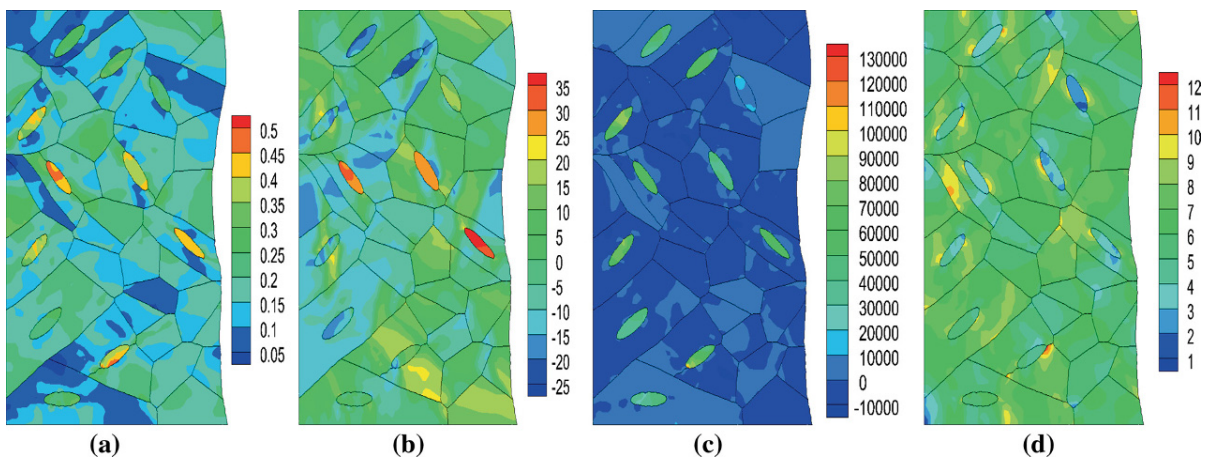

Fig. 8 Behavior at a nominal strain of 20 % for the high Euler angle case of (15°, 25°, 35°), **a** shear slip, **b** lattice rotation, **c** normalized total interaction dislocation density, **d** normalized normal stress

martensite and retained austenite, for the high Euler angle case at a nominal strain of 20 %, are shown in Fig. 7. The most active slip system in martensite is $(2\bar{1}1)[11\bar{1}]$, and the most active slip system in the retained austenite pockets is $(1\bar{1}1)[\bar{1}01]$, which are not the same as the cube orientation case and the low Euler angle case.

Some of $\{110\}<111>$ slip systems in martensite, which are compatible with f.c.c. austenite slip systems, are highly active, as indicated by the high immobile and mobile dislocation densities. This compatibility can result in plastic accumulation as shown in Fig. 8a. The accumulated plastic slip, lattice rotation, normalized total interaction dislocation density, and normalized normal stress, for this high Euler angle case at a nominal strain of 20 %, are shown in Fig. 8. The accumulated plastic slip is uniformly distributed in martensite, and the maximum value is 0.5. The maximum lattice rotation is 35°, and it occurs in the retained austenite pockets (Fig. 8b). The normalized total interaction dislocation densities have an average value of 5.0×10^4 in the retained austenite pockets, and -1.0×10^4 in the martensite blocks (Fig. 8c). This indicates that fewer dislocation-density junctions form in retained austen-

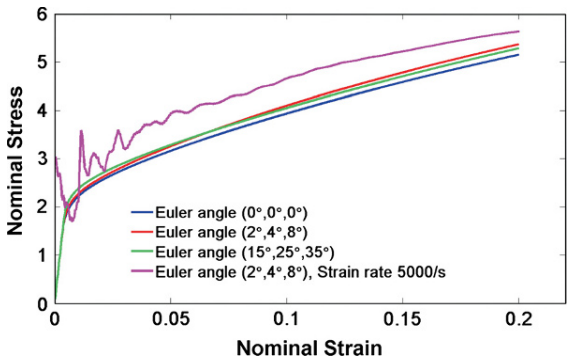

Fig. 9 Nominal stress–strain curves for different parent austenite orientations

ite, and more dislocation-density junctions are annihilated in martensite, which can result in shear strain localization and soften the material. The maximum normalized normal stress is 12, and occurs in the martensitic blocks. The stresses are significantly lower than that of the cube orientation by approximately 14 %, and 40 % lower than the low Euler angle case. These changes in behavior are due to the compatibility of the slip systems and the domination of annihilation processes due to dislocation-density interactions at the interface of martensite blocks and retained austenite pockets.

4.3 Dynamic behavior

In this section, we investigate the dynamic behavior of martensitic steel with retained austenite. The martensite blocks and retained austenite volume fractions and distributions are as before, and the parent austenite low Euler angle orientation of (2°, 4°, 8°) was used. A displacement load at a steep slope, such that a nominal strain rate of 5,000 s^{-1} results along the tensile axis (Fig. 1). The nominal strain-rate was obtained by scaling the nominal strain over an appropriate time-scale. The nominal stress–strain curves are shown in Fig. 9. The coscillations at high strain rate occur due to stress wave reflection along the free and fixed boundary.

The accumulated plastic slip, lattice rotation, normalized temperature (normalized by the martensite melting temperature of 1,700 K) and the normalized normal stresses, at a nominal strain of 10 %, are shown in Fig. 10. The maximum plastic slip is 0.4, and it occurs within the retained austenite pockets. The large values of plastic slip are as a result of geometrical softening and thermal softening. The geometrical softening occurs due to large lattice rotations (Fig. 10b). The maximum value of lattice rotation is 25°. In addition to the geometrical softening, adiabatic heating results in a thermal accumultion (Fig. 10c). It is assumed in our approach that the thermal accumulation is due to adiabatic heating, and the rate of change of temperature is obtained from the balance of energy as

$$\rho c_p \dot{T} = \chi \sigma'_{ij} D^p_{ij}, \tag{10}$$

where ρ is the mass density, c_p is the specific heat capacity, and χ is the fraction of plastic work transformed to heat energy, and is chosen as unity, which

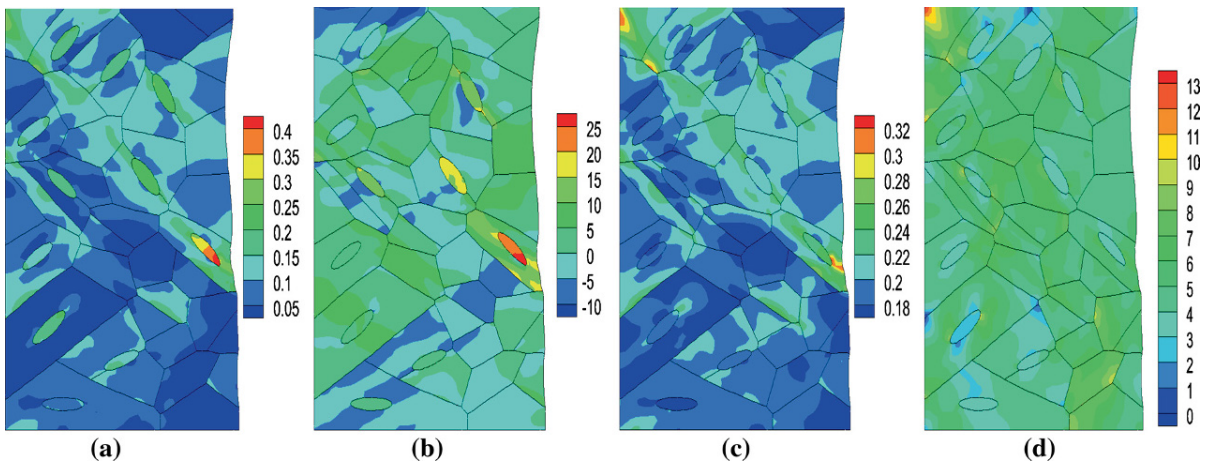

Fig. 10 Behavior at a nominal strain of 10 % for the low Euler angle case of (2°, 4°, 8°) at a strain rate of 5,000 s^{-1}, **a** shear slip, **b** lattice rotation, **c** normalized temperature, **d** normalized normal stress

indicates that all plastic work transforms into heat energy. The highest value of normalized temperature is 0.32, and it occurs at the interface of martensite blocks and retained austenite pockets. This increase in temperature further indicates that due to adiabtic heating, the material can soften. These two softening mechanisms would lead to shear strain localization, but it is a material competition with the dynamic strain hardening, which results in large normalized normal stresses of approximately 13 (Fig. 10d).

5 Conclusions

Newly developed dislocation-density based evolution equations are coupled to a multiple-slip crystal plasticity formulation, and a framework is established that relates immobile and mobile dislocation-density evolution to austenitic and martensitic crystallographic orientations and behavior. Specialized FE methodologies were then used to investigate the effects of retained austenite, parent austenite grain orientations, and quasi-static and dynamic loading rates on deformation and failure in martensitic steel.

For the cube orientation case, the dominant slip system in martensitic blocks is $(\bar{1}12)[1\bar{1}1]$, which is incompatible with the retained austenite (f.c.c.) slip systems. The interfaces between martensite blocks and retained austenite pockets can impede dislocation density transmission, and result in local strengthening at the interface. The dominant interaction mechanism was predicted to be the formation of dislocation junctions in austenite pockets, and the annihilation of dislocation junctions around the periphery of retained austenite in the martensitic blocks. This would result in the hardening of the austenitic pockets and the softening of martensitic blocks, which can render the martensitic blocks more susceptible to shear–strain localization. Furthermore, the high local stresses at the interface of martensite and retained austenite may induce a martensitic transformation, which can relax stress accumulations and inhibit crack nucleation.

In comparison with the cube orientation case, the low Euler orientation of the parent austenite grain does not change the dominant slip system, but exacerbates the incompatibility of slip system. This resulted in an increase of dislocation densities, plastic slips, and lattice rotations, and the formation of more dislocation junctions in retained austenite. The high Euler orientation of the parent austenite grain changes the dominant slip system in comparison with the cube orientation case and the low Euler angle case. Some of $\{110\}<111>$ slip systems, which are compatible with f.c.c. austenite slip systems, were activated. This results in the annihilation of more dislocation junctions in martensite, and subsequently softens the material.

The effects of dynamic load have been investigated for a strain rate of $5{,}000\,s^{-1}$. The coupled effects of the geometrical and thermal softening accelerate shear strain localization. As the strain-rate is increased, the material strain hardens, and this is a competing effect with the thermal and geometrical softening mechanisms.

Acknowledgments Support from the Office of Naval Research through Grant N000140510097 is gratefully acknowledged.

References

Asaro RJ, Rice JR (1977) Strain localization in ductile single-crystals. J Mech Phys Solids 25:309

Ashmawi W, Zikry MA (2003) Grain boundary effects and void porosity evolution. Mech Mater 35:537

Bulatov VV, Hsiung LL, Tang M, Arsenlis A, Bartelt MC, Cai W, Florando JN, Hiratani M, Rhee M, Hommes G, Pierce TG, de la Rubia TD (2006) Dislocation multi-junctions and strain hardening. Nature 440:1174

Byun TS, Hashimoto N, Farrell K (2004) Temperature dependence of strain hardening and plastic instability behaviors in austenitic stainless steels. Acta Mater 52:3889

Devincre B, Hoc T, Kubin L (2008) Dislocation mean free paths and strain hardening of crystals. Science 320:1745

Franciosi P, Berveiller M, Zaoui A (1980) Latent hardening in copper and aluminum single-crystals. Acta Metall 28:273

Hatem TM, Zikry MA (2009) Shear pipe effects and dynamic shear-strain localization in martensitic steels. Acta Mater 57:4558

Jacques P, Furnemont Q, Pardoen T, Delannay F (2001) On the role of martensitic transformation on damage and cracking resistance in trip-assisted multiphase steels. Acta Mater 49(1):139

Kubin L, Devincre B, Hoc T (2008a) Towards a physical model for strain hardening in fcc crystals. Mater Sci Eng A 483–484:19

Kubin L, Devincre B, Hoc T (2008b) Modeling dislocation storage rates and mean free paths in face-centered cubic crystals. Acta Mater 56:6040

Ma XP, Wang LJ, Liu CM, Subramanian SV (2012) Microstructure and properties of 13Cr5Ni1Mo0.025Nb0.09V0.06N super martensitic stainless steel. Mater Sci Eng A 539:271

Madec R, Kubin LP (2008) Second order junctions and strain hardening in bcc and fcc crystals. Scripta Mater 58:767

Moor E, Lacroix S, Clarke AJ, Penning J, Speer JG (2008) Effect of retained austenite stabilized via quench and partitioning on

the strain hardening of martensitic steels. Metall Mater Trans A 39A:2586
Morito S, Huang X, Furuhara T, Maki T, Hansen N (2006) The morphology and crystallography of lath martensite in alloy steels. Acta Mater 54(19):5323
Morito S, Tanaka H, Konishi R, Furuhara T, Maki T (2003) The morphology and crystallography of lath martensite in Fe-C alloys. Acta Mater 51(6):1789
Nakagawa H, Miyazaki T (1999) Effect of retained austenite on the microstructure and mechanical properties of martensitic precipitation hardening stainless steel. J Mater Sci 34:3901
Park ES et al (2004) Formation of reversed austenite during tempering of 14Cr-7Ni-0.3Nb-0.7Mo-0.03C super martensitic stainless steel. Metals Mater Int 10(6):521
Shanthraj P, Zikry MA (2011) Dislocation density evolution and interactions in crystalline materials. Acta Mater 59:7695
Shanthraj P, Zikry MA (2012) Dislocation-density mechanisms for void interactions in crystalline materials. Int J Plast 34:154
Shibata A, Takahashi H, Tsuji N (2012) Microstructural and crystallographic features of hydrogen-related crack propagation in low carbon martensitic steel. ISIJ Int 52(2):208
Song YY, Ping DH, Yin FX, Li XY, Li YY (2010) Microstructural evolution and low temperature impact toughness of a Fe-13%Cr-4%Ni-Mo martensitic stainless steel. Mater Sci Eng A 527:614
Thomas G (1978) Retained austenite and tempered martensite embrittlement. Metall Trans A 9A:439
Zhang K, Zhang M, Guo Z, Chen N, Rong Y (2011) A new effect of retained austenite on ductility enhancement in high-strength quenching-partitioning-tempering martensitic steel. Mater Sci Eng A 528:8486
Zikry MA (1994) An accurate and stable algorithm for high strain-rate finite strain plasticity. Comput Struct 50:337
Zikry MA, Kao M (1996) Inelastic microstructural faulure mechanisms in crystalline materials with high angle grain boundaries. J Mech Phys Solids 44(11):1765

DOI 10.1007/978-3-319-04397-5_17
Reprinted from *International Journal of Fracture*, DOI 10.1007/s10704-013-9883-7

ORIGINAL PAPER

Crack nucleation from a notch in a ductile material under shear dominant loading

A. Ghahremaninezhad · K. Ravi-Chandar

Received: 28 April 2013 / Accepted: 16 September 2013 / Published online: 16 October 2013

Abstract We examine the *nucleation* of a crack from a notch under a dominant shear loading in Al 6061-T6. The specimen is loaded in nominally pure shear over the gage section in an Arcan specimen configuration. The evolution of deformation is monitored using optical and scanning electron microscopy. Quantitative measurements of strain are made using the 2nd phase particles as Lagrangian markers which enable identification of the true (logarithmic) strains to levels in the range of two. Electron microscopy reveals further that the 2nd phase particles do not act as nucleation sites for damage in the regions of pure shear deformation. The initial notch is shown to "straighten out", forming a new, sharper notch and triggering failure at the newly formed notch. Numerical simulations of the experiment, using the conventional Johnson–Cook model and a modified version based on grain level calibration of the failure strains, reveal that it is necessary to account for large local strain levels prior to the nucleation of a crack in order to capture the large deformations observed in the experiment.

A. Ghahremaninezhad · K. Ravi-Chandar (✉)
Center for Mechanics of Solids, Structures and Materials,
The University of Texas at Austin, Austin,
TX 78712-0235, USA
e-mail:fracture@mail.ae.utexas.edu

Present affiliation:
A. Ghahremaninezhad
Department of Civil, Architectural,
and Environmental Engineering, University of Miami,
Coral Gables, FL 33146, USA

Keywords Fracture · Damage · Voids · Grain-level strain measurement

1 Introduction

The problem of ductile failure under shear has been addressed in numerous studies in the literature. Through the early investigations of Orowan (1948), Tipper (1949), Rogers (1960), Puttick (1960), and others, it is now established that failure occurs initially as a concentration of shear localization, followed by the nucleation of a large number of voids within this shear zone and their eventual coalescence. However, such failures were typically examined and observed in specimens with a triaxiality value—the ratio of the mean stress to the equivalent—in the range of 0.5–1. For pure shear loading, it is not evident how eventual failure occurs; the inability to trigger the localization that precedes failure under pure shear (low triaxiality) is rather well-known. Recently, Bao and Wierzbicki (2004), Barsoum and Faleskog (2007), Beese et al. (2010), and others have investigated failure of structural materials under conditions of high shear and low triaxiality. While the earlier results of Hancock and Mackenzie (1976) and Johnson and Cook (1985) reported a monotonic increase in the strain-to-fracture with decreasing triaxiality, the results of Beese et al. (2010) and Barsoum and Faleskog (2007) indicate a nonmonotonic dependence, with a cusp at different triaxiality levels for different materials and a drop in the strain-to-failure at lower triaxialities.

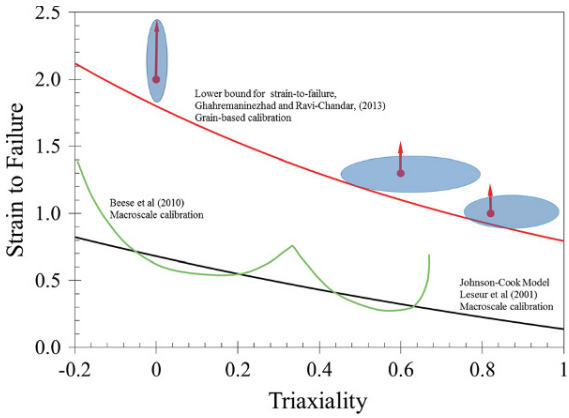

Fig. 1 Variation of strain-to-failure with triaxiality. The *blue* ellipses indicate the range of strain levels obtained from grain level measurements without damage at the levels of triaxiality indicated. Based on these, a lower bound estimate for the failure strain is suggested by the *red line*

In recent papers, (Ghahremaninezhad and Ravi-Chandar 2012, 2013; Haltom et al. 2013), we explored the appearance of very large deformations at the level of grains without development of damage in the form of voids in Al 6061-T6 alloy. Nevertheless, we reported that the final failure surface had a dimpled appearance and hence must have formed through the nucleation, growth and coalescence of voids in a localized manner. Based on these observations and measurements, a lower-bound strain-to-failure criterion that has an exponential dependence on the triaxiality as predicted by the cavity growth model of Rice and Tracey (1969) was proposed for this material, as shown in Fig. 1. The lower bound estimate of the strain-to-failure was fitted with an exponential of the form $\varepsilon_f = D_2 \exp(D_3 \sigma_m/\sigma_e)$, where σ_m is the mean stress and σ_e is the effective stress; the parameters of the model for Al 6061-T6 were found through a simple fitting procedure: $D_2 = 1.8$ and $D_3 = -0.82$. In contrast, the calibration by Lesuer et al. (2001), of the Johnson–Cook (JC) model $\varepsilon_f = D_1 + D_2 \exp(D_3 \sigma_m/\sigma_e)$ based on macroscopic strain measurements yielded the following parameters: $D_1 = -0.77$, $D_2 = 1.45$ and $D_3 = -0.47$; this failure curve is also plotted in Fig. 1. The calibration of the modified Mohr–Coulomb criterion of Beese et al. (2010) for the same Al 6061-T6 alloy is also shown in Fig. 1 for comparison. It is clear that the estimate of the lower bound for the strain-to-failure provided by Ghahremaninezhad and Ravi-Chandar (2013) is significantly higher than the traditional JC model and the modified Mohr–Coulomb model; however, it should be noted that the lower bound can be used appropriately only when the length scales at which this criterion is used are on the order of the grain size.

In this article, our main objective is to examine the predictive ability of the proposed grain level failure criterion. We accomplish this through an example dealing with crack nucleation from a notched specimen under dominant shear loading. Specifically, we consider an Arcan specimen with a notch machined by electric discharge machining to have a notch angle of $\sim 90°$ and a notch radius of $\sim 160\mu$m, subject it to a pure shear loading in the Arcan fixture, and follow the development of deformation and failure in the notch region as well as in the interior through interrupted tests. There is a body of literature that deals with a similar problem, where mixed-mode fracture initiation in specimen configurations that are dominated by the stress concentration associated with a single crack is considered. For example, Aoki et al. (1990), explored ductile failure under mixed mode loading in Al 5083-O aluminum alloy; mixed mode loading was applied in a compact-tension-shear configuration to explore crack initiation from a fatigue pre-crack. They found that one region of the pre-crack blunted, while the other region sharpened; the extent of such blunting was characterized by measuring the zone in which stretching occurred and shown to depend on the degree of mode mixity; furthermore, crack initiation was reported from the blunted crack tip since the triaxiality was higher in this region. Ghosal and Narasimhan (1994, 1996) performed finite element studies of crack response under mixed mode loading using a Gurson model that incorporates material damage, and examined various aspects of the problem, such as the debonding of inclusions, the deformation and sharpening of the notch, and the dependence of the fracture toughness on the mode mixity. They also report the separation of the pre-crack into blunted and sharpened corners, with significant void growth occurring in the blunted region. The goal of the studies cited above and others was mainly to investigate the dependence of the fracture toughness, characterized in terms of the J-integral, on mode-mixity. In contrast, the present paper is concerned with examining the evolution of deformation and the initiation of failure in specimens without an initial crack; we examine the development of deformation and damage in Al 6061-T6 leading up to the nucleation of a crack from a notch under shear dominant loading. In particular, we are interested (i) in

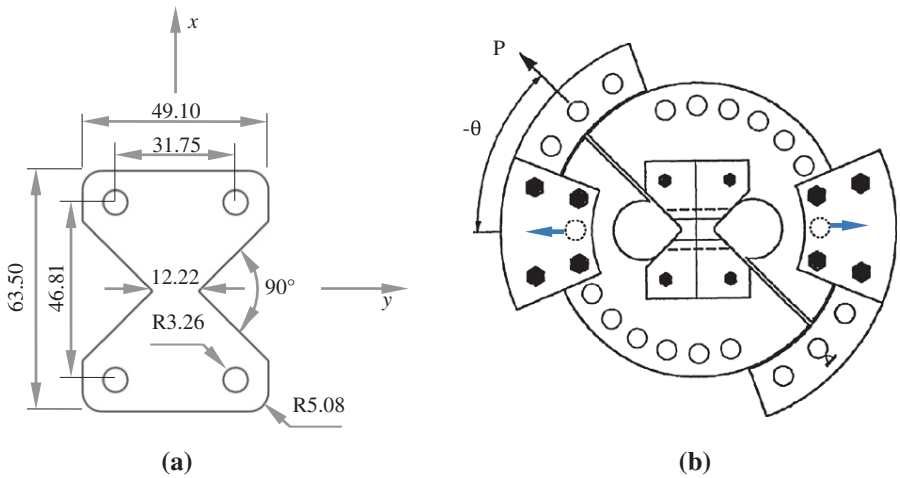

Fig. 2 **a** Geometry of the Arcan specimen (all dimensions are in mm). Rolling direction is along the x-direction. **b** Geometry of the fixture used in the modified Arcan tests (reproduced from Hung and Liechti 1999). The two *arrows* from the *dotted circles* indicate pure shear loading; other pairs of holes result in shear plus compression or tension

understanding the role of the second phase particles in generating damage, (ii) in determining how cracks are nucleated from the notch and (iii) finally in examining the use of the failure criterion presented in Ghahremaninezhad and Ravi-Chandar (2013), to simulate the nucleation of a crack.

This paper is organized as follows: experiments under pure shear loading of a notched Arcan-type specimen are described in Sect. 2. These experiments provide macroscopic measurements of the load-elongation response along with grain level strain measurements in order to provide local strain estimates. The nucleation/onset of crack initiation is also examined in this section through optical and scanning electron microscopy of interrupted tests. The response of the notched specimen is modeled in Sect. 3 using the finite element code ABAQUS, endowed with a JC model as calibrated by Lesuer et al. (2001) and also modified with the results of Ghahremaninezhad and Ravi-Chandar (2013). The main conclusions are summarized in Sect. 4.

2 Experimental results

The Arcan specimens were cut from the same rolled sheet stock (2.44 mm thick) used for the tension and shear tests described in Ghahremaninezhad and Ravi-Chandar (2012, 2013), with the rolling direction oriented along the x-direction as indicated in Fig. 2a. These specimens were loaded in a modified Arcan fixture shown in Fig. 2b. The grains are platelets that are about 14 μm thick, 46 μm wide, and 39 μm long, as characterized by Ghahremaninezhad and Ravi-Chandar (2012). In an effort to examine the grain level strains during the deformation stages, without the post-test measurements reported in our prior investigations, the specimen was polished and etched *prior to testing* using Weck's etchant (details of the specimen preparation for metallographic observations are given in Ghahremaninezhad and Ravi-Chandar 2012). The nominal stress (force/initial gage cross-sectional area) versus normalized crosshead displacement[1] response curves from one pure shear test is shown in Fig. 3. Note that the beginning of the load drop in the experiment corresponds to initiation of a crack from the notch regions and that the material continues to exhibit a strain hardening response. In order to examine the evolution of deformation and failure at the grain level, the loading on this specimen was interrupted at different stages along the overall response at points indicated as A, B, and C and then taken to final failure at D; at each of these stages, optical micrographs of the x–y plane near the center of the specimen were taken at high resolution and stitched together using the Microsoft Image Composite Editor (ICE) software, to explore a large spatial domain. Damage evolution in another specimen at approximately the same load-

[1] The compliance of the loading system has not been removed; therefore the slope of the linear region is significantly smaller than the slope of the specimen alone and the normalized displacements should not be compared directly to those determined in the simulations reported later in Sect. 4. Macroscopic scale digital image correlation (DIC) could provide adequate measurements to prescribe realistic boundary conditions arising from the loading system compliance, but this was not considered essential to the objectives of the present work.

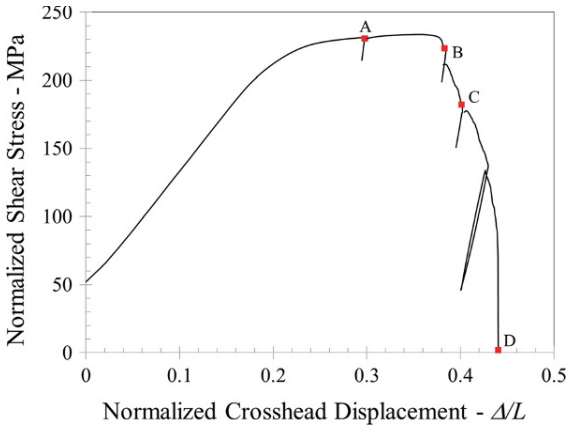

Fig. 3 Normalized shear stress (force/initial gage cross-sectional area) versus crosshead displacement (Δ) normalized by the gage length ($L = 12.22$ mm: see Fig. 2a) response of Arcan specimen. Points A and B correspond points when the specimen was unloaded and examined under an optical microscope. The point D corresponds to complete failure of the specimen. Note that the load drop near the point B is due to initiation of crack growth from the notch tips and not due to material softening

ing stages was investigated through scanning electron microscopy of the region near the notch tip. The optical microscopy focused on regions in between the two notches, while the scanning electron microscopy was restricted to the region of the notch, where very large strains are expected. These examinations eliminate polishing and etching effects and allow unhindered examination of the actual deformation and failure mechanisms; specifically, the actual strains can be determined by following the same set of grains over many deformed stages of the specimen.

2.1 Grain level measurement of local deformation

Real-time measurements of grain level deformation using digital image correlation (DIC) methods (Héripré et al. 2007; Carroll et al. 2013), X-ray tomography (Bay et al. 1999; Limodin et al. 2011), and other methods are becoming increasingly simpler to use and interpret; nevertheless, the strain levels encountered in almost all of these applications is quite small. For example, Héripré et al. (2007) examined polycrystalline plasticity and measured strains in the range of 0.1 using a square grid of gold deposited on the material; Carroll et al. (2013) used DIC to determine local strains in the vicinity of a fatigue crack; the strains measured in the vicinity of the crack were only about 0.01 even when the fatigue crack was extended. In contrast, the equivalent strains encountered in elastic-plastic fracture under monotonic loading are at least one to two orders of magnitude greater; true strains in the range of two have been estimated in some cases prior to nucleation of damage and fracture (see Fig. 1). In this range of strains, DIC typically encounters difficulties arising from image registry. Therefore, a more direct method of strain measurement is desirable. Here we develop a tracking scheme that is perhaps tedious, but still manageable; we present measurements from the interrupted tests, but in principle, this can be accomplished continuously if a loading stage attached to a microscope is used.

The principle of the method is quite simple: we follow the movement of the 2nd phase particles as a function of the global deformation and extract estimates of the local strain; it turns out that human eyes are surprisingly good in pattern recognition, and one can track the same set of particles even after substantial deformation, and sometimes, even when some of the particles break up into smaller pieces. The 2nd phase particles in the sheet are typically about 2 μm wide and about 10 μm long (see Figure 14 of Ghahremaninezhad and Ravi-Chandar 2012). A high magnification image of one such particle at Stage *B* of the loading is shown in Fig. 4; the initial orientation of the particles is such that the long direction of the particle is oriented

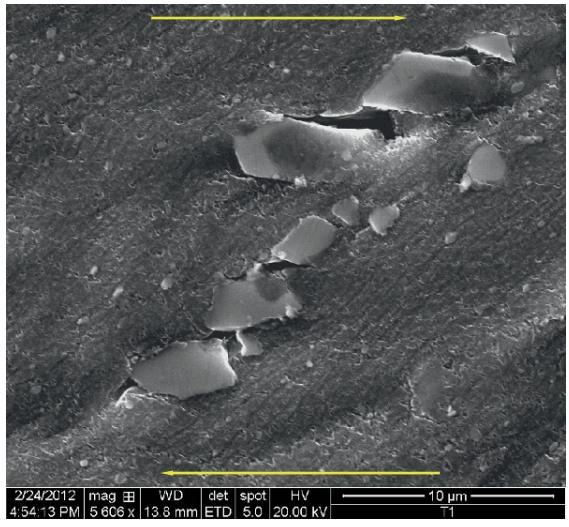

Fig. 4 SEM image of a 2nd phase particle after loading to Stage A. The long particle was initially aligned along the rolling direction (vertical), but broke into smaller fragments that rotate subsequently with the shear deformation. Note that there is no cavity formation; furthermore, the "gap" between the broken pieces is "filled" by the matrix flowing into it during the shearing deformation

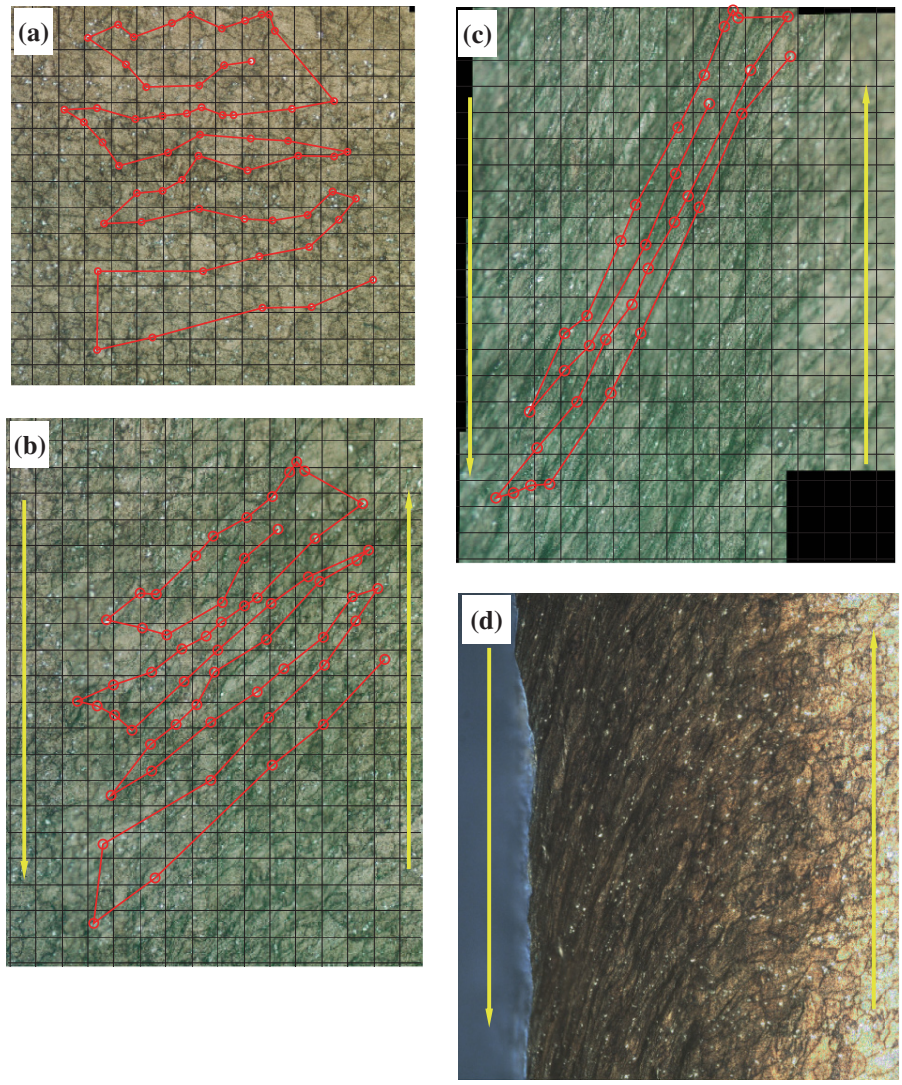

Fig. 5 Optical micrographs of the midsection of the specimen **a** undeformed, **b** Stage A, **c** Stage B and **d** Stage D. The direction of shear is indicated by the *yellow arrows*. 58 particles identified between the undeformed specimen and Stage A and 33 particles identified between Stages A and B are tracked by the *red circles* and *lines*. The image in **d** was obtained with a z-focus microscope to be able to focus on the significantly deformed specimen

with the rolling direction (vertical in this image). From the shear loading in the direction indicated by the yellow arrows, the particle rotates, and breaks into many smaller pieces. Furthermore, the pieces exhibit relative rotation with respect to each other indicating that after breaking, they continue to rotate with the deformation of the matrix. There is no indication of cavitation or other types of damage; in fact, it appears that the matrix simply "fills in" any gaps caused by the relative movement of the broken particles. Furthermore, there appears to be no failure mechanism that is generated at this scale under such shear loading; we note that failure of the specimen occurs eventually as a result of the deformation near the notch, which occurs at a triaxiality that is quite different from the pure shear state as we describe later.

Figure 5 shows four images of a region near the center of the specimen taken with an optical microscope; the grid lines superposed on the images represent squares that are 50 μm to a side. All of these images were obtained by taking multiple high resolution images at 200$\times$ magnification and stitching them together to be able to observe a much larger region. Figure 5a–d correspond to the (a) unstrained specimen, (b) loading Stage A, (c) loading Stage B, and (d) final failure. The images were then viewed under their original magnification in order to identify the same set of 2nd phase particles in each image; 58 particles were iden-

tified between the unstrained image and Stage A, and are indicated by the open red circular symbols and connecting lines in Fig. 5a, b. However, only 33 of these particles could be tracked between loading Stages A and B, mainly because the out of plane deformation of the specimen caused defocusing of some of the particles and made it difficult to identify the particles; the identified particles are indicated by the open red circular symbols and connecting lines in Fig. 5c. Finally, it was nearly impossible to identify the corresponding points between Stage B and final failure; however, if we draw tangent lines to the grain boundaries, the shear angle γ can be determined easily as indicated by the red line in Fig. 5d. This figure also indicates that the region of high shear strains is localized to an extremely narrow zone along the center of the specimen. Comparing Fig. 5a–c, it is evident that the positions of the 2nd phase particles changes significantly, but their relative order does not change, strongly suggesting that these particles are simply carried by the flow of the matrix aluminum, and therefore they represent good Lagrangian markers to evaluate the strain variation averaged over a length scale much larger than the particle size.

The deformation of the Arcan specimen can be written as $x_1 = X_1 + kX_2$, $x_2 = aX_2$, where $k = \tan\gamma$ is the amount of shear, and a is the stretch perpendicular to the direction of shear; then, the principal stretches may be estimated directly from the deformation observed in Fig. 5. These measurements indicate that the maximum principal true (logarithmic) strain (averaged over a $500 \times 500\,\mu\text{m}^2$ square) increases from 0.41 to 0.93 to 2 as we go from Stage A to Stage B to final failure. These strain levels are commensurate with the values shown in Fig. 1 for pure shear conditions. However, while the strain levels in that figure were inferred from a statistical estimate of changes in the grain dimensions, the present measurements represent a direct measurement of the strain using the *same* 2nd phase particles as Lagrangian markers. While we have only identified the average strain over a small region, the method used here clearly has the potential to resolve strains on a much smaller level, approaching grain level identification; this, however, requires that the experiments be performed in a scanning electron microscope that can image over large depth variations or in an optical microscope with a z-focus capability to reconstruct full three-dimensional image of the surface. These approaches are currently under investigation; the image in Fig. 5d was obtained using the z-focus capability of the microscope to focus over the entire field of view.

In order to explore the origins of failure under predominantly shear loading, the unstrained specimen and deformed specimens at Stages A and B, were examined in a scanning electron microscope. These observations are discussed in the next section.

2.2 Scanning electron microscopy of deformation

The development of deformation near the notch tip is rather spectacular, and is easily tracked using scanning electron microscopy. Figure 6a shows SEM images of the regions near one of the notches corresponding to the undeformed configuration. The line segment *a–b–c–d* of total length 1,400 μm identifying the edge of the specimen through the notch tip is marked by the green line in this figure for easy visual identification of the deformation. The spatial position of two additional points on the surface of the specimen near the notch (points e and f) are also marked in this figure. The deformed shape of the line *a–b–c–d* as well as the current position of the points e and f are traced through the two stages of deformation A and B in Fig. 6b, c. The specimen was initially polished to be flat and etched to reveal the grain boundaries; in the secondary electron imaging mode, the contrast in the undeformed SEM images is low. Nevertheless, some features resulting from an uneven coating of Au–Pd on the surface are visible even in the undeformed specimen, and these features are used to identify the deformation. As the deformation progresses, there is some out-of-plane deformation of the grains due to local three dimensionality of the deformation and stress fields, and this results in improved contrast in the last stage image shown in Fig. 6c. For ease of visualization of the deformation, the corresponding points in each image in Fig. 6a–c are traced by the solid and dashed lines with the arrows. We record the following important observations:

– Comparing Fig. 6a, b, the deformation of the line *a–b–c–d* is most striking; this line is seen to have "unfolded", with the segments *a–b* and *c–d* hardly exhibiting any changes, while the segment *b–c* has rotated significantly and elongated substantially. The EDM-cut notch at *c* has disappeared completely, and a "new", sharper notch has been created at the point *b*. Although we have only examined the notch at these two stages, it is apparent that as the "old"

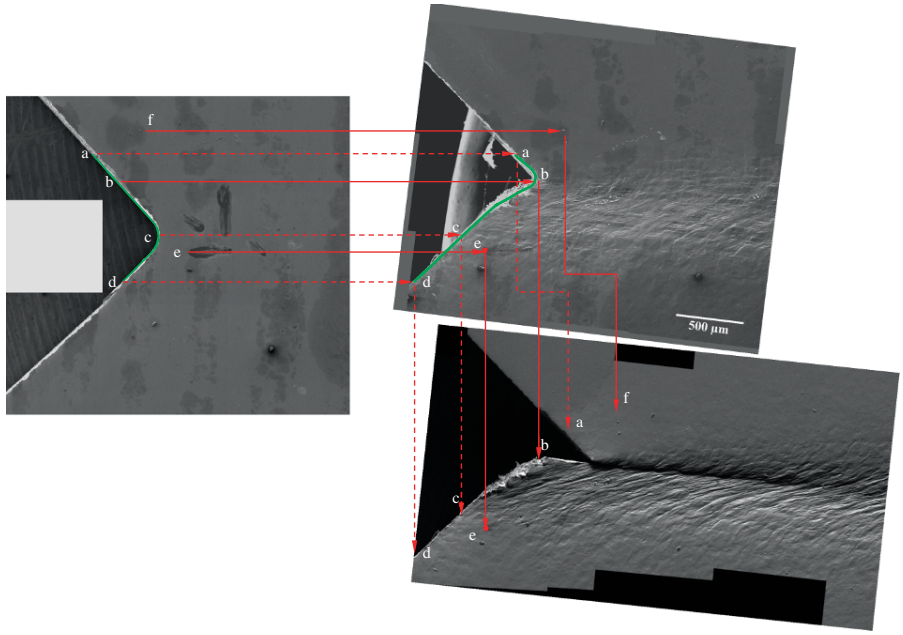

Fig. 6 SEM images of the **a** undeformed, **b** Stage A and Stage B specimens. The deformation of the notch region and selected points on the specimen is tracked between the three images. The *green line* labeled *a–b–c–d* identifies the initial notch. With deformation, it is seen that the initial notch straightens out and a new notch is formed. The contrast in **c** arises from nonuniform out-of-plane deformation of the grains. A crack is nucleated near the point *b* after significant plastic deformation

notch at *c* straightened out, the "current" notch must have moved progressively from the point *c* to the point *b*.

- It is evident that the deformation along the line *a–b–c–d* is not uniform; the deformation of each line segment can be estimated using the change in the length of each line segment. While the true logarithmic strains in segments *a–b* and *c–d* are about −0.05 and 0.13 respectively, segment *b–c* experiences rather large extensional strains, as high as 0.31.
- Segment *b–c* lies on the side of the band of high deformation extending in *x*-direction across the gage length of the specimen. This band is visualized easily in SEM images in Fig. 6c as the region of high contrast, arising from large out-of-plane-deformation at the level of the grains.
- At some loading between Fig. 6b, c, a crack was nucleated between the segment *a–b*, but very close to the point *b*, and grew towards the opposite notch region. A high-magnification image of the region near the crack in Fig 6c is shown in Fig. 7; the contrast seen in these images is from the fact that the random orientation of the grains relative to the plane of shear loading causes different out-of-plane displacements in each grain. Recalling that the rolling direction is vertical relative to the orientation of this image, it appears that there are large grain rotations, and grain elongations. Furthermore, there are many

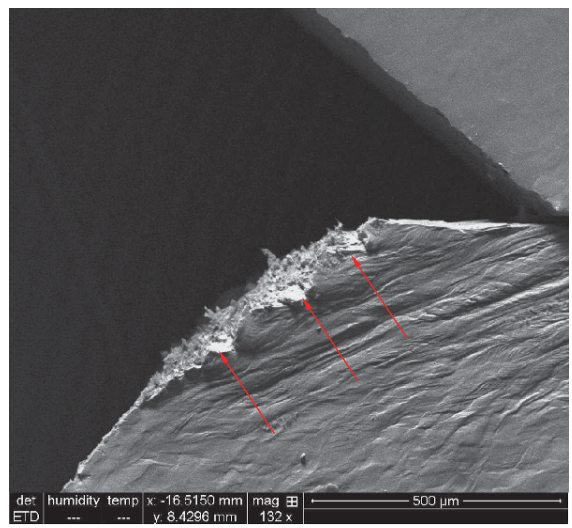

Fig. 7 A high magnification SEM image showing the crack extension from the current notch tip corresponding to the Stage *B*. The location of attempted crack nucleation at lower strain levels is marked with *arrows*. Surface relief arising from anisotropic grain deformation in the band of high deformation is easily seen in the SEM image

locations of attempted nucleation of cracks along the boundary *b–c* as identified by the red arrows in the figure.

- The crack borders on the region of highly localized shear deformation in the Arcan specimen and *does not nucleate from the original notch*. Eventually,

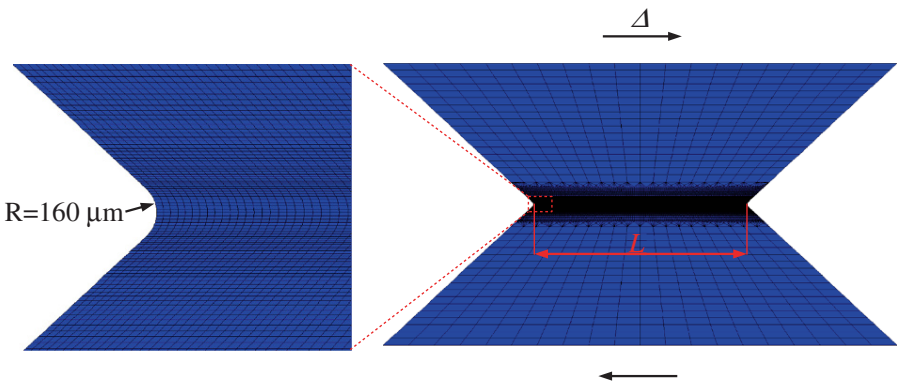

Fig. 8 **a**, **b** Schematics showing the detailed and overall mesh, respectively, used in the simulations. Fine mesh was used near the notch tip

when the crack nucleates and grows from the newly-formed notch, the material in this vicinity is quite different from the initial material in terms of its grain size, grain boundary morphology, anisotropy etc, suggesting that grain boundary morphology and texture evolution will be important aspects in dictating the onset of fracture. The stress-state is also significantly different from the initial state corresponding to the pure shear loading, as will become apparent from the numerical simulations in the next section.

These observations provide ample support to the idea that significant plastic deformation, to levels indicated by the lower bound estimate in Fig. 1, must occur prior to the onset of damage and nucleation of a crack in ductile materials such as Al 6061-T6. In the next section, we explore such crack nucleation through numerical simulations.

3 Numerical simulations

Numerical simulations of the Arcan specimen were performed in the finite element software ABAQUS/Explicit 6.11-2 with the objective of using the grain-based failure criterion of Ghahremaninezhad and Ravi-Chandar (2013) to predict the nucleation of the crack in the Arcan specimen. The geometry of the Arcan specimen was discretized using fine linear brick elements with reduced integration (C3D8R) in the gage section and gradually coarsened elements in the regions towards the outer boundaries of the specimen. The gage section consisted of 200 × 40 × 88 elements in $x \times y \times z$ directions. A mesh size of 34 × 16 × 61 μm in $x \times y \times z$ directions is used near the notch tip to explicitly represent the EDM cut notch radius, which is about 160 μm in the simulations. Since our previous studies have shown that deformation localization and failure in the polycrystalline aluminum alloy being studied here occur at a scale that is on the order of the grain size, it is expected that the finite radius of the notch will most likely influence the failure initiation in this specimen. Figure 8a, b show the detail of the fine mesh near the notch tip, and specimen geometry and overall mesh used in the simulation, respectively. The material model described in Ghahremaninezhad and Ravi-Chandar (2012, 2013) is used: specifically, the flow theory of plasticity with Hill's anisotropy is used. The Lankford parameters were measured experimentally and are given in Table 3 of Ghahremaninezhad and Ravi-Chandar (2012); the power-law hardening rule given in Eq. (2) of this article was used **only up to the Considère strain** and then extrapolated for larger strain levels through an inverse procedure where the global uniaxial response from experiments and simulations are matched (the true-stress vs strain curve is given in Figure 21b of Ghahremaninezhad and Ravi-Chandar 2012). It is recognized that an accurate prediction of material response will require the use of more complex phenomenological plasticity models where the effects of texture evolution are considered; additionally, crystal plasticity models can also provide insights into the evolution of material deformation and texture evolution in polycrystalline materials. This has been attempted recently in an effort to analyze crack tip fields in a polycrystalline material under mixed tension and shear loading (Sreeramulu et al. 2010). All these models consist of several material parameters and require lengthy material calibration programs and were not pursued in the present work.

A simple ductile failure model based on relating the strain-to-failure versus triaxiality was assumed using two calibration methods, namely, Johnson–Cook and

grain-based methods as described in detail in our previous work (Ghahremaninezhad and Ravi-Chandar 2013); these two failure criteria are shown in Fig. 1. The JC failure model is based on material parameter calibration due to Lesuer et al. (2001), which is based on strain measurements using specimen dimensions as the gage length, whereas our calibration of the grain-based failure model uses statistical measurements of strains using grain size change with respect to an intrinsic grain size as the gage length. More information on the methodology of grain based measurements is provided in (Ghahremaninezhad and Ravi-Chandar 2011, 2012, 2013). The JC failure model as implemented in ABAQUS/Explicit 6.11-2 through a cumulative damage approach is used. A linear strength degradation with a displacement to failure of $1\,\mu$m is assumed for elements at which failure initiation has been triggered, and element deletion was used to remove the elements once load carrying capacity of elements falls below 25% of that in undamaged condition; we note that while this may not be the best procedure to reach final failure, our main objective is to show that plasticity alone is responsible for much of the response of the structure and that failure processes are initiated very late, and then progress very rapidly. A semi-automatic mass scaling scheme with a minimum time increment of 10^{-6} s was used to increase the computational efficiency. Macroscopic quasi-static in-plane shear loading was simulated by specifying a constant velocity in y-direction at the top surface and zero velocity at the bottom surface. The out-of-plane displacements (z direction) at both end surfaces were fixed at zero.

A brief discussion of the length scales involved in using the failure criteria is important prior to dealing with the results of the simulations. The impetus to define strain-to-failure as a material property is driven primarily by the need to use such a parameter in numerical simulations of structural reliability. The dilemma one must face immediately is that it is essential to include a length scale in constructing such a definition; this has been considered in the literature in different ways. The most common method involves measuring the reduction in the cross-sectional area at the location of failure and using this measurement to define the strain-to-failure. This sets the scale to the characteristic transverse dimension, which is typically on the order of a few millimeters. The JC model is typically characterized at this scale. However, the model is considered a material failure model and typically reported (and commonly used) without restrictions on the size of the discretization. If we admit that they are to be used only at the scale of calibration—specimen thickness—then elements should be about a few mm large; one cannot even represent the notch in our problem with such an element size! Our simulations will demonstrate that a macroscopic failure model calibrated at the thickness scale cannot reproduce the experimentally observed behavior. A more recent method that has been used to calibrate failure models is a hybrid technique where a numerical simulation is performed to mimic the global response of the specimen and the strain at the critical element in the simulation mesh is taken to be the strain-to-failure; in this case, the failure strain is considered to be calibrated to this specific mesh size. The works of Barsoum and Faleskog (2007); Beese et al. (2010) and Dunand and Mohr (2010) fall in this category. The main drawback in this method is that it is very sensitive to the constitutive model used in the numerical simulations; as can be seen from Fig. 1 for the Al 6061, the mesh-based calibration of the failure strains does not differ too greatly from the calibration of Lesuer et al. (2001). In contrast, our calibration arises from direct measurements of the local strain on the verge of final failure, and is associated with a length scale that is of the order of a few grains.

3.1 Prediction of Johnson–Cook and grain-based failure models

The variation of the nominal stress (force/gage cross sectional area) versus normalized boundary displacement (Δ/L) obtained from the simulations using the JC and grain-based methods are shown in Fig. 9. L is the distance between the notches as shown in Fig. 8 and is used as the "gage" length. A sequence of images showing the equivalent plastic strain and triaxiality contours near the notch for the JC model calibrated by Lesuer et al. (2001) are shown in Fig. 10; these correspond to the stage B and crack initiation as marked in Fig. 9.Another sequence of images showing the equivalent plastic strain and triaxiality contours near the notch tip for the grain-based failure model is shown in Figs. 11 and 12, corresponding to stages B–C–D–E marked by diamond symbols in Fig. 9. The location of the notch tip at mid thickness of the specimen in the undeformed and deformed configurations is marked with a white square in Figs. 10a, b and 11 and by the black squares in

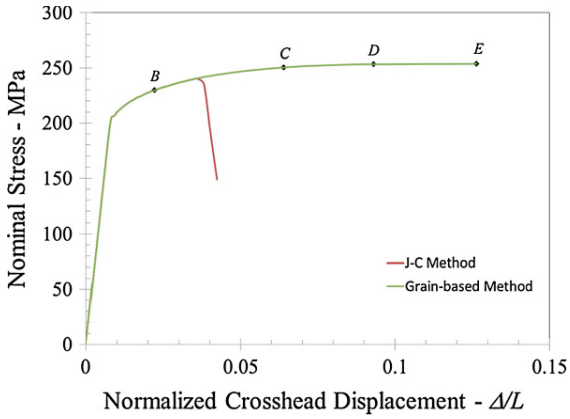

Fig. 9 Nominal Stress (force/cross sectional area) versus normalized boundary displacement (Δ/L) from simulation using grain-based and J–C methods. Points *B*–*E* denote nominal displacement levels at which the plastic strain and triaxiality are evaluated for display in Fig. 12. The point *E* corresponds to the onset of crack nucleation observed in experiments

Fig. 12 to aid in tracking deformation in the vicinity of the notch. The simulation based on the grain-based failure criterion exhibits many features that are observed in the experiments, while the JC criterion triggers failure prematurely and does not allow significant deformation near the notch region:

– It can be seen from Fig. 9 that the JC method predicts a premature failure in the specimen at a very small overall displacement in comparison to that observed in the experiments. From Fig. 10, it is seen that the notch has not deformed (or "unfolded") as observed in the experiments. The strains developed in the interior of the specimen are also significantly smaller than observed in the experiments.
– Figure 10b shows the prediction of JC method at a stage where a crack has nucleated just below the initial notch tip, and extended partially into the specimen. As can be seen from this figure, the region below the notch is where the triaxiality develops to levels of 0.5 and hence the failure strain of the JC model is reached rather early in the overall deformation. The inability of JC method to obtain a prediction of failure nucleation arises from the fact that average failure strains calibrated from measurements over a large gage length that is on the order of ten millimeters (characteristic specimen cross-sectional dimensions) is now applied

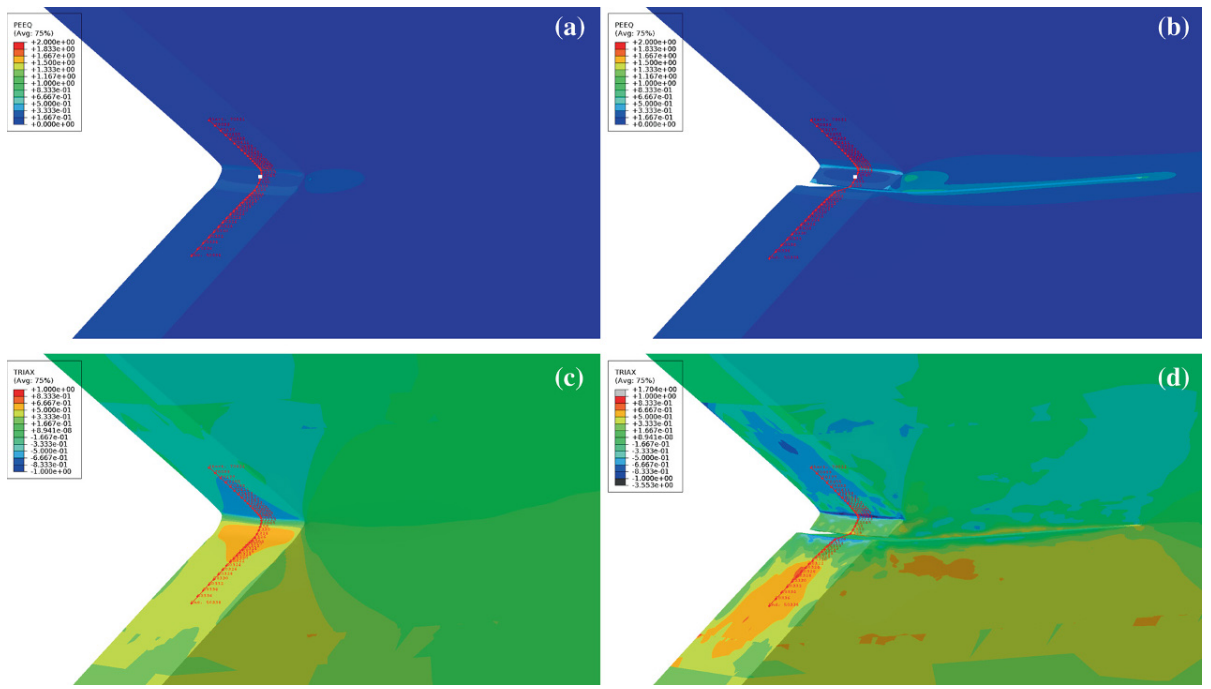

Fig. 10 **a**, **b** Plastic strain contours and deformed configurations, and **c**, **d** triaxiality contours corresponding to the stage *B* and crack initiation, respectively, as marked in Fig. 9, obtained from the JC model. The *white squares* (in **a**, **b**) indicate the location of the initial notch tip

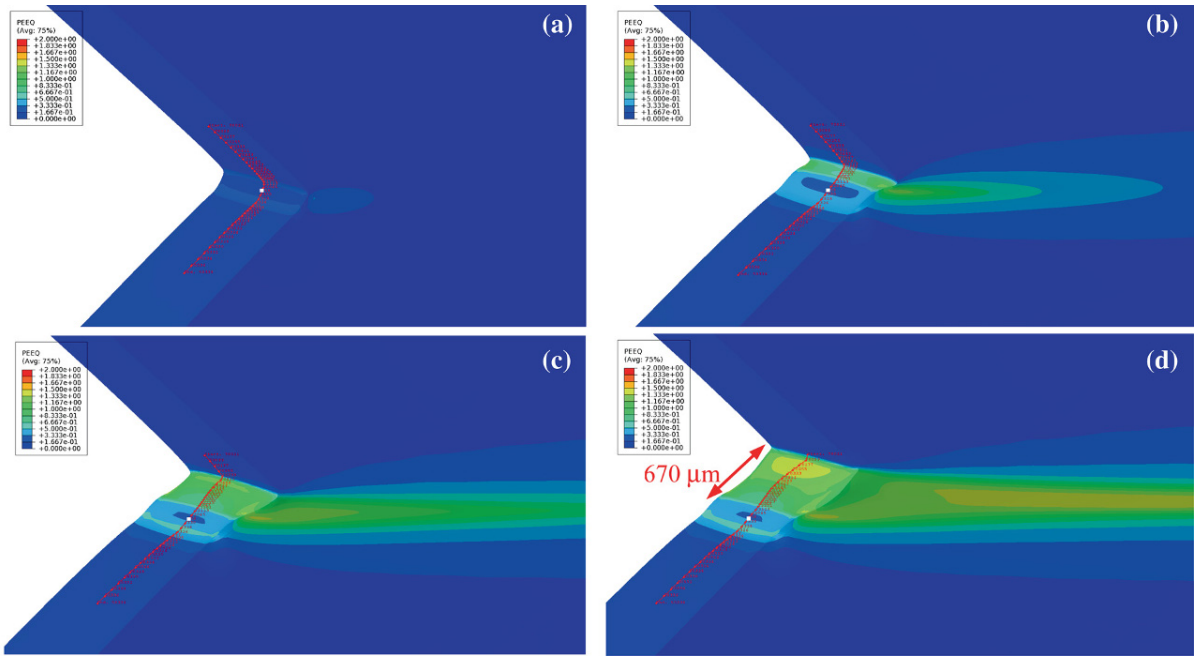

Fig. 11 **a–d** Plastic strain contours and deformed configurations at various stages of deformation (*B–E*) indicated in Fig. 9, obtained from the simulation using grain-based model. The *white squares* indicate the location of the initial notch tip

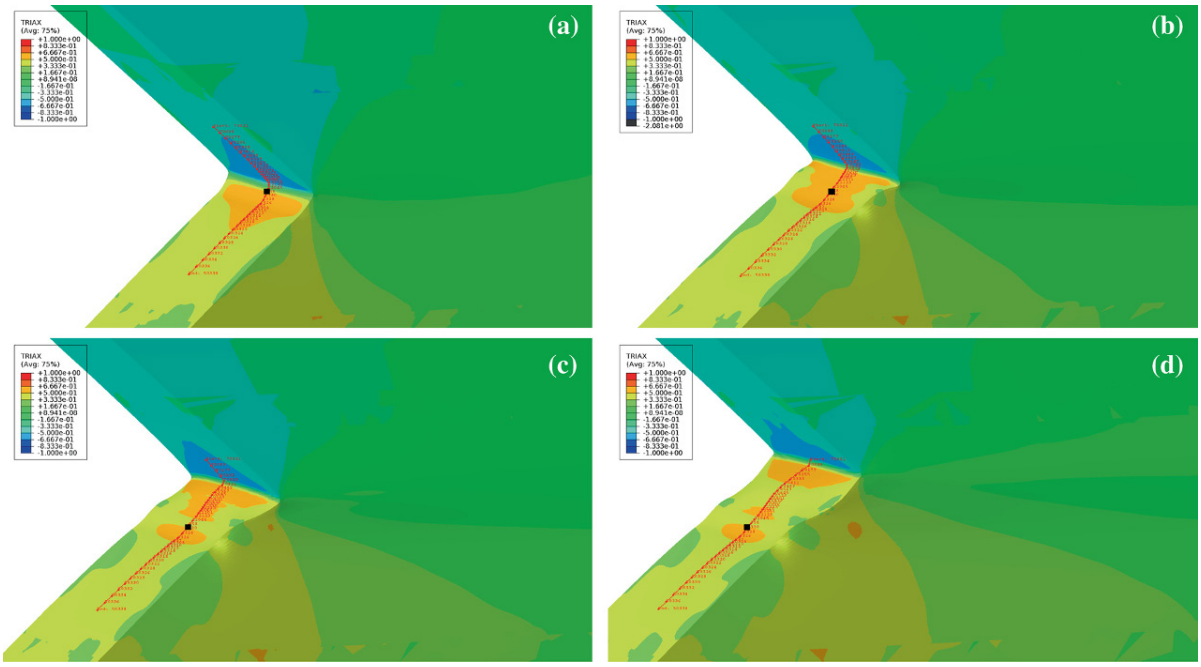

Fig. 12 **a–d** Triaxiality contours at various stages of deformation (*B–E*) as indicated in Fig. 9, obtained from the simulation using grain-based model. The *black square indicates* the initial location of the notch tip

on a scale that is a couple of orders of magnitude smaller.

– The sequence of images shown in Fig. 11 clearly indicates that the initial notch does not trigger failure when the grain-based calibration of the strain-to-failure model is used, but rather large deformation of the notch tip region occurs with the resulting "unfolding" of the notch and the development of a new notch, similar to what was observed in the experiments.
– The stage *E* in Fig. 11 corresponds to the onset of crack initiation in the experiment as determined by comparing the distance between the initial and current notch tips in the experiments; this distance was measured to be 670 μm (Line *b*–*c* in Fig. 6b). The force-displacement response from the grain-based simulation shown in Fig. 9 was terminated corresponding to the arrival of the notch at the location of crack nucleation observed in the experiment; at this point, the cumulative damage parameter was at the threshold of failure. The global response in Fig. 9 indicates a hardening behavior with no indication of macroscopic localization up to the onset of nucleation of the crack, as observed in the experiment.
– A line of length 1,850 μm drawn on the notch surface symmetrically along the middle of the thickness of the specimen as shown in Figs. 10 and 11 is used to quantify the evolution of equivalent plastic strain and stress triaxiality near the notch tip. This line is analogous to the notch line *a*–*b*–*c*–*d* that was used to identify the deformation in the experiments. First, this line also "unfolds" with overall deformation of the specimen and results in large strains of about 1.3 in the vicinity of the notch. Second, we obtain a quantitative measure of the true strain and triaxiality at points *a*, *b*, *c*, and midway between *b* and *c* (hereafter denoted as *bc*) along this line (approximating the path *d*–*c*–*b*–*a* from the experiment) during the overall deformation, and plot them in Fig. 13. It is seen from Fig. 13b that at any loading stage, the segment of the line above the current location of the notch experiences a state of compression with a triaxiality of −0.5 and the segment below the current location of the notch experiences a state of tension with a triaxiality of about 0.5. This can be readily seen from the values of triaxiality at points *a*, *c*, which lie above and below the current notch tip, respectively, during the deformation. As a point passes through the current notch tip, the stress state exhibits a rather sharp transition from compression to tension; this is demonstrated in Fig. 13b by tracking the triaxiality of the points *b* and *bc*.
– The peak plastic strain is always at the *current* notch tip; however, as the notch moves from *c* towards *b*, there is a monotonic increase in the magnitude of the peak plastic strain as the overall loading increases. Eventually, the combination of the triaxiality and plastic strain cross over the lower-bound failure envelope and crack initiation occurs. A dashed line in Fig. 13a, b marks the stage corresponding to failure as observed in the experiments. The path taken by material points at *b*, *c* and *bc* in the space of triaxiality and plastic strain is shown in Fig. 14. The triaxiality at the point *c* reaches about 0.5, but as the notch straightens out, the deformation at *c* stops at a strain level of about $\varepsilon_p = 0.3$; therefore, ductile failure is not triggered at this point. The triaxiality at the point *bc* changes rapidly as the notch approaches this

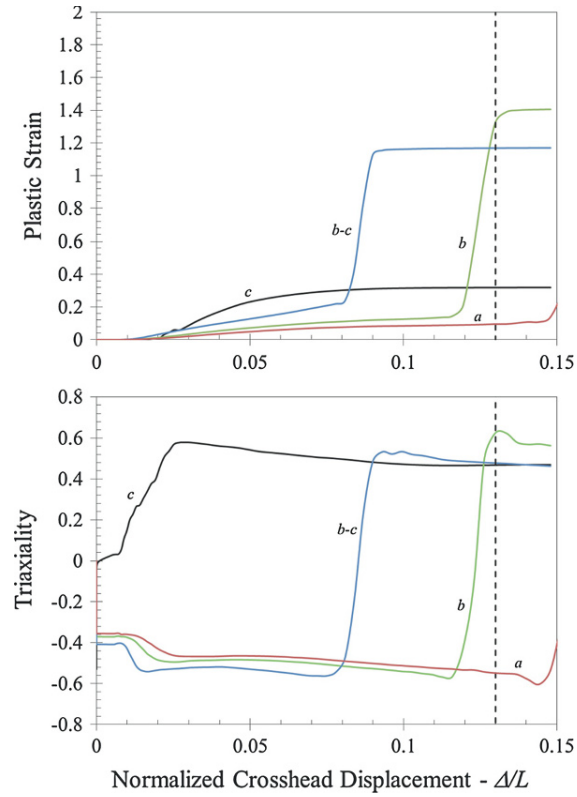

Fig. 13 Variation of (**a**) equivalent plastic strain and **b** triaxiality at points *a*, *b*, *c* and the midway point between *b* and *c* denoted by *bc* as marked in Fig. 6, obtained from the simulation using grain-based model. The dashed line marks the stage at which failure was observed in the experiments

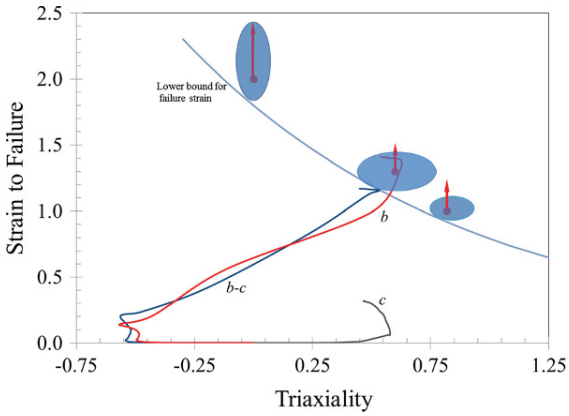

Fig. 14 The equivalent plastic strain-triaxiality variation at points *b*, *bc* and *c*, as marked in Fig. 6 superposed onto the lowerbound strain-to-failure curve obtained from grain-based measurements

point; at this stage, the plastic strain reaches about $\varepsilon_p = 1.16$; this is shown in Fig. 14 by the blue line that stops just short of penetrating the failure envelope; we note that the grain level calibration is a lower bound estimate based on statistical estimates. As the notch moves further up, the triaxiality-plastic strain state at point *b* crosses over the lower bound failure envelope and results in the crack nucleation as observed in the experiment. Therefore, even though the simulation was terminated corresponding to the arrival of the notch at the location of crack nucleation observed in the experiment, at this point, the cumulative damage parameter was at the threshold of failure, indicating that the grain-based calibration of the failure model yields an acceptable prediction of failure.

– The strains in the interior of the specimen are significantly greater than at the notch root, but they develop under a significantly lower triaxiality since these regions are under pure shear; such large strains were examined in our previous work (Ghahremaninezhad and Ravi-Chandar 2013).

The results of the numerical simulations, in comparison to the experiments indicate that the grain-based method is able to provide a realistic lower bound prediction of material deformation and failure until large strains. The primary focus here has been to elucidate a picture of material deformation and the onset of failure initiation under shear loading in ductile polycrystalline Al 6061-T6. For quantitative prediction of fracture initiation and growth, the incorporation of initial anisotropy of flow, and the further evolution of the material structure with its attendant effects on material properties is required. In particular, final fracture that appears along the reoriented microstructure needs additional considerations not included in the lower-bound failure criterion indicated in Fig. 1. While we adopted the JC method for comparison, the general outcome of this study applies to other failure models such as modified Gurson-Tvergaard-Needleman (Nahshon and Hutchinson 2008) and modified Mohr Coulomb model (Beese et al. 2010) unless these models are recalibrated with measurements on the scale of the grains; the ability of the material to deform plastically to a very large equivalent plastic strain prior to the onset of damage is an essential ingredient that must be introduced in any model.

4 Conclusions

Nucleation of a crack from a notch under dominant shear loading is considered in this article. In addition to following the load-displacement at the specimen level, optical and scanning electron microscopy are used in interrupted tests to track the deformation and failure in Al 6061-T6. The main conclusions are as follows:

– The local variation of strains can be determined to very large strain levels by proper identification and tracking of the 2nd phase particles that are distributed randomly in the material. The resolution can be improved by improving the spatial and temporal sampling during the experiments.
– The 2nd phase particles break up into smaller pieces but rotate and move with the flow of the matrix material. They do not form cavities or trigger other damage mechanisms; the matrix flows into any gaps that form as a result of differential rotation of the broken 2nd phase particles.
– The initial notch does not cause crack nucleation from its root, but triggers substantial deformation that results in the migration of the notch along the upper part of the notch; large grain deformations and rotations are observed in this vicinity. Crack nucleation occurs from the newly-formed notch, where the material is quite different from the initial material in terms of its grain size, orientation, anisotropy etc, suggesting that texture and grain boundary morphology evolution will be important aspects in dictating the onset of fracture.

– Numerical simulations using the JC model, calibrated on the basis of strains measured over length scales of specimen cross-sectional dimensions underestimate the ability of the material to strain; contradictory to experimental observations, these simulations predict premature nucleation of the crack at the root of the initial notch. It is noted that the typical calibrations of the JC model are performed at the scale of the specimen cross-section, but are always presented as material failure models, without any associated length scale.
– Numerical simulations using the grain-based calibration of Ghahremaninezhad and Ravi-Chandar (2012, 2013) are able to provide a realistic lower bound prediction of material deformation and failure until large strains. It must be emphasized that in order to use the gain-based calibration, the finite element discretization must also be at a comparable length scale—of the size of a few grains.

The lower bound failure envelope in Fig. 1 provides appropriate predictions of the plastic response of the material; for complete failure analysis, additional modeling of the changes in texture, failure properties etc is needed.

Acknowledgments This work was performed during the course of an investigation into ductile failure under two related research programs funded by the Office of Naval Research: MURI project N00014-06-1-0505-A00001 and FNC project: N00014-08-1-0189. This support is gratefully acknowledged.

References

Aoki S, Kishimoto K, Yoshida T, Sakata M, Richard HA (1990) Elastic-plastic fracture behavior of an aluminum alloy under mixed mode loading. J Phys Mech Solids 38:195–213

Bao Y, Wierzbicki T (2004) On fracture locus in the equivalent strain and stress triaxiality space. Int J Mech Sci 46:81–98

Barsoum I, Faleskog J (2007) Rupture in combined tension and shear: experiments. Int J Solids Struct 44:1768–1786

Bay BK, Smith TS, Fyhrie DP, Saad M (1999) Digital volume correlation: three-dimensional strain mapping using X-ray tomography. Exp Mech 39:217–226

Beese AM, Luo M, Li Y, Bai Y, Wierzbicki T (2010) Partially coupled anisotropic fracture model for aluminum sheets. Eng Fract Mech 77:1128–1152

Carroll JD, Abuzaid WZ, Lambros J, Sehitoglu H (2013) On the interactions between strain accumulation, microstructure and fatigue crack behavior. Int J Fract 180:223–241

Dunand M, Mohr D (2010) Hybrid experimental-numerical analysis of basic ductile fracture experiments for sheet metals. Int J Solids Struct 47:1130–1143

Ghahremaninezhad A, Ravi-Chandar K (2011) Ductile failure in polycrystalline OFHC copper. Int J Solids Struct 48:3299–3311

Ghahremaninezhad A, Ravi-Chandar K (2012) Ductile failure behavior of polycrystalline Al 6061–T6. Int J Fract 174:177–202

Ghahremaninezhad A, Ravi-Chandar K (2013) Ductile failure behavior of polycrystalline Al 6061-T6 under shear dominant loading. Int J Fract 180:23–39

Ghosal AK, Narasimhan R (1994) A finite element analysis of mixed-mode fracture initiation by ductile failure mechanisms. J Phys Mech Solids 42:953–978

Ghosal AK, Narasimhan R (1996) Numerical simulations of hole growth and ductile fracture initiation under mixed-mode loading. Int J Fract 77:281–304

Haltom SS, Kyriakides S, Ravi-Chandar K (2013) Ductile failure under combined shear and tension. Int J Solids Struct 50:1507–1522

Hancock JW, Mackenzie AC (1976) On the mechanisms of ductile failure in high-strength steels subjected to multi-axial stress-states. J Mech Phys Solids 24:147–169

Héripré E, Dexet M, Crépin J, Gélébart L, Roos A, Bornert M, Caldemaison D (2007) Coupling between experimental measurements and polycrystal finite element calculations for micromechanical study of metallic materials. Int J Plast 23:1512–1539

Hung S-C, Liechti KM (1999) Finite element analysis of the Arcan specimen for fiber reinforced composites under pure shear and biaxial loading. J Compos Mater 33:1288–1317

Johnson GR, Cook WH (1985) Fracture characteristics of three metals subject to various strains, strain rates, temperatures and pressures. Eng Fract Mech 21:31–48

Lesuer DR, Kay GJ, LeBlanc MM (2001) Modeling large strain, high-rate deformation in metals. UCRL-JC-134118, Lawrence Livermore National Laboratory

Limodin N, Rethore J, Buffiere J-Y, Hild F, Ludwig W, Rannou J, Roux S (2011) 3D X-ray microtomography volume correlation to study fatigue crack growth. Adv Eng Mater 13:186–193

Nahshon K, Hutchinson JW (2008) Modification of the Gurson model for shear. Eur J Mech A Solids 27A:1–17

Orowan E (1948) Fracture and strength of solids. Rep Prog Phys 12:185–232

Puttick KE (1960) Shear component of ductile failure. Philos Mag 5:759–762

Rice JR, Tracey DM (1969) On ductile enlargement of voids in triaxial stress fields. J Mech Phys Solids 17:201–217

Rogers HC (1960) The tensile fracture of ductile metals. Trans Metall Soc AIME 218:498–506

Sreeramulu K, Sharma P, Narasimhan R, Mishra RK (2010) Numerical simulations of crack tip fields in polycrystalline plastic solids. Eng Fract Mech 77:1253–1274

Tipper CF (1949) The fracture of metals. Metallurgia 39:133–137