Journal of
Mechanics of
Materials and Structures

JOURNAL OF MECHANICS OF MATERIALS AND STRUCTURES

http://www.jomms.org

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

PLANE HARMONIC ELASTO-THERMODIFFUSIVE WAVES IN SEMICONDUCTOR MATERIALS

JAGAN NATH SHARMA AND NAVEEN THAKUR

The aim of this article is to give a detailed account of the plane harmonic generalized elasto-thermo-diffusive (ETNP) waves in semiconductive materials. The shear (purely transverse) waves get decoupled from the rest of the motion and remain independent of the influence of other fields. These waves propagate without dispersion and attenuation in semiconductors. The coupled system of partial differential equations, governing the rest of the interacting fields, has been solved to obtain a complex secular equation. According to the frequency equation, four coupled longitudinal waves, namely, the quasither-moelastic (QTE), quasielastodiffusive (QEN/QEP), quasithermodiffusive (QTN/QTP), and quasithermal (T-mode), can exist and propagate in an infinite semiconductor. The complex secular equation of plane harmonic waves in semiconductors is solved by using Descartes' algorithm and the irreducible case of Cardan's method in order to obtain phase velocities and attenuation coefficients of all possible coupled waves. The thermoelastic (ET), elastodiffusive (EN/EP) and thermodiffusive (TN/TP) waves have also been investigated as special cases. The derived theoretical results have been illustrated and verified numerically for germanium (Ge) and silicon (Si) semiconductors. The computed phase velocity and attenuation profiles have been presented graphically.

1. Introduction

Certain substances like germanium, silicon, carbon etc. are neither good conductors like copper nor insulators like glass. In other words, the resistivity (10^{-4} to $0.5\,\Omega$m) of these materials lies between conductors and insulators. Such substances are classified as semiconductors. Semiconductors have some useful properties and are being extensively used in electronic circuits. For instance, transistor $-a$ semiconductor device is fast replacing bulky vacuum tubes in almost all applications. A semiconductor has negative temperature coefficient of resistance i.e. the resistance of a semiconductor decreases with increase in temperature and vice-versa. Wave motion is a form of disturbance, which travels through a medium due to the repeated periodic motion of particles about their mean positions. The motion being handed over from one particle to the other. The waves which can be produced or propagated in a material medium, are called elastic (mechanical) waves. In case of large frictional forces present in the medium, the wave motion dies out soon. [Maruszewski 1986a; 1986b; 1987a] presented theoretical considerations of the simultaneous interactions of elastic, thermal and diffusion of charge carriers' fields in semiconductors. The problems of interaction of various fields were formulated mathematically by [Maruszewski 1989; 1986c; 1987b] based on the following assumptions:

(i) All the considerations are made in the framework of the phenomenological model.

(ii) The electric neutrality of the semiconductor is satisfied.

Keywords: semiconductors, relaxation time, electrons and holes, waves, germanium and silicon.

(iii) The magnetic field effect can be ignored.

(iv) The mass of the charge carrier fields is negligible.

 (v) The surface heat sources are neglected.

(vi) The electric field with in the boundary layer is very weak and can be neglected.

(vii) The recombination functions of electrons and holes are selected on the basis of facts that take care of defects and hence the concentration values of the charge carrier fields according to [Many et al. 1965].

(viii) The surface of the semiconductor is free of mechanical loading.

(ix) The temperature $T_0 =$ constant is the uniform reference temperature and $\theta = T - T_0$, is the temperature change of the body.

 (x) The concentrations of electrons and holes satisfy the conditions $N = n - n_0$, $P = p - p_0$, where n, p and n_0, p_0 are respectively the nonequilibrium and equilibrium values of electrons and holes concentrations.

Maruszewski [1989] studied the propagation of thermodiffusive surface waves in the semiconductors with relaxation of heat and charge carrier fields. The secular equation of coupled elastic, thermal and diffusive waves is obtained and illustrated by considering two particular cases of elastodiffusive (EN) and thermodiffusive (TN) waves.

The governing equations in classical dynamic coupled thermoelasticity are wave-type (hyperbolic) equations of motion and a diffusion type (parabolic) equation of heat conduction. Therefore, it is seen that a part of the solution of energy equation extends to infinity implying that if a homogeneous isotropic elastic medium is subjected to thermal or mechanical disturbance, the effects of temperature and displacement fields are experienced at an infinite distance from the source of disturbance. This shows that part of the disturbance has an infinite velocity of propagation, a physically impossible phenomenon. With this drawback in mind, some researches such as [Lord and Shulman 1967; Green and Lindsay 1972; Dhaliwal and Sherief 1980] and [Chandrasekharaiah 1986], modified the Fourier law of heat conduction and constitutive relations to obtain a hyperbolic equation for heat conduction. These models include the time needed for acceleration of heat flow in addition to the coupling between temperature and strain fields. According to the investigations of [Ackerman et al. 1966; Guyer and Krumhansl 1966; Ackerman and Overton 1969], these theories have also been supported with the experimental exhibition of actual occurrence of 'second sound' at low temperature and small intervals of time. [Banerjee and Pao 1974] investigated the propagation of plane harmonic waves in infinitely extended anisotropic thermoelastic solids by taking into account the thermal relaxation time. Extensive studies of wave propagation in heat conducting elastic solids under the influence of thermal relaxation time in "infinite velocity" and "finite velocity" descriptions, have been carried out by many investigators such as [Scott 1989; Chadwick 1979; Sharma et al. 2000; Sharma 1986] and [Sharma and Singh 1989; 1990].

The present article is aimed at giving a detailed account of the plane harmonic generalized thermoelastic waves in infinite semiconductor materials in context of the mathematical model formulated by [Maruszewski 1989]. The basic equations are first nondimensionalized and then solved by adopting the approach of [Achenbach 1973] after decoupling the shear waves' (purely transverse) motion. The shear waves are not affected by thermal and charge carrier fields and also remain independent from the

rest of the motion. The frequency equation for rest of the motion reveals that, in general, there are four longitudinal waves namely, a quasithermoelastic (QTE), a quasithermal (T-mode) and two quasidiffusive (elastodiffusive QEN/QEP and thermodiffusive QTN/QTP waves), which can propagate in such semiconductor media. These waves are coupled and get modified due to thermal variations, thermal relaxation time, and life/relaxation times of electron and hole fields. The complex secular equation of coupled waves is then solved by using Descartes' algorithm along with irreducible Cardan's method. The analytical results so obtained have been verified numerically and illustrated graphically in case of Ge and Si materials.

2. Formulation of the problem

We consider an unbounded, homogeneous, isotropic, thermoelastic semiconductor at a uniform temperature T_0 in an undisturbed state. Let $\vec{u}(x_1, x_2, x_3, t) = (u_1, u_2, u_3)$ and $\theta(x_1, x_2, x_3, t)$ be the displacement vector and temperature change of the medium at any time t, respectively. The basic governing equations of motion and heat conduction, in the absence of body forces and heat sources, for such materials as given by [Maruszewski 1989] are

$$\mu u_{i,jj} + (\lambda + \mu) u_{j,ij} - \rho \ddot{u}_i - \lambda^n N_{,i} - \lambda^p P_{,i} - \lambda^T \theta_{,i} = 0, \tag{1}$$

$$K\theta_{,ii} + m^{nq} N_{,ii} + m^{pq} P_{,ii}$$
$$- \left(1 + t^Q \frac{\partial}{\partial t}\right)(\rho C_e \dot{\theta} + \rho T_0 \alpha^n \dot{N} + \rho T_0 \alpha^p \dot{P} + T_0 \lambda^T \dot{u}_{k,k}) - \rho(a_1^n \dot{N} + a_1^p \dot{P})$$
$$= \left(a_1^n \left(\frac{\rho}{t_n^+}\right) N + a_1^p \left(\frac{\rho}{t_p^+}\right) P\right), \tag{2}$$

$$\rho D^n N_{,ii} + m^{qn} \theta_{,ii} - \rho \left(1 - a_2^n T_0 \alpha^n + t^n \frac{\partial}{\partial t}\right) \dot{N} - a_2^n (\rho C_e \dot{\theta} + \rho T_0 \alpha^p \dot{P} + T_0 \lambda^T \dot{u}_{k,k})$$
$$= -\left(1 + t^n \frac{\partial}{\partial t}\right)\left(\frac{\rho}{t_n^+}\right) N, \tag{3}$$

$$\rho D^p P_{,ii} + m^{qp} \theta_{,ii} - \rho \left(1 - a_2^p T_0 \alpha^p + t^p \frac{\partial}{\partial t}\right) \dot{P} - a_2^p (\rho C_e \dot{\theta} + \rho T_0 \alpha^n \dot{N} + T_0 \lambda^T \dot{u}_{k,k})$$
$$= -\left(1 + t^p \frac{\partial}{\partial t}\right)\left(\frac{\rho}{t_p^+}\right) P, \tag{4}$$

where the notation

$$a_1^n = \frac{a^{Qn}}{a^Q}, \qquad a_1^p = \frac{a^{Qp}}{a^Q}, \qquad a_2^n = \frac{a^{Qn}}{a^n}, \qquad a_2^p = \frac{a^{Qp}}{a^p},$$

$$P = p - p_0, \qquad N = n - n_0, \qquad \lambda^T = (3\lambda + 2\mu)\alpha_T,$$

is used. The field variables have been subjected to only those assumptions (except (v), (vi) and (viii)) of [Maruszewski 1989] that are applicable and relevant in the present context of an infinite description

of semiconductors. Here λ, μ are Lame parameters; ρ is the density of the semiconductor; λ^n, λ^p are the elastodiffusive constants of electrons and holes; α_T is the coefficient of linear thermal expansion of the material; K is the thermal conductivity; α^p, α^n are thermodiffusive constants of holes and electrons; $a^{Qn}, a^{Qp}, a^Q, a^n, a^p$ are the flux-like constants; and D^n, D^p are the diffusion coefficients of electron and holes. The quantities $m^{nq}, m^{pq}, m^{qn}, m^{qp}$ are the Peltier–Seebeck–Dufour–Soret-like constants; t^Q, t^n and t^p are the relaxation times of heat, electron and hole fields, respectively; C_e is the specific heat; t_n^+, t_p^+ denotes the life times of the carriers' fields; and n, p and n_0, p_0 are the nonequilibrium and equilibrium values of electrons and holes, respectively. The comma notation is used for spatial derivatives and a superposed dot represents differentiation with respect to time.

We define the quantities

$$x_i' = \frac{\omega^* x_i}{c_1}, \qquad t'l = \omega^* t, \qquad \theta' = \frac{\theta}{T_0}, \qquad P' = \frac{P}{p_0},$$

$$N' = \frac{N}{n_0}, \qquad u_i' = \frac{\rho \omega^* c_1}{\lambda^T T_0} u_i, \qquad t^{Q'} = t^Q \omega^*, \qquad t^{p'} = t^p \omega^*,$$

$$t^{n'} = t^n \omega^*, \qquad t_n^{+'} = t_n^+ \omega^*, \qquad t_p^{+'} = t_p^+ \omega^*, \qquad \delta^2 = \frac{c_2^2}{c_1^2},$$

$$\varepsilon_T = \frac{\lambda^{T^2} T_0}{\rho C_e (\lambda + 2\mu)}, \qquad \omega^* = \frac{C_e(\lambda + 2\mu)}{K}, \qquad c_1^2 = \frac{\lambda + 2\mu}{\rho},$$

$$c_2^2 = \frac{\mu}{\rho}, \qquad k = \frac{K}{\rho C_e}, \qquad \bar{\lambda}_n = \frac{\lambda^n n_0}{\lambda^T T_0}, \qquad \bar{\lambda}_p = \frac{\lambda^P p_0}{\lambda^T T_0},$$

$$\varepsilon^{qn} = \frac{m^{qn} T_0}{\rho D^n n_0}, \qquad \varepsilon^{qp} = \frac{m^{qp} T_0}{\rho D^p p_0}, \qquad \varepsilon_n = \frac{a_2^n K T_0}{\rho n_0 D^n}, \qquad \varepsilon_p = \frac{a_2^p K T_0}{\rho p_0 D^p},$$

$$\varepsilon^{pq} = \frac{m^{pq} p_0}{K T_0}, \qquad \varepsilon^{nq} = \frac{m^{nq} n_0}{K T_0}, \qquad a_0^n = \frac{a_1^n n_0}{C_e T_0}, \qquad a_0^p = \frac{a_1^p p_0}{C_e T_0},$$

$$\alpha_0^n = \frac{\alpha^n n_0}{C_e}, \qquad \alpha_0^p = \frac{\alpha^P p_0}{C_e}. \tag{5}$$

Here ε_T and k are the thermoelastic coupling parameter and thermal diffusivity. Upon introducing the scalar point potential function ϕ and vector point potential function $\vec{\psi}$ defined by the relations

$$\vec{u} = \nabla \phi + \nabla \times \vec{\psi}, \qquad \nabla \cdot \vec{\psi} = 0 \tag{6}$$

into Equations (1)–(4), along with the quantities in (5), we obtain

$$\nabla^2 \phi - \ddot{\phi} - \overline{\lambda_n} N - \overline{\lambda_p} P - \theta = 0, \tag{7a}$$

$$-\varepsilon_T\nabla^2(\dot{\phi}+t^Q\ddot{\phi})+\varepsilon^{nq}\nabla^2N-\left\{\frac{\alpha_0^n t^Q\partial 2}{\partial t^2}+(a_0^n+\alpha_0^n)\frac{\partial}{\partial t}+\frac{a_0^n}{t_n^+}\right\}N$$

$$+\varepsilon^{pq}\nabla^2 P-\left\{\frac{\alpha_0^p t^Q\partial 2}{\partial t^2}+(a_0^p+\alpha_0^p)\frac{\partial}{\partial t}+\frac{a_0^p}{t_p^+}\right\}P+\nabla^2\theta-(\dot{\theta}+t^Q\ddot{\theta})=0, \tag{7b}$$

$$-\varepsilon_n\varepsilon_T\nabla^2\dot{\phi}+\nabla^2N-\frac{K}{\rho C_e D^n}\left(-\frac{1}{t_n^+}+\left(1-\frac{\in_n\alpha_0^n D^n}{k}-\frac{t^n}{t_n^+}\right)\frac{\partial}{\partial t}+\frac{t^n\partial 2}{\partial t^2}\right)N$$

$$-\varepsilon_n\alpha_0^p\dot{P}-\varepsilon_n\dot{\theta}+\varepsilon^{qn}\nabla^2\theta=0, \tag{7c}$$

$$-\varepsilon_p\varepsilon_T\nabla^2\dot{\phi}+\nabla^2P-\frac{K}{\rho C_e D^p}\left(-\frac{1}{t_p^+}+\left(1-\frac{\in_p\alpha_0^p D^p}{k}-\frac{t^p}{t_p^+}\right)\frac{\partial}{\partial t}+\frac{t^p\partial 2}{\partial t^2}\right)P$$

$$-\varepsilon_p\alpha_0^n\dot{N}-\varepsilon_p\dot{\theta}+\varepsilon^{qp}\nabla^2\theta=0, \tag{7d}$$

$$\nabla^2\vec{\psi}=\frac{1}{\delta^2}\ddot{\vec{\psi}}. \tag{7e}$$

The last equation of (7e) corresponds to purely transverse waves which get decoupled from rest of the motion and are not affected by the thermal and charge carrier fields. These waves travel with nondimensional velocity $'\delta'$ without dispersion, attenuation, or damping. We drop this motion in the following analysis unless stated otherwise. Equations (7c) and (7d) can be further simplified under the assumption that the considered semiconductor is of relaxation type. For such materials, according to [Maruszewski 1989], the diffusion approximation of the physical process ceases to be obligatory and the relaxation/life times t^n, $t_n^+(t^p, tp^+)$ become comparable to each other in their values $(t^n=tn^+, t^p=tp^+)$.

3. Solution of the problem

We may take plane harmonic wave solutions as

$$(\phi,\theta,N,P)=(\overline{\phi,\theta,N},\overline{P})\exp\{i\omega(v^{-1}x_r n_r-t)\}, \quad r=1,2,3. \tag{8}$$

The use of solution (8) in the coupled system of equations (7a)–(7d), after straightforward algebraic reductions and manipulations, leads to the following characteristic equation

$$\xi^4-A\xi^3+B\xi^2-C\xi+D=0, \qquad \xi=v^{-2}, \tag{9}$$

where

$$A = \frac{1 + \left(\Delta_1 + \Delta_2 + (1 + \varepsilon_T)\Delta_3\right)}{\Delta},$$

$$B = \frac{\left(\Delta_1' + (1 + \varepsilon_T)(\Delta_2' + \Delta_3') + \Delta_4' + \Delta_5' + \Delta_6'\right)}{\Delta},$$

$$C = \frac{\left((1 + \varepsilon_T)\Delta_1'' + \Delta_2'' + \Delta_3'' + \Delta_4''\right)}{\Delta},$$

$$D = \frac{\Delta_1'''}{\Delta}.$$

(10)

Here the quantities $\Delta_i (i = 1, 2, 3)$, $\Delta_i' (i = 1, 2, 3, 4, 5, 6)$, $\Delta_i'' (i = 1, 2, 3, 4)$, Δ_1''' and Δ are defined in the Appendix. Equation (9), being a fourth degree polynomial equation in ξ, has four roots, and hence in general, there are four ETNP waves: a QTE, a T-mode, and two quasidiffusive waves (QEN/QEP and QTN/QTP), in addition to purely transverse waves which can propagate in such semiconductive materials. The secular equation (9) with complex coefficients A, B, C, and D contains complete information regarding the wave number, frequency, phase velocity and attenuation coefficient of these waves. In order to solve the complex secular equation (9) we use Descartes' algorithm outlined below:

Shifting the roots of secular equation (9) by a factor of $\frac{A}{4}$ to eliminate the second term, we obtain

$$\zeta^4 + H\zeta^2 + G\zeta + I = 0,$$

(11)

where

$$\zeta = \xi - \frac{A}{4}, \qquad H = B - \frac{3A^2}{8}, \qquad G = \frac{AB}{2} - \frac{A^3}{8} - C, \qquad I = D + \frac{A^2 B}{16} - \frac{3A^4}{256} - \frac{AC}{4}.$$

Factoring Equation (11) into two quadratic factors, we have

$$\zeta^4 + H\zeta^2 + G\zeta + I = (\zeta^2 + l\zeta + m)(\zeta^2 - l\zeta + n).$$

(12)

Comparing the coefficients of various powers of ζ in (12) on both sides, we get

$$m + n = l^2 + H, \qquad n - m = \frac{G}{l}, \qquad mn = I.$$

(13)

Eliminating m and n from Equation (13), we obtain

$$Z^3 + 2HZ^2 + (H^2 - 4I)Z - G^2 = 0,$$

(14)

where $Z = l^2$. Being cubic with complex coefficients, Equation (14) can be solved by using the irreducible case of Cardan's method with the help of De Moivre's theorem. We again shift the roots of (14) by a factor of $\frac{-2H}{3}$ in order to obtain the standard cubic as

$$Y^3 - 3H^* Y - G^* = 0,$$

(15)

where

$$Y = Z + \frac{2H}{3}, \qquad H^* = \frac{(H^2 + 12I)}{9}, \qquad G^* = G^2 - \frac{8HI}{3} + \frac{2H^3}{27}.$$

(16)

Let the roots of Equation (15) be of the type

$$Y = U + V \tag{17}$$

so that $U^3 + V^3 = G^*$, $U^3 V^3 = H^{*3}$.

We may find the cube roots with the help of De Moivre's theorem, as shown below:

Let

$$U^3 = \frac{G^* + \sqrt{G^* - 4H^{*3}}}{2} = L + iM, \quad L, M \in R. \tag{18}$$

Then the values of U are given by

$$U_k = r^{1/3} \left(\cos \frac{2k\pi + \Phi}{3} + i \sin \frac{2k\pi + \Phi}{3} \right), \quad k = 0, 1, 2, \tag{19}$$

where $r = \sqrt{L^2 + M^2}$ and $\Phi = \tan^{-1}(M/L)$. Having determined U, the values of V can be obtained from the relation $UV = H^*$ which further leads to the required values of Y and hence to the values of $l^2 = Z = Y - 2H/3$. One of the (convenient) values of l so obtained is then used to evaluate m and n by Equation (13). Using the values of m, n and l, the reduced secular Equation (11) is factored into two quadratic factors of the type (12), which are further solved to obtain the four roots ζ_i, $i = 1, 2, 3, 4$. The complex roots of secular equation (9) are obtained from the relation $\xi_i = \zeta_i + A/4$, for $i = 1, 2, 3, 4$. This leads to the determination of the complex phase velocities as

$$v_i = \frac{1}{\sqrt{\xi_i}}, \quad \text{for } i = 1, 2, 3, 4. \tag{20}$$

In general, v is complex, and hence we may write

$$v^{-1} = V^{-1} + i\omega^{-1}Q \tag{21}$$

so that the exponent part $\exp[i\omega(v^{-1}x_p n_p - t)]$ of solutions (8) can be rewritten as

$$\exp\{i\omega(V^{-1}x_p n_p - t) - Qx_p n_p\}.$$

This implies that V is the phase speed and Q the attenuation coefficient of ETNP waves.

Upon using representation (21) in Equation (20) we can obtain the phase velocity (V_i) and the attenuation coefficient (Q_i) of different modes of wave propagation. We obtain

$$V_i = \frac{1}{\text{Re}\sqrt{\xi_i}}, \quad Q_i = \omega I_m(\sqrt{\xi_i}), i = 1, 2, 3, 4. \tag{22}$$

4. Special cases of wave solutions

In this section we consider some special cases of wave propagation such as EN/EP, TN/TP and ET waves, in semiconductor materials.

4.1. *Elastodiffusive (EN/EP) waves.* If we confine our discussion concerning the reciprocal dynamical interactions of elastic and electron diffusion fields to the propagation of EN waves, the system of equations (7) in the absence of thermal and hole charge carrier fields ($P = \theta = 0$, $\varepsilon_T = 0 = \varepsilon^{nq}$, $\alpha_0{}^n = 0 = a_0{}^n$) along with the solution to (8) leads to the secular equation

$$(\xi - p_1^2)(\xi - p_2^2) = 0, \tag{23}$$

where $p_1^2 + p_2^2 = 1 + \tau_n^{*'}$, $p_1^2 p_2^2 = \tau_n^{*'}$.

Here $\tau_n^{*'}$ is obtained from τ_n^* defined in the Appendix, on setting $\alpha_0^n = 0$. In general, the roots $\xi = p_1^2$, p_2^2 are complex and hence waves are attenuated in space. The corresponding phase velocity depends directly on the relaxation and life times of the electrons. For relaxation type semiconductors ($t^n = t_n^+$), the quantity $\tau_n^{*'}$ becomes real and so are the roots $\xi = p_1^2$, p_2^2. Therefore for such semiconductors EN waves propagate without dispersion, damping, and attenuation, which is in agreement with [Maruszewski 1989]. The amplitude ratios of the waves in this case are related by

$$\frac{\bar{N}}{\bar{\bar{\phi}}} = k^2(v^2 - 1)/\bar{\lambda}_n. \tag{24}$$

Upon using representation (21), the phase velocities and attenuation coefficients of EN waves are obtained as

$$V_i = \frac{1}{\text{Re}(p_i)}, \qquad Q_i = \omega I_m(p_i), \quad i = 1, 2. \tag{25}$$

The EP-waves can also be discussed in a similar manner by omitting the influence of thermal and electron fields ($N = \theta = 0$, $\varepsilon_T = 0 = \varepsilon^{pq}$, $\alpha_0^p = 0 = a_0^p$). The corresponding results can be obtained by replacing (N, n) with (P, p) in the above analysis.

4.2. *TN/TP waves.* Here we confine our discussion concerning the reciprocal dynamical interactions of the thermal and electron diffusion fields to the propagation of TN waves, and omit the elastic and hole charge carrier fields ($\phi = 0 = P$, $\alpha_0^p = 0 = \varepsilon_p$, $\varepsilon^{qp} = 0 = \varepsilon_T$). The system of equations (7) with the help of solution (8) in this case leads to the secular equation

$$(\xi - q_1^2)(\xi - q_2^2) = 0, \tag{26}$$

where

$$q_1^2 + q_2^2 = \frac{\tau^Q + \tau_n^* - \varepsilon^{qn}\tau_n' - i\omega^{-1}\varepsilon_n\varepsilon^{nq}}{1 - \varepsilon^{nq}\varepsilon^{qn}}, \qquad q_1^2 q_2^2 = \frac{\tau^Q\tau_n^* - i\omega^{-1}\varepsilon_n\tau_n'}{1 - \varepsilon^{nq}\varepsilon^{qn}}. \tag{27}$$

The complex phase velocity can be obtained from Equation (26) as

$$v_i^{-1} = q_i, \quad i = 1, 2. \tag{28}$$

In this case the waves are attenuated in space and damped with time even for relaxation type of semiconductors. Upon using representation (21) the real phase velocities and attenuation coefficients of TN waves are obtained as

$$V_i = \frac{1}{\text{Re}(q_i)}, \qquad Q_i = \omega \, \text{Im}(q_i), \quad i = 1, 2. \tag{29}$$

Here the amplitude ratios are given by

$$\frac{\bar{\theta}}{\bar{N}} = \frac{\tau_n' v^2 - \varepsilon^{nq}}{1 - \tau^Q v^2} = \frac{\tau_n^* v^2 - 1}{\varepsilon^{qn} - i\omega^{-1}\varepsilon_n v^2}. \tag{30}$$

The results pertaining to TP waves can be obtained from the above analysis by replacing N, n with P, p after setting ($\phi = 0 = N$, $\alpha_0^n = 0 = \varepsilon_n$, $\varepsilon^{qn} = 0 = \varepsilon_T$) in the governing equations (7).

4.3. _ET waves._ When a complete equilibrium state of electron and hole concentration is established the system becomes charge-free. Here, we confine our discussion concerning the reciprocal dynamical interactions of elastic and thermal fields in the absence of electron and hole fields ($N = 0 = P$, $\varepsilon_n = \varepsilon_p = 0$, $\varepsilon^{qn} = 0 = \varepsilon^{qp}$) to the propagation of ET waves.

In this case the system of equations (7) governing the interaction along with solution (8) leads to the secular equation for ET waves as under

$$(\xi - a_1^2)(\xi - a_2^2) = 0, \tag{31}$$

where

$$a_1^2 + a_2^2 = 1 + (1 + \varepsilon_T)\tau^Q, \qquad a_1^2 a_2^2 = \tau^Q \tag{32}$$

and τ^Q is defined in the Appendix. The secular equation (31) gives us $v_i^{-1} = \pm a_i$, $i = 1, 2$.

These quantities are also complex so the waves are attenuated in space and damped with time. Upon using representation (21) the real phase velocities and attenuation coefficients of ET waves are obtained as

$$V_i = \frac{1}{\text{Re}(a_i)}, \qquad Q_i = \omega \, \text{Im}(a_i), \quad i = 1, 2. \tag{33}$$

In this case the amplitude ratios are given by

$$\frac{\bar{\theta}}{\bar{\phi}} = k^2(v^2 - 1). \tag{34}$$

This type of wave motion has already been discussed by many authors such as [Chandrasekharaiah 1986; Chadwick 1979; Chadwick and Seet 1970; Sharma et al. 2000; Sharma 1986; Sharma and Singh 1989; 1990; Scott 1989].

5. Numerical results and discussion

In this section the values of phase velocity and attenuation coefficient of various partial wave modes have been computed numerically from the analytical results obtained above for Ge and Si materials under the assumption that the semiconductor considered is of relaxation type. In such a case, the diffusion approximation of the physical processes ceases to be obligatory, and t_n, t_n^+, t_p, t_p^+ become comparable to each other in their values so that $t_n = t_n^+$, $t_p = t_p^+$. The physical data for these materials is given in Table 1. To understand the interactions of various fields considered in thermoelastic semiconductors, the nondimensional phase velocities and attenuation coefficients of different modes of wave propagation have been obtained and computed numerically for Ge and Si materials, and their profiles are plotted on log-linear scale against nondimensional frequency (ω) in Figures 1, 2–7, 8. The phase velocity and attenuation coefficient profiles in special cases of ET, EN/EP and TN/TP waves have also been computed

Coefficient	Unit	Value (Ge)	Value (Si)	Reference
λ	Nm^{-2}	0.48×10^{11}	0.64×10^{11}	[Maruszewski 1989]
μ	Nm^{-2}	0.53×10^{11}	0.65×10^{11}	[Maruszewski 1989]
ρ	Kgm^{-3}	5.3×10^3	2.3×10^3	[Maruszewski 1989]
t_n^+	s	$< 10^{-5}$	$< 1.4 \times 10^{-6}$	[Maruszewski 1989]
t_p^+	s	$< 10^{-5}$	$< 10^{-5}$	[Maruszewski 1989]
D^n	$\mathrm{m^2 s^{-1}}$	10^{-2}	0.35×10^{-2}	[Maruszewski 1989]
D^p	$\mathrm{m^2 s^{-1}}$	0.5×10^{-2}	0.125×10^{-2}	[Maruszewski 1989]
K	$\omega \mathrm{m^{-1} K^{-1}}$	60	150	[Sze 1981]
C_e	$\mathrm{jKg^{-1} K^{-1}}$	310	700	[Sze 1981]
α_T	$\mathrm{K^{-1}}$	5.8×10^{-6}	2.6×10^{-6}	[Sze 1981]
$n_0 = p_0$	$\mathrm{m^{-3}}$	10^{20}	10^{20}	[Zambuto 1989]
α^p	$\mathrm{m^2 / s}$	1.3×10^{-3}	5×10^{-3}	[Zambuto 1989]
α^n	$\mathrm{m^2 / s}$	3.4×10^{-3}	1×10^{-2}	[Zambuto 1989]
m^{nq}	$\mathrm{vk^{-1}}$	1.4×10^{-5}	1.4×10^{-5}	[Lal 1995]
m^{pq}	$\mathrm{vk^{-1}}$	-0.004×10^{-6}	-0.004×10^{-6}	[Lal 1995]
m^{qn}	$\mathrm{vk^{-1}}$	1.4×10^{-5}	1.4×10^{-5}	[Lal 1995]
m^{qp}	$\mathrm{vk^{-1}}$	-0.004×10^{-6}	-0.004×10^{-6}	[Lal 1995]

Table 1. Physical data of germanium and silicon.

and represented graphically on log-linear scales in Figures 9–15. The numerical results are found to be in close agreement with the theoretical analysis.

Figure 1 shows the phase velocity (V_1) profiles of QTE waves with frequency in Ge and Si semiconductor materials. The variations of phase velocity at low frequency ($\omega \ll 1$) limits are quite small as compared to that at high-frequency ($\omega \gg 1$). The phase velocity in both the cases increases sharply in the frequency range $0.3 \le \omega \le 100$ to attain its maximum value at $\omega = 100$ for Ge and at $\omega = 30$ in case of Si, and then becomes steady after a slight decline in its value for $\omega \ge 100$. Although the effect of relaxation time of heat transportation is negligibly small, it is still more significant at high frequency (isothermal) conditions than at low frequency (isentropic) limits, especially when $\omega \ge 10$. This also shows that thermal relaxation (second sound) effects are short lived. The behavior of dispersion curves for Ge and Si materials is similar except that the magnitude of velocity in the latter one is quite small. Figure 2 shows nondimensional attenuation-frequency curves of QTE waves. As evident from Figures 1 and 2, the behavior of nondimensional attenuation coefficient (Q1) is similar to that of nondimensional velocity (V_1) for Ge. It assumes maximum value at $\omega = 100$ in higher frequency regime and is significantly affected by thermal relaxation time of heat transportation. The attenuation profile of silicon (Si) is linear everywhere except in the range $1 \le \omega \le 100$, where it has Gaussian character with mean value at $\omega = 10$.
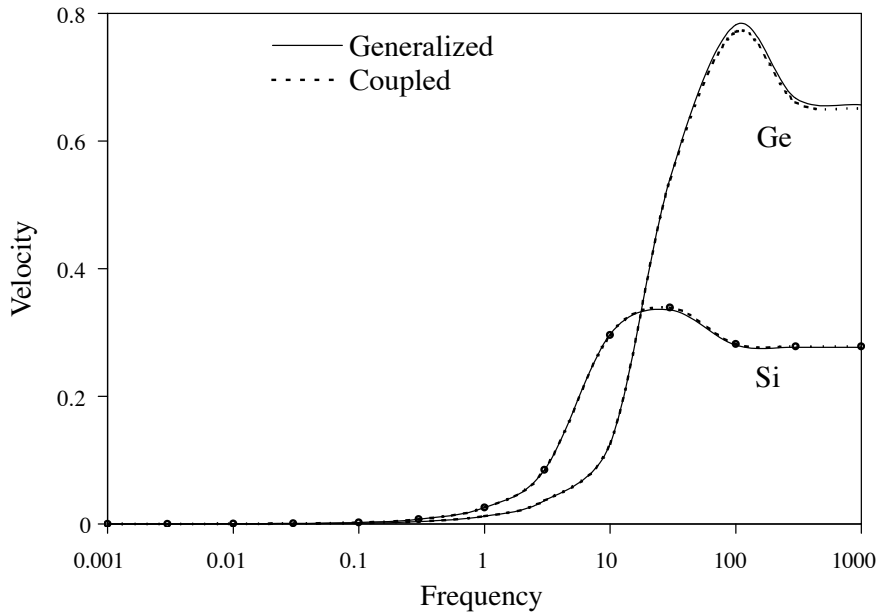
Figure 1. Phase velocity profile of QTE waves.
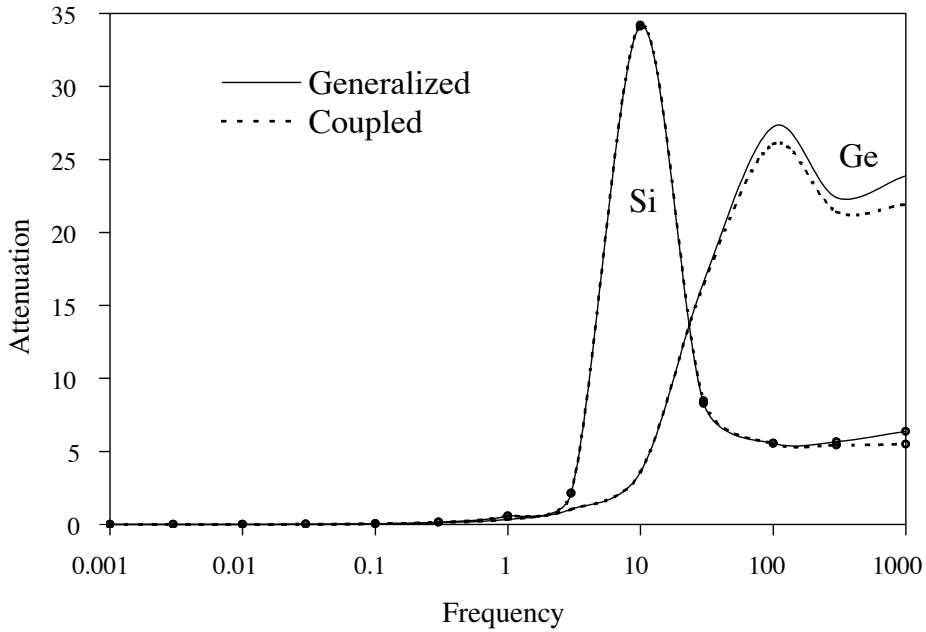
Figure 2. Attenuation coefficient profile of QTE waves.

Figure 3 shows nondimensional phase velocity (V_2) profiles of QEN/QEP waves with nondimensional frequency (ω) in Ge and Si semiconductor materials. For the $0 \leq \omega \leq 10$ frequency range, the variations
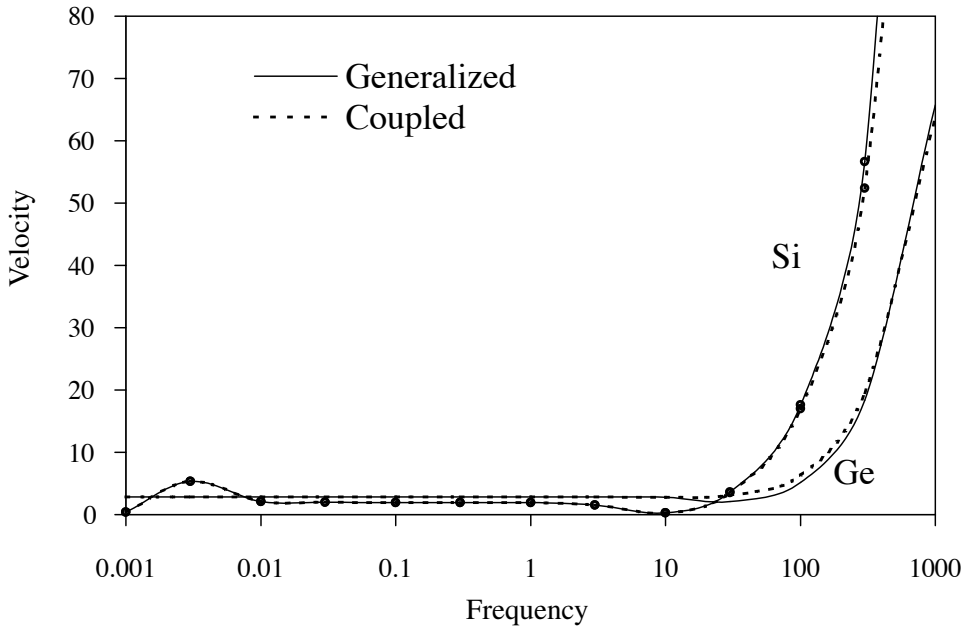
Figure 3. Phase velocity profile of QEN/QEP waves.
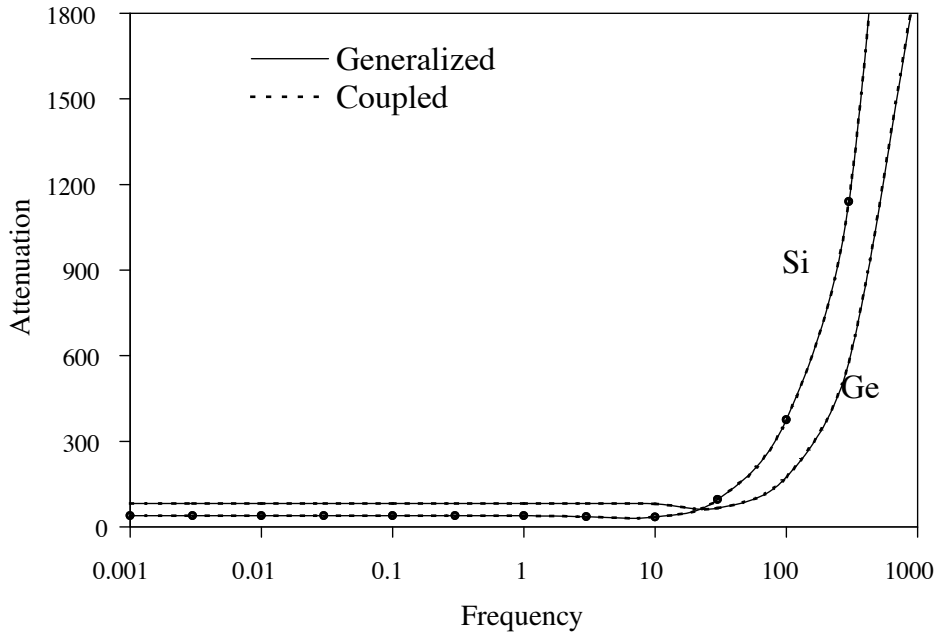
Figure 4. Attenuation coefficient profile of QEN/QEP waves.

of phase velocity are almost linear in Ge material and increase logarithmically at higher values of the frequency afterwards. The phase velocity profile for silicon (Si) semiconductor is slightly dispersive for $0 \leq \omega \leq 10$ in contrast to that of Ge which is linear and hence nondispersive in this range of frequency
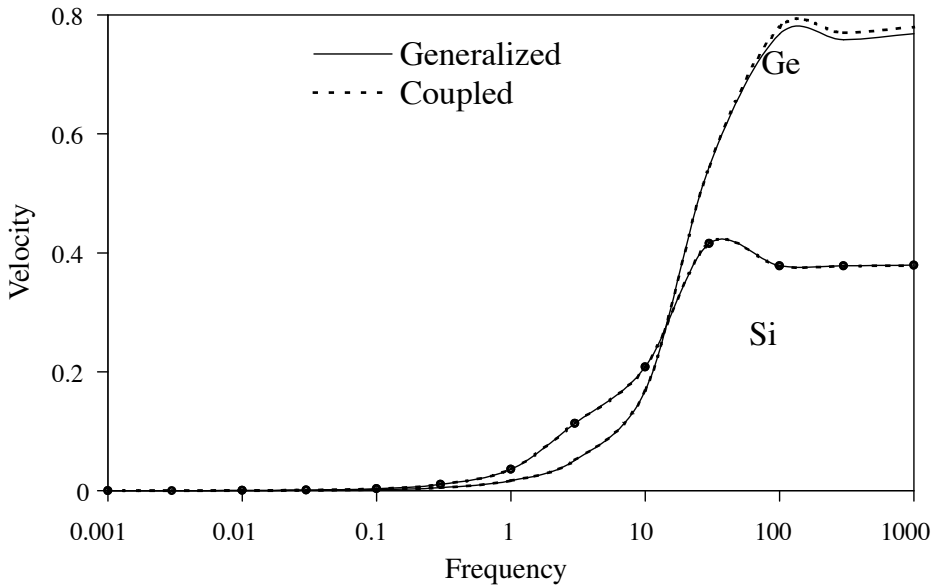
Figure 5. Phase velocity profile of QTN/QTP waves.
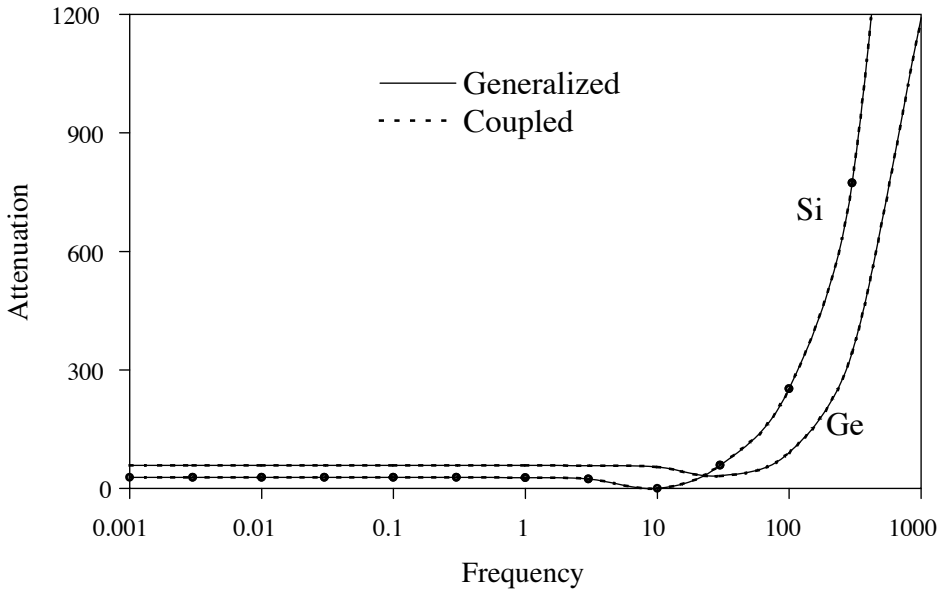
Figure 6. Attenuation coefficient profile of QTN/QTP waves.

values. The effect of thermal relaxation on the transportation of charge carrier fields is observed to be prominent for higher values ($\omega \geq 10$) of the frequency. From Figure 4, we see that the behavior of attenuation coefficient (Q_2) is same as that of the phase velocity in Figure 3, except that the thermal relaxation has negligibly small effect in this case. The variations of the attenuation coefficient are

linear for the frequency range $0 \leq \omega \leq 10$ and increase logarithmically afterwards in both Ge and Si semiconductors. Here the attenuation profiles for thermal relaxation and nonrelaxation cases overlap each other, meaning that thermal relaxation phenomenon does not affect the attenuation coefficient. The critical value of frequency pertaining to the change of this behavior of phase velocity and attenuation coefficient in both the semiconductors (Ge and Si) is observed to be at $\omega = 30$ rather than at $\omega = 1.0$.

Figure 5 shows the phase velocity (V_3) profile of QTN/QTP waves in Ge and Si semiconductors. It is observed that the phase velocity is almost negligible in the frequency range $0 \leq \omega \leq 0.3$, and suffers a sharp increase in its value in $0.3 < \omega \leq 10$. This slightly decreases after attaining its maximum value at $\omega = 100$ for Ge and at $\omega = 30$ in the case of Si to become steady afterwards. The effect of thermal relaxation time is quite pertinent to phase velocity at higher frequencies in the case of Ge, but it has virtually no effect on Si semiconductors. The magnitude of phase velocity in the germanium semiconductor is much higher than that of the silicon one. Figure 6 represents attenuation coefficient profiles of QTN/QTP waves for Ge and Si materials. The behavior of the attenuation coefficient (Q_3) in this case is more or less similar to that of Q_2 as represented by Figure 4 except for the variations in magnitude.

Figure 7 shows the phase velocity (V_4) profiles of T-mode waves in Ge and Si semiconductor materials with respect to nondimensional frequency. The phase velocity in silicon (Si) material has Gaussian behavior with mean value at $\omega = 10$ in the frequency range $0.3 \leq \omega \leq 30$. The phase velocity profile of Ge has linear variations in $0 \leq \omega \leq 3$ but is subject to dispersion beyond $\omega \geq 3$. The effect of thermal relaxation on phase velocity is clearly visible in the case of Si, but is quite small for the Ge semiconductor. Figure 8 represents the attenuation coefficient profile of T-mode waves for Ge and Si materials. The behavior of attenuation coefficient (Q_4) profiles in this case is more or less similar to that of Q_3 represented by the profiles in Figure 6 except for certain variations in magnitude and its prominence in this case.

Figure 9 shows phase velocity and attenuation coefficient profiles for EN/EP waves with respect to frequency. The nondimensional phase velocity in Ge and Si increases sharply in the frequency range $0 \leq \omega \leq 1$ and becomes linear for $\omega \geq 1$ in both materials. This means that the elasto-diffusive waves are significantly influenced by electron and hole charge carrier fields in the low frequency limit, but remain unaffected by such fields at higher frequencies in both materials. The variations of attenuation coefficient are noticed to be significant, but quite small, in the frequency range $0 \leq \omega \leq 0.1$ for Ge and Si materials. The magnitude of the attenuation coefficient in the silicon (Si) semiconductor is greater than that of the Ge, but it varies in Gaussian manner with mean value at $\omega = 0.01$ in both materials. The attenuation coefficient profiles disappear for $\omega \geq 0.1$ in the materials seen in Figure 9.

Figure 10 represents the nondimensional phase velocity and attenuation coefficient profiles of electron/hole diffusive (N/P) waves with frequency. We see that the phase velocity varies linearly with the frequency in the range $0 \leq \omega \leq 0.1$, and then increases logarithmically for $0.1 \leq \omega \leq 10$ before becoming steady/fixed at $\omega \geq 10$. The amplitude of phase velocity in Ge is larger than that of the case of Si material. The behavior of the attenuation coefficient of electron/hole diffusive (N/P) waves is similar to that of phase velocity for electron/hole diffusive waves except for the fact that the magnitude of the former has larger values in the case of Si as compared to those of the Ge semiconductor. Thus, in the Si semiconductor these waves are subjected to more attenuation than in the case of Ge.
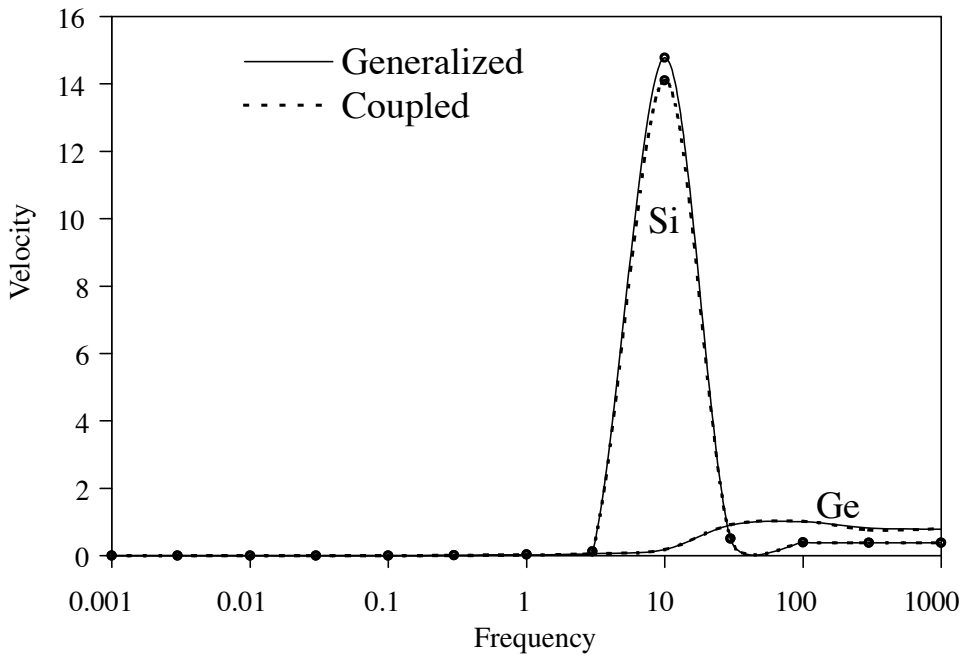
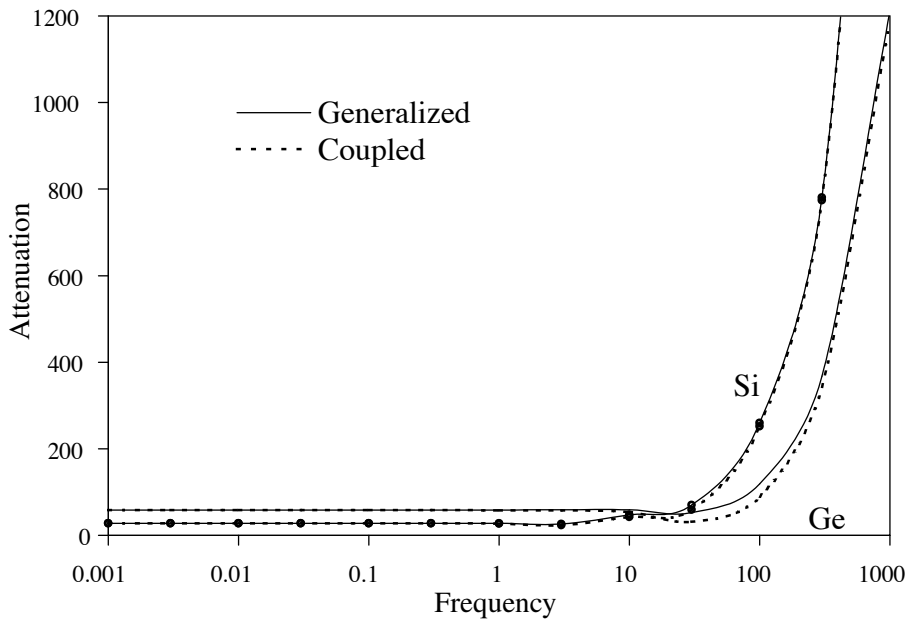**Figure 7.** Phase velocity profile of T-mode waves.

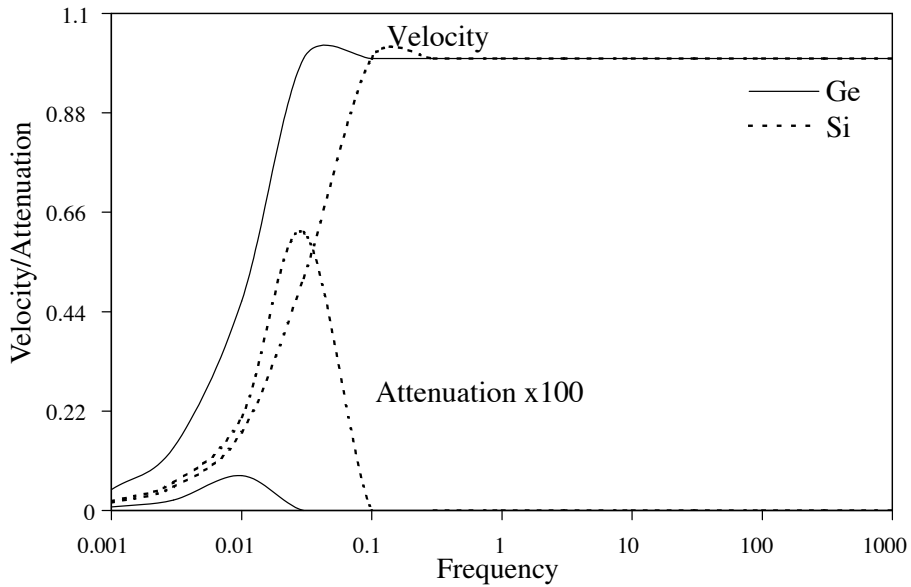**Figure 8.** Attenuation coefficient profile of T-mode waves.

Figure 9. Phase velocity and attenuation coefficient profiles of elasto-diffusive (EN/EP) waves.
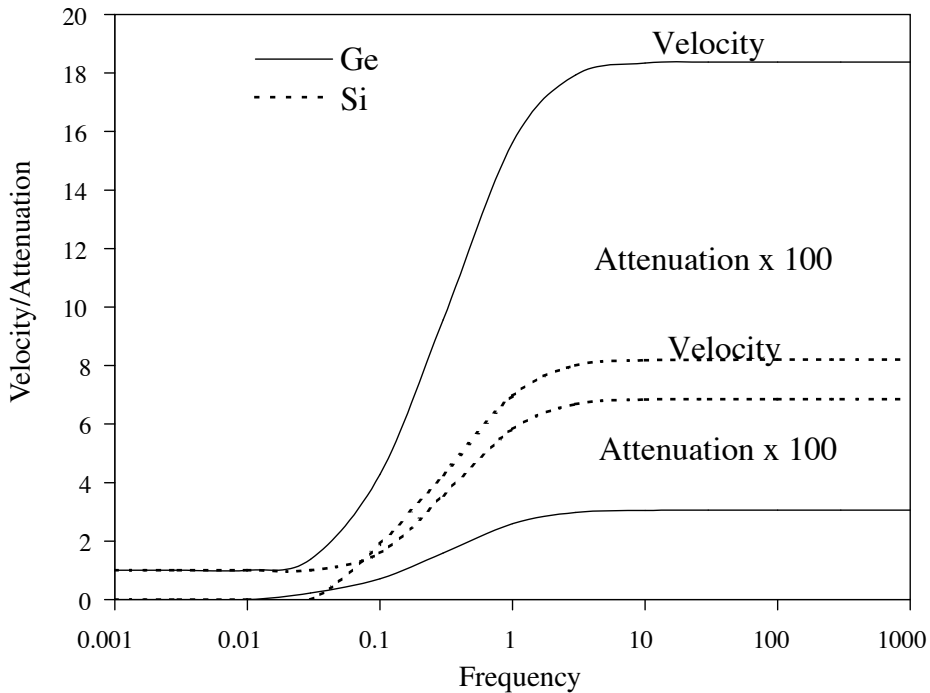

Figure 10. Phase velocity and attenuation coefficient profiles of electron/hole diffusive (N/P) waves.

Figure 11. Phase velocity and attenuation coefficient profiles of thermodiffusive (TN/TP) waves.

Figure 11 depicts phase velocity and attenuation coefficient profiles of N/P waves. In the frequency range $0 \leq \omega \leq 10$, a zigzag type of behavior of phase velocity profiles is noticed and it varies linearly for $\omega \geq 10$. For higher frequency ($\omega \geq 10$) velocity, the profiles for both materials overlap each other, meaning that there is no physical distinction between Ge and Si profiles. It is further observed that there is a sharp increase in the values of attenuation coefficient in the frequency range $0.01 \leq \omega \leq 10$, which become steady/fixed at $\omega \geq 10$ in both semiconductors. Figure 12 shows the phase velocity and attenuation coefficient profiles of electron and hole (N/P) diffusive waves with respect to nondimensional frequency. The behavior of these profiles is more or less similar to that in Figure 9, except for some variations in the behavior of phase velocity in $0 \leq \omega \leq 0.03$ and that in the magnitude of attenuation coefficient in addition to its existence beyond $\omega \geq 0.1$ here. The peak values of attenuation are observed to be at $\omega = 0.03$ and $\omega = 0.1$ for Ge and Si, respectively, instead of $\omega = 0.01$ in Figure 9.

Figure 13 deals with the phase velocity profiles of ET waves in Ge and Si materials. The nondimensional phase velocity in Ge materials has value unity at low frequency ($\omega \ll 1$) range, which comes down to its isothermal value at high-frequency ($\omega \gg 1$) limits. Thus in the Ge semiconductor, the phase velocity lies between isentropic and isothermal values as already established by many authors, such as [Chadwick 1979; Chadwick and Seet 1970; Sharma et al. 2000] and [Sharma and Singh 1989]. For the Si material the phase velocity varies linearly throughout with value unity at all frequencies and hence ET waves in this case travel without dispersion, irrespective of isentropic or isothermal conditions. Figure 14 represents attenuation coefficient profiles of ET waves. The nature of the attenuation profiles is the same

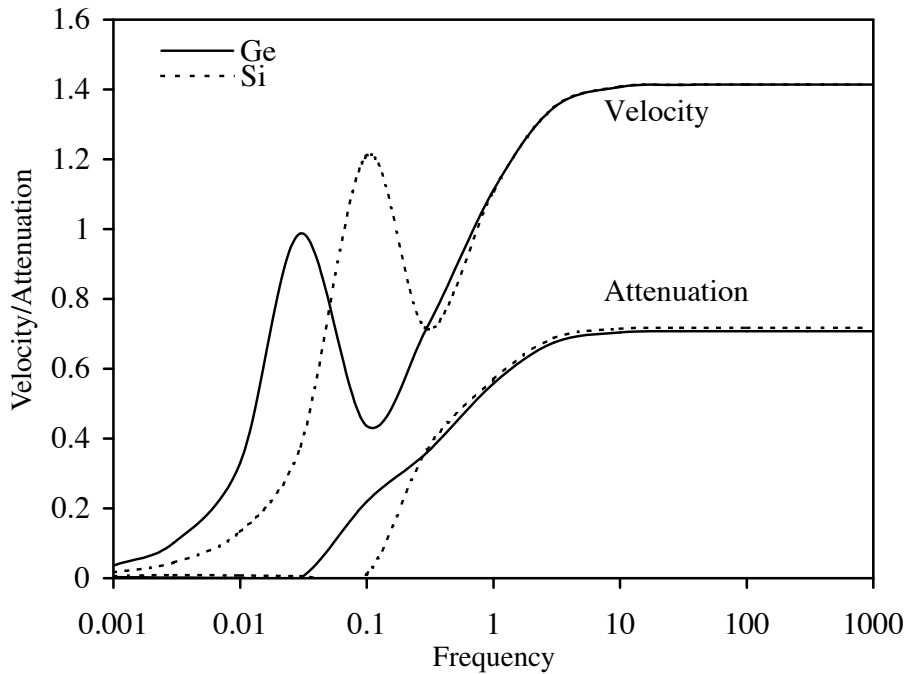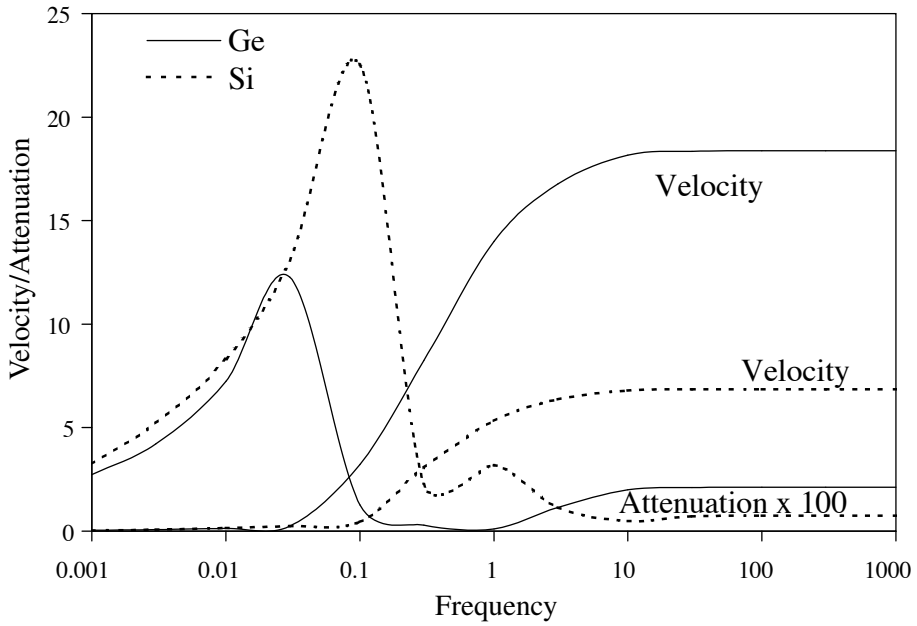**Figure 12.** Phase velocity and attenuation coefficient profiles of electron/hole diffusive (N/P) waves.
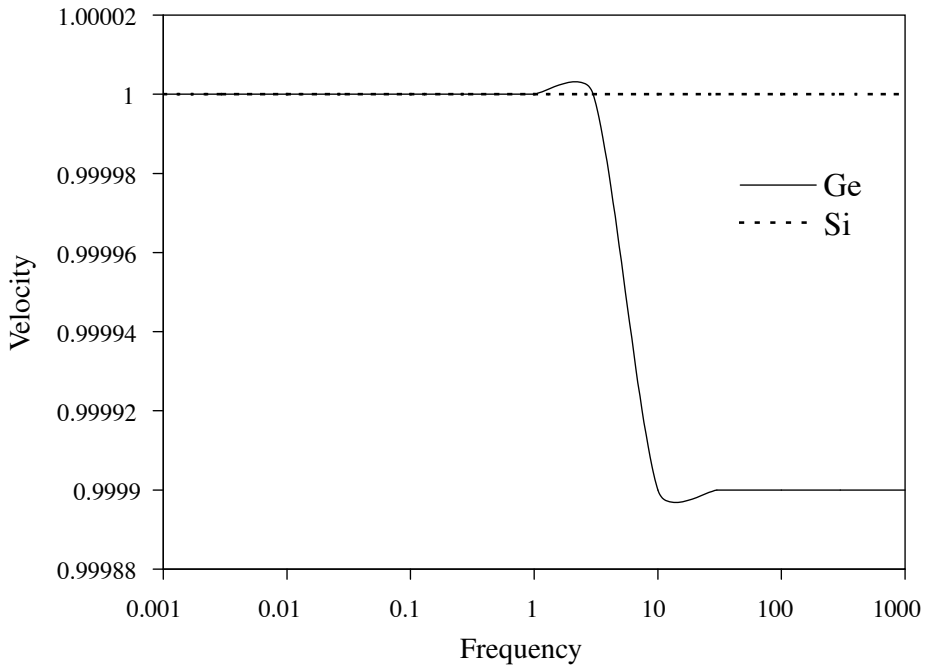
Figure 13. Phase velocity profile of ET waves.
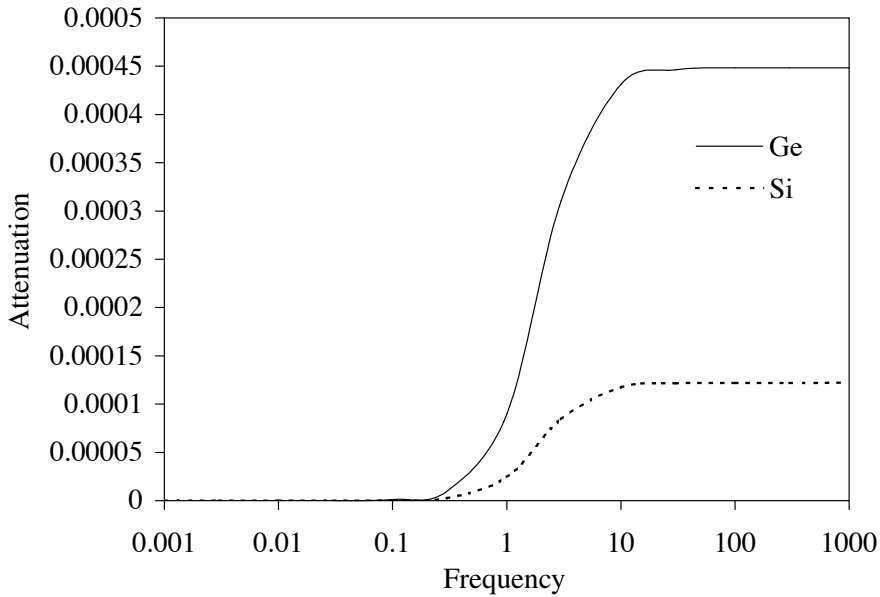
Figure 14. Attenuation coefficient profile of ET waves.

as that in Figure 10 except that waves have less attenuation in the case of Si than Ge here, in addition to negligible magnitude of attenuation coefficient. The nondimensional phase velocity and attenuation coefficient profiles in the case of Ge and Si semiconductors are given in Figure 15. Both the quantities
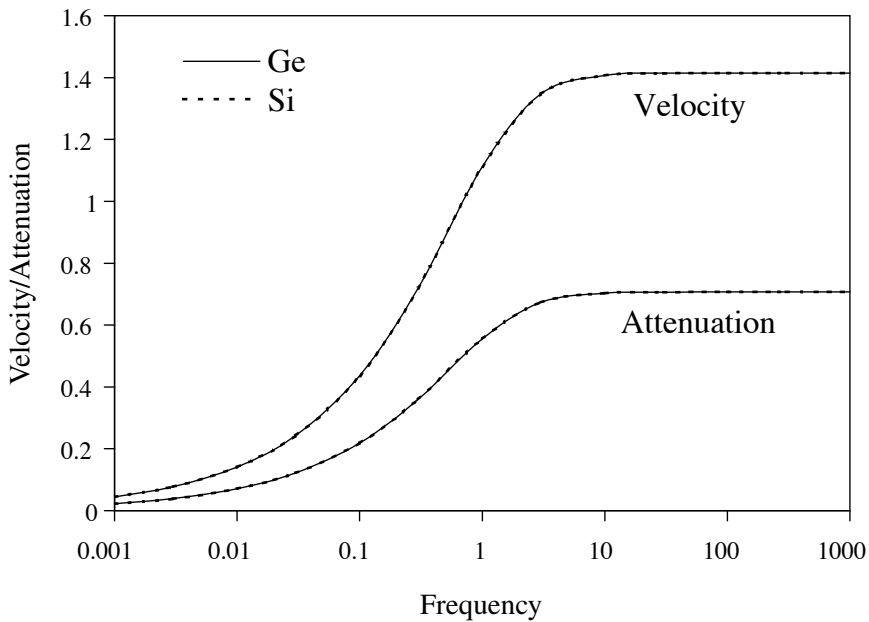
Figure 15. Phase velocity and attenuation coefficient profiles of T-mode waves.

are observed to vary logistically and the phenomenon is closer to wave motion than diffusion one. The profiles in case of Ge and Si materials overlap each other because of negligibly small distinction in their behavior and difference in their values.

The comparison of Figures 1, 2, 9, 13, and 14 suggests that the interactions of mechanical, thermal and electron/hole charge carrier fields have attributed to significant modifications in the values of phase velocity and attenuation coefficients of elastic, thermal and diffusive waves in the low and high frequency ranges. While the phase velocity has been lower down in its value in both germanium and silicon semiconductors, the attenuation coefficient has increased manifold and shifts in the value of critical frequency (ω^c) from $\omega^c = 1$ to $\omega^c = 10$ in case of Si and to $\omega^c = 100$ for Ge semiconductor. Similarly, the comparison of Figures 3, 4, 10 and 12 leads to the conclusion that the interaction of all the above fields with each other results in a fourfold increase in the phase velocity and an increase of four orders of magnitude in the attenuation coefficient of QEN/QEP waves in addition to phase shifts/changes. Figures 5, 6, 7, 8, 11, and 15 reveal that the magnitude of nondimensional phase velocity and attenuation coefficient of thermal waves has increased tenfold to that of T-mode due to the considered effect of various fields. The nature of this quantity has become closer to wave phenomena in contrast to diffusion as in Figure 15.

Appendix

The quantities

$$\Delta_i, \quad i = 1, 2, 3,$$
$$\Delta_i', \quad i = 1, 2, 3, 4, 5, 6,$$
$$\Delta_i'', \quad i = 1, 2, 3, 4,$$
$$\Delta_1''' \quad \text{and} \quad \Delta,$$

used in equations (10) are defined as

$$\Delta = 1 - \varepsilon^{nq}\varepsilon^{qn} - \varepsilon^{pq}\varepsilon^{qp}, \qquad \Delta_3 = \tau^Q - \varepsilon_n\varepsilon^{nq} - \varepsilon_p\varepsilon^{pq},$$

$$\Delta_1 = (1 - \varepsilon^{qn}\varepsilon^{nq})(\tau_p^* + i\omega^{-1}\varepsilon_p\varepsilon_T\overline{\lambda_p})$$
$$- \varepsilon^{qp}\{\tau_p' + \overline{\lambda_p}\tau^Q\varepsilon_T - i\omega^{-1}\varepsilon_n\varepsilon^{nq}(\alpha_0^n + \overline{\lambda_p}\varepsilon_T)\},$$

$$\Delta_2' = (\tau^Q - \varepsilon_n\varepsilon^{nq})(\tau_p^* + i\omega^{-1}\varepsilon_p\varepsilon_T\overline{\lambda_p})$$
$$- \varepsilon_p\{\tau_p' + \overline{\lambda_p}\tau^Q\varepsilon_T - i\omega^{-1}\varepsilon_n\varepsilon^{nq}(\alpha_0^n + \overline{\lambda_p}\varepsilon_T)\},$$

$$\Delta_4' = \tau_p^*(1 - \varepsilon^{qn}\varepsilon^{nq}) - \varepsilon^{qp}(\tau_p' - i\omega^{-1}\varepsilon_n\varepsilon^{nq}\alpha_0^n),$$

$$\Delta_6' = \tau^Q - i\omega^{-1}(\varepsilon_n\varepsilon^{nq} + \varepsilon_p\varepsilon^{pq}),$$

$$\Delta_3'' = \tau_p^*(\tau^Q - i\omega^{-1}\varepsilon_n\varepsilon^{nq}) - i\omega^{-1}\varepsilon_p(\tau_p' - i\omega^{-1}\varepsilon_n\varepsilon^{nq}\alpha_0^n),$$

$$\Delta_2'' = (\tau_n^* - \varepsilon^{qn}\tau_n')(\tau_p^* - \varepsilon^{qp}\tau_p')$$
$$- \{(i\omega^{-1}\varepsilon_p\alpha_0^p - \varepsilon^{qp}\tau_n')(i\omega^{-1}\varepsilon_n\alpha_0^n - \varepsilon^{qn}\tau_p')\},$$

$$\Delta_1' = \left\{(\tau_p^* + i\omega^{-1}\varepsilon_p\varepsilon_T\overline{\lambda_p}) - \varepsilon^{qp}(\tau_p' + \varepsilon_T\overline{\lambda_p}\tau^Q)\right\}$$
$$\times \left\{(\tau_n^* + i\omega^{-1}\varepsilon_n\varepsilon_T\overline{\lambda_n}) - \varepsilon^{qn}(\tau_n' + \varepsilon_T\overline{\lambda_n}\tau^Q)\right\}$$
$$- \left\{i\omega^{-1}\varepsilon_p(\alpha_0^p + \overline{\lambda_n}\varepsilon_T) - \varepsilon^{qp}(\tau_n' + \varepsilon_T\overline{\lambda_n}\tau^Q)\right\}$$
$$\times \left\{(i\omega^{-1}\varepsilon_n(\alpha_0^n + \overline{\lambda_p}\varepsilon_T) - \varepsilon^{qn}(\tau_p' + \varepsilon_T\overline{\lambda_p}\tau^Q)\right\},$$

$$\Delta_1''' = \left(\tau_n^* - i\omega^{-1}\varepsilon_n\frac{\tau_n'}{\tau^Q}\right)\left(\tau_p^* - i\omega^{-1}\varepsilon_p\frac{\tau_p'}{\tau^Q}\right) + \omega^{-2}\varepsilon_n\varepsilon_p\left(\alpha_0^n - \frac{\tau_p'}{\tau^Q}\right)\left(\alpha_0^p - \frac{\tau_n'}{\tau^Q}\right),$$

$$\Delta_1'' = \left((\tau_p^* + i\omega^{-1}\varepsilon_p\varepsilon_T\overline{\lambda_p}) - \frac{\varepsilon_p(\tau_p' + \varepsilon_T\overline{\lambda_p}\tau^Q)}{\tau^Q}\right)$$
$$\times \left((\tau_n^* + i\omega^{-1}\varepsilon_n\varepsilon_T\overline{\lambda_n}) - \frac{\varepsilon_n(\tau_n' + \varepsilon_T\overline{\lambda_n}\tau^Q)}{\tau^Q}\right)$$
$$- \left(i\omega^{-1}\varepsilon_p(\alpha_0^p + \varepsilon_T\overline{\lambda_n}) - \frac{\varepsilon_p(\tau_n' + \varepsilon_T\overline{\lambda_n}\tau^Q)}{\tau^Q}\right)$$
$$\times \left(i\omega^{-1}\varepsilon_n(\alpha_0^n + \varepsilon_T\overline{\lambda_p}) - \frac{\varepsilon_n(\tau_p' + \varepsilon_T\overline{\lambda_p}\tau^Q)}{\tau^Q}\right),$$

where

$$\tau^Q = t^Q + i\omega^{-1},$$

$$\varepsilon_n' = \frac{\varepsilon_n}{\varepsilon^{qn}},$$

$$\varepsilon_p' = \frac{\varepsilon_p}{\varepsilon^{qp}},$$

$$\tau_n' = t^Q\alpha_0^n + i\omega^{-1}(\alpha_0^n + a_0^n) - a_0^n\omega^{-2}/t_n^+,$$

$$\tau_p' = t^Q\alpha_0^p + i\omega^{-1}(\alpha_0^p + a_0^p) - a_0^p\omega^{-2}/t_p^+,$$

$$\tau_n^* = \frac{K}{\rho C_e D^n}\left(t^n + i\omega^{-1}\left(1 - \frac{\in_n \alpha_0^n D^n}{k} - \frac{t^n}{t_n^+}\right) + \frac{1}{\omega^2 t_n^+}\right),$$

$$\tau_p^* = \frac{K}{\rho C_e D^p}\left(t^p + i\omega^{-1}\left(1 - \frac{\in_p \alpha_0^p D^p}{k} - \frac{t^p}{t_p^+}\right) + \frac{1}{\omega^2 t_p^+}\right).$$

The other coefficients Δ_2, Δ_3', Δ_5' and Δ_4'' can be written from Δ_1, Δ_2', Δ_4' and Δ_3'' respectively, by replacing n with p and vice-versa.

Acknowledgments

References

[Achenbach 1973] J. D. Achenbach, *Wave propagation in elastic solids*, North-Holland, Amsterdam, 1973.

[Ackerman and Overton 1969] C. C. Ackerman and W. C. Overton, "Second sound in solid helium-3", *Phys. Rev. Lett.* **22**:15 (1969), 764–766.

[Ackerman et al. 1966] C. C. Ackerman, B. Bertman, H. A. Fairbank, and R. A. Guyer, "Second sound in solid helium", *Phys. Rev. Lett.* **16**:18 (1966), 789–791.

[Banerjee and Pao 1974] D. K. Banerjee and Y. H. Pao, "Thermoelastic waves in anisotropic solids", *J. Acoust. Soc. Am.* **56**:5 (1974), 1444–1454.

[Chadwick 1979] P. Chadwick, "Basic properties of plane harmonic waves in prestressed heat conducting elastic material", *J. Therm. Stresses* **2** (1979), 193–214.

[Chadwick and Seet 1970] P. Chadwick and L. T. C. Seet, "Wave propagation in transversely isotropic heat conducting elastic materials", *Mathematica* **17** (1970), 255–274.

[Chandrasekharaiah 1986] D. S. Chandrasekharaiah, "Thermoelasticity with second sound: a review", *Appl. Mech. Rev.* **39** (1986), 355–376.

[Dhaliwal and Sherief 1980] R. S. Dhaliwal and H. H. Sherief, "Generalized thermoelasticity for anisotropic media", *Q. Appl. Math.* **38** (1980), 1–8.

[Green and Lindsay 1972] A. E. Green and K. A. Lindsay, "Thermoelasticity", *J. Elasticity* **2**:1 (1972), 1–7.

[Guyer and Krumhansl 1966] R. A. Guyer and J. A. Krumhansl, "Thermal conductivity, second sound and phonon, hydrodynamic phenomena in nonmetallic crystals", *Phys. Rev.* **148**:2 (1966), 778–788.

[Lal 1995] S. Lal, *Fundamental physics*, S. Vikas, New Delhi, 1995.

[Lord and Shulman 1967] H. W. Lord and Y. Shulman, "A generalized dynamical theory of thermoelasticity", *J. Mech. Phys. Solids* **15**:5 (1967), 299–309.

[Many et al. 1965] A. Many, J. Goldstein, and N. B. Grover, *Semiconductor surfaces*, North-Holland, Amsterdam, 1965.

[Maruszewski 1986a] B. Maruszewski, "Electro-magneto-thermo-elasticity of extrinsic semiconductors: Classical irreversible thermodynamic approach", *Arch. Mech.* **38** (1986), 71.

[Maruszewski 1986b] B. Maruszewski, "Electro-magneto-thermo-elasticity of extrinsic semiconductors. Extended irreversible thermodynamic approach", *Arch.Mech.* **38** (1986), 83.

[Maruszewski 1986c] B. Maruszewski, *Termodynamiczne podstawy elektrotermodyfuzji magnetotermodyfuzji wosrodku ciaglym*, Wydawnictwa Naukoweftechniki Poznanskiej, Poznań, 1986.

[Maruszewski 1987a] B. Maruszewski, "Coupled evolution equations of deformable semiconductors", *Int. J. Eng. Sci.* **25**:2 (1987), 145–153.

[Maruszewski 1987b] B. Maruszewski, "Heat and charge carrier relaxation in deformable semiconductors", in *Electromagnetomechanical interactions in deformable solids and structures* (Tokyo, 1986), edited by Y. Yamamoto and K. Miya, North-Holland, Amsterdam, 1987.

[Maruszewski 1989] B. Maruszewski, "Thermodiffusive surface waves in semiconductors", *J. Acoust. Soc. Am.* **85**:5 (1989), 1967–1977.

[Scott 1989] N. H. Scott, "The stability of plane waves in generalized thermoelasticity", pp. 623–628 in *Elastic wave propagation: Proceedings of the Second IUTAM-IUPAP Symposium on Elastic Wave* (Galway, 1988), edited by M. F. McCarthy and M. A. Hayes, North-Holland, Amsterdam, 1989.

[Sharma 1986] J. N. Sharma, "On the low and high-frequency behavior of generalized thermoelastic waves", *Arch. Mech.* **38** (1986), 665–673.

[Sharma and Singh 1989] J. N. Sharma and H. Singh, "Generalized thermoelastic waves in anisotropic media", *J. Acoust. Soc. Am.* **85**:4 (1989), 1407–1413.

[Sharma and Singh 1990] J. N. Sharma and H. Singh, "Propagation of generalized thermoelastic waves in cubic crystals", *Arch. Mech.* **42** (1990), 19–30.

[Sharma et al. 2000] J. N. Sharma, V. Kumar, and S. P. Sud, "Plane harmonic waves in orthorhombic thermoelastic materials", *J. Acoust. Soc. Am.* **107**:1 (2000), 293–305.

[Sze 1981] S. M. Sze, *Physics of semiconductor devices*, 2nd ed., Wiley, New York, 1981.

[Zambuto 1989] M. Zambuto, *Semiconductor devices*, McGraw-Hill, New York, 1989.

JAGAN NATH SHARMA: jns@recham.ernet.in
Department of Applied Sciences, National Institute of Technology, Hamirpur 177 005, Himachal Pradesh, India

NAVEEN THAKUR: rinkalthakur@yahoo.com
Department of Applied Sciences, National Institute of Technology, Hamirpur 177 005, Himachal Pradesh, India

THERMOMECHANICAL FORMULATION OF STRAIN GRADIENT PLASTICITY FOR GEOMATERIALS

JIDONG ZHAO, DAICHAO SHENG AND IAN F. COLLINS

Constitutive formulation of strain gradient plasticity for geomaterials via a thermomechanical approach is investigated in this paper. It is demonstrated that, by defining two thermodynamical potentials (a free-energy function and a rate of dissipation function), the entire constitutive behavior of a decoupled strain-gradient-dependent material may be determined. The elastic relations are dependent on the free-energy function, while the plastic yielding and flow rule are determined by the dissipation function in conjunction with the free-energy function. Yield surfaces in both dissipative stress and true stress spaces may be derived without difficulty. Nonassociative flow rules and possible micromechanical mechanisms for the difference between plastic work and rate of plastic dissipation are interpreted for gradient-dependent materials. Using the obtained formulations and choosing appropriate thermodynamical functions, a wide variety of strain gradient plasticity models in the literature are recovered. Typical features associated with geomaterials, such as pressure and Lode-angle dependency, are addressed in detail. This paper provides a general thermodynamically-consistent framework of developing strain gradient plasticity models for geomaterials.

1. Introduction

A large number of microscale experiments on metallic materials have been done in recent years and have demonstrated strong size effects in solids (see [Fleck et al. 1994; Nix and Gao 1998; Tsagrakis and Aifantis 2002] and references therein). In most cases, the introduction of one or more gradient-dependent internal length scales of the deformation field is necessary to qualitatively and quantitatively interpret the behavior of size effects. Classical continuum theories fail to address the problem due to their local assumptions and lack of length scale(s) in the constitutive descriptions. The same reason accounts for their inability to determine the shear band size and describe the post-localization behavior in localization problems, which are frequently observed in metals and geomaterials. Modeling of strain localization by classical theories may generally lead to such consequences as the loss of ellipticity for the governing equations and spurious mesh-dependency problems in computation. In fact, length scales exist prevalently in a material and the loads applied to it, in the form of either the characteristic size of the homogeneous deformation domain (*internal length scale*), or the wavelength of a harmonic external load (*external characteristic length scale*) [Eringen and Kafadar 1976]. The internal length scale is closely related to the microscopically discontinuous structures or particles underlying a macroscopically continuous material body, such as the atomic lattice spacing in crystals, the grain size of polycrystals, and the grain diameter in granular materials, etc., whereas the external length scale is dependent on the properties of a loading force. The application regime of classical continuum theories is actually bounded

by the ratio of the external characteristic length scale over the internal length scale. When this ratio is much larger than 1, the details of the interactions between the microstructures can be neglected, and classical local continuum theories may be successfully applied. However, when this ratio approaches 1, the influence of microstructures is no longer negligible and classical models become inadequate to account for the mechanical behavior. Consequently gradient-enhanced generalized continuum theories are needed to take microstructural effects into account [Chambon et al. 2004].

A variety of gradient-enhanced theories have thus been proposed in the literature to address the problems mentioned above associated with classical theories and to account for the influence of microstructures, such as the Cosserat (micropolar) continuum theory [Cosserat and Cosserat 1909], the theory of micromorphic continua [Toupin 1962; Mindlin 1964; 1965; Germain 1973], the nonlocal gradient plasticity pioneered in [Aifantis 1984; 1987; Zbib and Aifantis 1988b; 1988a] and Fleck and Hutchinson's [1993; 1997] flow theory of gradient plasticity and its further extensions. For a comprehensive review of gradient theories, see [Chambon et al. 2004; Voyiadjis and Abu Al-Rub 2005]. In general, the existing gradient theories typically include higher-order gradient terms with coefficients that represent length-scale measures of microstructural deformation in the constitutive equations, and are mostly employed to investigate metal plasticity behavior. Compared with metals, the gradient-dependent characteristics in geomaterials are also evident. Geomaterials often contain microstructures such as grains, microvoids and microcracks. The basic micromechanical particles in geomaterials are much larger than the corresponding particles in metals, so microstructure is of much greater significance to continuum models of geomaterials than it is to metals. Higher gradient theories with internal length scales that relate the microstructures in geomaterials with the macroscopic mechanical behavior are thus necessary. The gradient dependency in geomaterials is also experimentally evident. For example, the experimental observation of localization phenomena in granular soils has indicated that the deformed material is usually characterized by strong spatial density variation. Formation of shear bands, a finite-thickness material zone with increased porosity, has been observed in sands [Vardoulakis and Graf 1985]. In such a region of strong spatial variation, higher gradients of appropriate physical properties of the material should be of vital importance. In addition, many other features of geomaterials can be attributed to micromechanical behaviors of structures on the microscale. By using gradient-dependent constitutive models with internal length scales, the influence of these microstructures may be appropriately addressed.

However, due to the distinctive features exhibited by geomaterials in contrast with metals, direct application of the gradient theories developed for metals to geomaterials may result in great deviations. Typical geomaterials like rocks, soils and some other granular materials have long been recognized to be associated with special properties and mechanical behaviors, such as the property of multiphase media, nonlinear elasticity, isotropic hardening behavior, pressure-sensitive frictional behavior, plastic volumetric changes during plastic loading course, nonassociated flow laws, and progressive strength degradation. In developing gradient models for geomaterials, these features have to be addressed.

There have been a limited number of gradient-enhanced constitutive models specially developed for and applied to geomaterials. In this connection, micropolar Cosserat continuum theories have long been applied to granular materials [Mühlhaus 1986; Mühlhaus and Vardoulakis 1987; Pijaudier-Cabot and Bazant 1987] by including microrotational terms into constitutive formulations. However, introducing only microrotational terms in a micropolar continuum is not sufficient since it cannot describe the dilative deformation in a shear band or other localized bifurcation modes such as compaction bands

[Olsson 1999]. Vardoulakis and Aifantis [1989; 1991; 1994] extended the gradient plasticity model to account for frictional/dilatant materials by incorporating the second-order strain gradients into the flow rule, yield condition, and/or dilatancy condition. The principle of virtual work is used to derive the appropriate extra boundary conditions. Mühlhaus and Aifantis [1991] derived a variational formulation for the Aifantis' gradient plasticity in connection with shear band analysis. Fremond and Nedjar [1996] proposed a combined gradient and rate-dependent damage model for quasibrittle materials. Chambon et al. [1996; 1998; 2001; 2004] developed a single and multimechanism flow theory of gradient plasticity for geomaterials as a possible approach to the analysis of localized failure, based on the framework of [Germain 1973]. Matsushima et al. [2000; 2002] further presented a finite deformation version of their theory. Aifantis et al. [1999] and Oka et al. [2002] proposed a gradient-dependent viscoplastic constitutive model for water-saturated clay, with gradients of the volumetric viscoplastic strain being included in the constitutive equations, while di Prisco et al. [2002] modified a preexisting elastoviscoplastic model for granular soils according to gradient and nonlocal approaches. Based on the framework of [Mindlin 1964; Fleck and Hutchinson 1993; 1997], Zhou and coworkers [Zhou et al. 2002; Zhao et al. 2005] developed a strain-gradient-enhanced damage model for rock-like geomaterials with application to shear band analysis. [Zhao and Sheng 2006] further constructed a framework of strain gradient plasticity by an internal-variable approach with normality structures to account for the influence of microstructures on the overall macroscopic mechanical behavior of gradient-dependent solids. Effects of strain gradient have also been used to account for the friction-controlled interparticle slip and particle rearrangement for granular materials [Valanis 1996; Walsh and Tordesillas 2004].

In developing the above gradient-dependent constitutive relations for geomaterials, a general routine for constitutive development has been followed by proposing observation-based constitutive relations first and then imposing the laws of thermodynamics on these relations. However, the modern theory of thermomechanics, as expounded, for example, in [Ziegler 1975; 1983; Houlsby 1981; Maugin 1992; Coussy 1995; Collins and Houlsby 1997; Rajagopal and Srinivasa 1998a; 1998b; Houlsby and Puzrin 2000; Puzrin and Houlsby 2001], develops constitutive models by first guaranteeing the fulfillment of these laws. Additional internal variables other than the general state variables such as plastic strain or generalized stress variables will then be used to characterize the material behavior. It is suggested that the total constitutive relations be determined by merely two thermodynamic potentials: a specific free-energy function and a dissipation function. This theory places strong emphasis on the use of internal variables to describe the past history of the material. The first and second laws of thermodynamics are enforced in the formulation to accommodate the requirements for most of the constitutive models. This theory has also been systematically applied to soil mechanics modeling; see [Houlsby 1981; 1982; Collins 2005] and references therein). In particular, the thermomechanical framework has been proven to be useful in accommodating models for geotechnical materials which generally exhibit nonassociated plastic flow behavior, pressure-sensitivity, nonlinear elasticity and dilatancy. A variety of plasticity models like the Drucker–Prager criterion and critical state models have been recovered and extended.

Even though arguments still exist on some issues, this thermomechanical approach has shown great generality in constitutive developments with solid thermodynamic considerations. In recognition of the advantages associated with this approach, we will apply it to develop a framework of gradient theory for geomaterials, covering some of the important features of geomaterials associated with the influence of microstructures. This theory is essentially a continuum one in that the discrete granular structure

for granular materials is not specially treated. However, by taking into account gradient terms and introducing a length scale, the microstructures in the materials as well as the microscale inhomogeneities may be appropriately reflected. It is demonstrated that by the thermomechanical formulations developed in the next section, a large range of gradient-dependent constitutive models existing in literature may be recovered and special features of geomaterials may be conveniently addressed.

Due to the introduction of gradient terms, some second-order tensors that correspond to microlevel strains, as well as third-order tensors corresponding to microlevel strain gradients, may be chosen as state variables. However, it should be noted that, because of the dependence of elastic compliances on the stresses and internal variables for some materials ('*coupled*', as will be noted subsequently), the strain rates and strain gradient rates may be dependent on the current values of the microdeformations within the material body, such that their integrations, which make up the total strains and strain gradients, are generally path-dependent. In this case, elastic and plastic strains and strain gradients are no longer appropriate for selection as independent state variables. The only exception is for *decoupled* materials where the elastic compliances are independent of the internal variables. Such materials include most ductile metals and the critical state models with linear relationships in the double logarithmic $\ln e$-$\ln p$ space [Lubliner 1972; Collins and Houlsby 1997]. To avoid the excessive complication of the coupled case for materials, we hereafter restrict our discussions to *decoupled materials only*, for which the free-energy function may be defined by two separate terms that characterize the elastic free energy and stored plastic energy components. In this case, the elastic and plastic terms for both strain and strain gradient may be regarded as state variables.

2. Thermomechanical formulation for gradient-dependent geomaterials

2.1. *Development of constitutive relations.* In the following formulation, isothermal, small strain deformation and rate-independent processes are assumed. The mechanical behavior of a material is assumed to be determined once the (Helmholtz) free-energy function Ψ and the rate of dissipation function $\hat{\Phi}$ are specified. The former represents the stored energy, the latter the rate at which energy is being converted irreversibly into heat. Both functions are defined for unit volume. Since the dissipation rate is generally path dependent, it is different from the pure time-rate as expressed by $\dot{\Phi}$. Following the framework of strain gradient theory in [Mindlin 1964; Germain 1973; Fleck and Hutchinson 1993; 1997], the mechanical behavior of the material is assumed to be described by the strain $\boldsymbol{\varepsilon}$ and strain gradients $\boldsymbol{\eta}$, and their respective work-conjugate thermodynamic forces: Cauchy stress $\boldsymbol{\sigma}$ and higher-order stress $\boldsymbol{\tau}$. The first law of thermodynamics states that a variation in the free energy is equal to the variation in the work done on the unit domain and the heat flux into the domain such that

$$\sigma_{ij}\dot{\varepsilon}_{ij} + \tau_{ijk}\dot{\eta}_{ijk} = \dot{\Psi} + \hat{\Phi}, \tag{1}$$

where the free energy is a state function dependent on a set of state variables that can describe the past history of the material: $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n)$. Here the internal variable $\boldsymbol{\alpha}_i$ $(i = 1, \cdots, n)$ may be in the form of a scalar, vector, or tensor (of second order or higher). The generalized n internal variables in tensor form, as stated in [Collins and Houlsby 1997] and [Puzrin and Houlsby 2001], preserve the ability to model compression and shear effects separately in granular geomaterials, model anisotropy

features through use of multiple kinematic hardening laws, and describe the past history of an elastic-plastic material. A more general form of the free-energy function has been suggested by Rice [1971] to be a functional of the time history of the internal variables. The Helmholtz free-energy function $\Psi = \Psi(\boldsymbol{\varepsilon}, \boldsymbol{\eta}, \boldsymbol{\alpha})$ is expressed in the compound space of strain and strain gradients. If expressed in an alternative space of Cauchy stress and higher-order stress, a Gibbs free energy function $\Omega = \Omega(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ may be found, related to the Helmholtz free-energy function by an appropriate Legendre transformation, such as

$$\Omega = \Omega(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \Psi(\boldsymbol{\varepsilon}, \boldsymbol{\eta}, \boldsymbol{\alpha}) - \boldsymbol{\sigma} : \boldsymbol{\varepsilon} - \boldsymbol{\tau} \vdots \boldsymbol{\eta}. \tag{2}$$

The rate of dissipation function is assumed to be dependent on dissipative internal variables $\boldsymbol{\alpha}$ as well as their rates $\dot{\boldsymbol{\alpha}}$. For rate independent materials, $\hat{\Phi}$ is a homogeneous first-degree functional in the space of $\dot{\boldsymbol{\alpha}}$. For purely frictional materials, the strength parameters are dimensionless frictional angles, which is different from the yield stress for metals. It is thus necessary to include some component(s) of the current stress and/or higher-order stress in the expression for the rate of dissipation function [Collins and Kelly 2002; Collins 2005]. In consequence, the rate of dissipative function for gradient-dependent materials is assumed to have an expression of this form: $\hat{\Phi} = \hat{\Phi}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \dot{\boldsymbol{\alpha}})$.

The key assumption of the thermomechanical approach for constructing constitutive relations is that they are fully determined by two functions: a thermodynamic potential, such as the Helmholtz free-energy function, or any of the related potentials derived from Legendre transformations, and the rate of dissipation function. Here it is assumed that all aspects of the isothermal constitutive behavior of such rate-independent gradient-enhanced materials may be uniquely defined by the two functions $\Psi(\boldsymbol{\varepsilon}, \boldsymbol{\eta}, \boldsymbol{\alpha})$ and $\hat{\Phi}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \dot{\boldsymbol{\alpha}})$. Thus Equation (1) may be rewritten as

$$\sigma_{ij}\dot{\varepsilon}_{ij} + \tau_{ijk}\dot{\eta}_{ijk} = \frac{\partial\Psi}{\partial\varepsilon_{ij}}\dot{\varepsilon}_{ij} + \frac{\partial\Psi}{\partial\eta_{ijk}}\dot{\eta}_{ijk} + \left(\frac{\partial\Psi}{\partial\boldsymbol{\alpha}_k}\otimes\dot{\boldsymbol{\alpha}}_k + \frac{\partial\hat{\Phi}}{\partial\dot{\boldsymbol{\alpha}}_k}\otimes\dot{\boldsymbol{\alpha}}_k\right),$$

where Euler's theorem for homogeneous functions has been used to rewrite $\hat{\Phi}$, and where $\otimes$ denotes an appropriate inner tensor operator according to the order of the internal variables. For example, if the second-order tensor is adopted for the internal variables, $\otimes$ implies the dyadic inner production operator ':'. It is then simple to obtain

$$\sigma_{ij} = \frac{\partial\Psi}{\partial\varepsilon_{ij}}, \qquad \tau_{ijk} = \frac{\partial\Psi}{\partial\eta_{ijk}}, \qquad \boldsymbol{q}^k = -\frac{\partial\Psi}{\partial\boldsymbol{\alpha}_k} = \frac{\partial\hat{\Phi}}{\partial\dot{\boldsymbol{\alpha}}_k} \tag{3}$$

and

$$\hat{\Phi} = \boldsymbol{q}^k \otimes \dot{\boldsymbol{\alpha}}_k \text{ for } \dot{\boldsymbol{\alpha}}_k \neq 0, \tag{4}$$

where the Cauchy stress $\boldsymbol{\sigma}$ and the higher-order stress $\boldsymbol{\tau}$ are sometimes called quasiconservative stresses [Ziegler and Wehrli 1987], which are conjugate with the state variables of strains and strain gradients, and the dissipative stresses $\boldsymbol{q}^k$ are the work conjugates of the internal variables. The dissipation surfaces represented by $\hat{\Phi}$ in the (compound) space of $\dot{\boldsymbol{\alpha}}_k$ are star-shaped with respect to the origin and convex, and the dissipative stresses $\boldsymbol{q}^k$ lie in their outward normal [Ziegler and Wehrli 1987]. In deriving these relations, a weak form of Ziegler's orthogonality hypothesis has been used.

Recalling the Gibbs free-energy function defined in Equation (2) and the results obtained in (3) and (4), one may easily obtain

$$\sigma_{ij} = \frac{\partial \Psi}{\partial \varepsilon_{ij}}, \quad \eta_{ijk} = \frac{\partial \Psi}{\partial \tau_{ijk}}, \quad \boldsymbol{q}^k = \frac{\partial \Omega}{\partial \boldsymbol{\alpha}_k}. \tag{5}$$

Hence decomposition of the rates of strain and strain gradient leads to

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}_{ij}^e + \dot{\varepsilon}_{ij}^p, \quad \dot{\eta}_{ijk} = \dot{\eta}_{ijk}^e + \dot{\eta}_{ijk}^p,$$

where

$$\dot{\varepsilon}_{ij}^e = \frac{\partial^2 \Omega}{\partial \sigma_{ij} \partial \sigma_{kl}} \dot{\sigma}_{kl}, \quad \dot{\eta}_{ijk}^e = \frac{\partial^2 \Omega}{\partial \tau_{ijk} \partial \tau_{lmn}} \dot{\tau}_{lmn}, \tag{6}$$

$$\dot{\varepsilon}_{ij}^p = \frac{\partial^2 \Omega}{\partial \sigma_{ij} \partial \boldsymbol{\alpha}_k} \otimes \dot{\boldsymbol{\alpha}}_k, \quad \dot{\eta}_{ijk}^p = \frac{\partial^2 \Omega}{\partial \tau_{ijk} \partial \boldsymbol{\alpha}_k} \otimes \dot{\boldsymbol{\alpha}}_k, \tag{7}$$

denote the elastic and plastic rates for strains and strain gradients, respectively. The coefficients before $\dot{\sigma}_{kl}$ and $\dot{\tau}_{lmn}$ in Equation (6) are the instantaneous elastic and higher order elastic compliances for the material. They are generally dependent on the Cauchy stresses, higher-order stresses and on the internal variables as well. Total elastic and plastic parts of strains and strain gradients may be obtained by integrations of the expressions (6) and (7) over the entire loading course.

As far as decoupled materials are concerned, we further choose ε_{ij}^p and η_{ijk}^p to be the internal variables. A general form of free-energy function depends on both the elastic and plastic strains and strain gradients like this: $\Psi = \Psi(\varepsilon_{ij}^e, \varepsilon_{ij}^p, \eta_{ijk}^e, \eta_{ijk}^p)$. For the decoupled case, it can be written as the sum of an elastic term $\Psi^e(\varepsilon_{ij}^e, \eta_{ijk}^e)$, which depends on the elastic strains and elastic strain gradients only, and a plastic term $\Psi^p(\varepsilon_{ij}^p, \eta_{ijk}^p)$, which depends only on the plastic strains and plastic strain gradients. That is,

$$\Psi(\varepsilon_{ij}^e, \varepsilon_{ij}^p, \eta_{ijk}^e, \eta_{ijk}^p) = \Psi^e(\varepsilon_{ij}^e, \eta_{ijk}^e) + \Psi^p(\varepsilon_{ij}^p, \eta_{ijk}^p). \tag{8}$$

As ε_{ij}^p and η_{ijk}^p have been chosen as the internal variables, from Equation (3), their thermodynamic conjugate dissipative stresses σ_{ij}^d and τ_{ijk}^d may be written as

$$\sigma_{ij}^d = \frac{\partial \hat{\Phi}}{\partial \dot{\varepsilon}_{ij}^p}, \quad \tau_{ijk}^d = \frac{\partial \hat{\Phi}}{\partial \dot{\eta}_{ijk}^p}. \tag{9}$$

With the preceding assumptions, the instantaneous elastic moduli for both strains and strain gradients are independent of the plastic strains and plastic strain gradients. Consequently, the total work rate may be decomposed into two parts:

$$\hat{W} = \hat{W}^e + \hat{W}^p, \tag{10}$$

where $\hat{W}^e = \dot{\Psi}^e(\varepsilon_{ij}^e, \eta_{ijk}^e)$ and $\hat{W}^p = \dot{\Psi}^p(\varepsilon_{ij}^p, \eta_{ijk}^p) + \hat{\Phi}$.

We have

$$\dot{\Psi}^e(\varepsilon_{ij}^e, \eta_{ijk}^e) = \sigma_{ij} \dot{\varepsilon}_{ij}^e + \tau_{ijk} \dot{\eta}_{ijk}^e;$$

thus

$$\sigma_{ij} = \frac{\partial \Psi^e}{\partial \varepsilon_{ij}^e}, \quad \tau_{ijk} = \frac{\partial \Psi^e}{\partial \eta_{ijk}^e}. \tag{11}$$

The last several equations since Equation (8) lead to the relations

$$\sigma_{ij} = \rho_{ij} + \sigma_{ij}^d, \quad \tau_{ijk} = \pi_{ijk} + \tau_{ijk}^d, \tag{12}$$

where

$$\rho_{ij} = \frac{\partial \Psi^p}{\partial \varepsilon_{ij}^p} \quad \text{and} \quad \pi_{ijk} = \frac{\partial \Psi^p}{\partial \eta_{ijk}^p} \tag{13}$$

are the shift stress and higher-order shift stress, which may present sound physical interpretations in developing anisotropic, kinematic hardening gradient models.

As we can see, several stress measures have been defined in the course of this section, including Cauchy stress, higher-order stress, dissipative stress, shift stress and higher-order shift stress. It is necessary to clarify these stresses here. First, as is well known, in conventional theories, *Cauchy stress* refers to the stress term at infinitesimal deformation, and is used to express Cauchy's first and second laws of motion (for the latter case, together with couple stress). Cauchy stress denotes force per unit area of the deformed solid. Other definitions of stress also appear in constitutive descriptions, including Kirchhoff, nominal (First Piola–Kirchhoff) and Material (Second Piola–Kirchhoff) stress tensors. The latter three stress measures consider forces acting on the undeformed solid, and require knowledge not only of the behavior of the deformed state, but also of the predeformation state. For a problem involving infinitesimal deformation, all the aforementioned stress measures are equal. However, here in this paper, we use *Cauchy stress* to represent the usual stress as in conventional theory, merely in distinguishing from the higher-order stresses that are introduced afterward, not distinguishing it from the aforementioned Kirchhoff and other stresses. In addition, the assumption of small strain made in the beginning of this section also cancels the differences between Cauchy stress and the other three terms.

Second, the term *higher-order stress* follows the usage of [Toupin 1962; Mindlin 1964; 1965; Fleck and Hutchinson 1997], and addresses the additional stresses required by strain gradient theories in addition to Cauchy stress. This higher-order stress is a more general stress that includes the *couple stress* as a subset case, as it may address both cases of rotation gradient and stretch gradient, while couple stress applies to rotation gradient case only. The third point regards *dissipative stress*. In passing, the dissipative stresses are defined as the conjugates of internal variables [Ziegler and Wehrli 1987]. In conventional thermomechanics, it may denote a vector, a second order tensor such as Cauchy stress, or a set of such vectors and tensors. While in the framework of strain gradient theory, third-order tensors may also appear (as is shown in Equation (9)), depending on the choice of internal variables. The occurrence of shift stress and higher-order shift stress is due to the nonidentity of true compound stress space (constituted by Cauchy stress and higher-order stress) and compound dissipative stress space (constituted by usual dissipative stress and higher-order dissipative stress). More information regarding this point can be found in [Collins and Houlsby 1997].

Plastic work and plastic dissipation. The plastic work rate and the rate of dissipation are generally different from each other, contrary to the long-held understanding in soil mechanics [Collins and Kelly 2002; Collins 2005]. The second law of thermodynamics states that the rate of dissipation can never be negative, but the sign of the plastic work increment is not restricted [Mroz 1973; Lubliner 1990]. Thus, the two rates are not equal to each other in the general case. The difference between them was called the *stored plastic work*, and was thought to be related to the frozen elastic energy on the microscale. As for

strain gradient characterized materials, the plastic work and dissipation rates are given by

$$\hat{W}^p = \sigma_{ij}\dot{\varepsilon}^p_{ij} + \tau_{ijk}\dot{\eta}^p_{ijk} = \sigma_{ij}\frac{\partial^2\Omega}{\partial\sigma_{ij}\partial\boldsymbol{\alpha}_k}\otimes\dot{\boldsymbol{\alpha}}_k + \tau_{ijk}\frac{\partial^2\Omega}{\partial\tau_{ijk}\partial\boldsymbol{\alpha}_k}\otimes\dot{\boldsymbol{\alpha}}_k,$$

$$\hat{\Phi} = \boldsymbol{q}^k\otimes\dot{\boldsymbol{\alpha}}_k = \frac{\partial\Omega}{\partial\boldsymbol{\alpha}_k}\otimes\dot{\boldsymbol{\alpha}}_k.$$

(14)

For a decoupled material, the plastic strain ε^p_{ij} and plastic strain gradient η^p_{ijk} are chosen to be the internal state variables, with

$$\sigma^d_{ij} \quad \text{and} \quad \tau^d_{ijk}$$

being their conjugate dissipative stresses. The difference between the plastic work rate and dissipation rate is

$$\hat{W}^s = \hat{W}^p - \hat{\Phi} = \left(\sigma_{kl}\frac{\partial^2\Omega}{\partial\sigma_{kl}\partial\varepsilon^p_{ij}} - \sigma^d_{ij}\right)\dot{\varepsilon}^p_{ij} + \left(\tau_{lmn}\frac{\partial^2\Omega}{\partial\tau_{lmn}\partial\eta^p_{ijk}} - \tau^d_{ijk}\right)\dot{\eta}^p_{ijk},$$

where $\hat{W}^s$ is the stored plastic work rate for the gradient-dependent material. For the fully decoupled case, $\hat{W}^s = \dot{\Psi}^p$, or alternatively, $\hat{W}^p = \dot{\Psi}^p + \hat{\Phi}$. From a micromechanical point of view, the stored plastic work in gradient-dependent materials may depend on the microscopic material properties as well as the imposed deformation state. As discussed in [Benzerga et al. 2003], for brittle or quasibrittle materials that may be characterized by single crystals, the behavior of stored plastic work during unloading processes is essentially controlled by the density of statistical stored dislocations (SSDs) as well as geometrical necessary dislocations (GNDs), of which the latter have long been attributed to the micromechanical origin of strain gradient effects. As for granular Cosserat materials, part of plastic work may be used to drive the interparticle rotation, and upon unloading, part of the rotation deformation remains unrecovered. The unrecoverable plastic work associated with this rotation part together with the frozen part contributed from strains makes up the total stored plastic work for granular Cosserat materials.

It is easy to see, for arbitrary plastic deformation increments, that the following conditions hold if and only if $\hat{W}^s$ becomes zero:

$$\sigma^d_{ij} = \sigma_{kl}\frac{\partial^2\Omega}{\partial\sigma_{kl}\partial\varepsilon^p_{ij}}, \qquad \tau^d_{ijk} = \tau_{lmn}\frac{\partial^2\Omega}{\partial\tau_{lmn}\partial\eta^p_{ijk}}.$$

(15)

Taking (5) into account, Equation (15) may be rewritten as

$$\boldsymbol{q}^k = (\boldsymbol{q}_\sigma, \boldsymbol{q}_\tau), \qquad \boldsymbol{q}_\sigma = \frac{\partial\boldsymbol{q}_\sigma}{\partial\sigma_{ij}}\sigma_{ij}, \qquad \boldsymbol{q}_\tau = \frac{\partial\boldsymbol{q}_\tau}{\partial\tau_{ijk}}\tau_{ijk},$$

(16)

where $\boldsymbol{q}_\sigma = \partial\Omega/\partial\boldsymbol{\alpha}_\varepsilon$, $\boldsymbol{q}_\tau = \partial\Omega/\partial\boldsymbol{\alpha}_\eta$. Equation (16) implies that $\boldsymbol{q}^k$ are homogeneous functions of degree one in the compound space of Cauchy stress σ_{ij} and higher-order stress τ_{ijk}. In consequence, the rate of dissipation function $\hat{\Phi}$ is also a homogeneous first degree-function of degree one in the compound space of Cauchy stress and higher-order stress, such that the material may be called a *purely frictional material*. For a purely frictional material, there is no stored plastic work frozen in the elastic energy, and thus the plastic work is totally dissipated. Yield loci of such models always exhibit convex cones in three dimensions, as will be demonstrated in the next section for such criteria as the generalized Coulomb,

von Mises, Drucker–Prager and Matsuoka–Nakai criteria in the framework of the strain gradient theory. However, *frictional materials* refer to those whose dissipation function depends not only on the Cauchy stress and higher-order stress, but also on the function(s) of the plastic deformations (such as the consolidation pressure in soil mechanics). In conventional soil mechanics, these materials are generally treated by critical state models with typical yield loci of homothetic curves or surfaces, as in [Collins and Kelly 2002]. It will also be shown in the following sections, in the compound stress space of σ_{ij} and τ_{ijk}, that the yield loci of gradient-dependent frictional materials also exhibit a homothetic feature. A third category of materials is termed as *quasifrictional materials*, which, in addition to the true stresses, depends on material parameters that involve stresses (such as cohesion as for structured soils, or fracture toughness as in particle crushing).

Nonassociative flow rule. The nonassociated flow rule is a distinguished characteristic of geomaterials. In conventional plastic theory, the dissipation function is generally assumed to depend on the plastic strain and the plastic strain rate. Consequently, the assumption that the associated flow rule is normal to the yield surface is frequently made in this theory. However, geomaterials generally exhibit a quite different behavior when the direction of the plastic strain rate vector is not normal to the current yield surface. The most used associated flow rule for metals has been proved to be inadequate for geomaterials. Models with nonassociated flow rules have thus been developed in terms of separate yield functions and plastic potentials [Oda and Iwashita 1999; Lade 1988; Vardoulakis and Sulem 1995]. However, the general theory of thermomechanics can still be applied with necessary nonassociated flow rules whenever the dissipation potential depends on the current stress or on the total strain or elastic strain. It remains true for the case of frictional gradient-dependent geomaterials. As will be shown in the following, whenever the dissipation function has an explicit dependency on the true stresses, the normality property of flow rules in the dissipative stress space will be lost when being transferred to true stress space.

We follow the same procedure used in [Collins 2005] to derive the nonassociative flow rule for gradient-dependent geomaterials. As has been stated in Equation (3), the dissipative stresses q^k are expressed as the derivative of $\hat{\Phi}$ with respect to $\dot{\alpha}_k$. It is assumed a dual function $\tilde{F}(\sigma, \tau, \alpha, q)$ may be found by a Legendre transformation of $\hat{\Phi}$, such that $\dot{\alpha}_k$ may be expressed in terms of q^k. As has been stated, $\hat{\Phi}$ is a homogeneous function of degree one in α for rate-independent materials, since there is no characteristic time. For such a function the above transformation is singular. In such a case, the value of the Legendre dual of $\hat{\Phi}$ is identically zero [Collins and Houlsby 1997]:

$$\tilde{F}(\sigma, \tau, \alpha, q) = 0. \tag{17}$$

In addition, the relation between $\dot{\alpha}_k$ and q^k is not unique. Instead of having a unique expression, the dual relation to Equation (3) may be presented in a form for the time derivatives of the internal variables as follows:

$$\dot{\alpha}_k = \dot{\lambda}\frac{\partial \tilde{F}(\sigma, \tau, \alpha, q)}{\partial q^k}. \tag{18}$$

Equations (17) and (18) present a yield condition and associated flow rule in generalized stress space (or alternatively dissipative stress space). And the yield function in Equation (17) displays an obvious dependence on the true stresses σ and τ, which is inherited from the original form of $\hat{\Phi}$ and is always necessary for frictional materials. In true compound stress space of σ and τ, the yield condition may be

obtained as follows:

$$F(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}_k) = \tilde{F}\left(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}_k, \boldsymbol{q}^k = \frac{\partial \Omega(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\alpha}_k)}{\partial \boldsymbol{\alpha}_k}\right) = 0. \tag{19}$$

As we have

$$\frac{\partial F}{\partial \boldsymbol{\sigma}} = \frac{\partial \tilde{F}}{\partial \boldsymbol{\sigma}} + \frac{\partial^2 \Omega}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\alpha}_k} \otimes \frac{\partial \tilde{F}}{\partial \boldsymbol{q}^k}, \qquad \frac{\partial F}{\partial \boldsymbol{\tau}} = \frac{\partial \tilde{F}}{\partial \boldsymbol{\tau}} + \frac{\partial^2 \Omega}{\partial \boldsymbol{\tau} \partial \boldsymbol{\alpha}_k} \otimes \frac{\partial \tilde{F}}{\partial \boldsymbol{q}^k}. \tag{20}$$

In connection with Equations (6), (18), (19) and (20), the flow rule for the plastic rates of strain and strain gradient in the true compound stress space may be obtained as

$$\dot{\varepsilon}_{ij}^p = \dot{\lambda} \frac{\partial F}{\partial \boldsymbol{\sigma}} + \frac{\partial \hat{\Phi}}{\partial \boldsymbol{\sigma}}, \qquad \dot{\eta}_{ijk}^p = \dot{\lambda} \frac{\partial F}{\partial \boldsymbol{\tau}} + \frac{\partial \hat{\Phi}}{\partial \boldsymbol{\tau}}, \tag{21}$$

where

$$\frac{\partial \hat{\Phi}}{\partial \boldsymbol{\sigma}} = -\dot{\lambda} \frac{\partial \tilde{F}}{\partial \boldsymbol{\sigma}}, \qquad \frac{\partial \hat{\Phi}}{\partial \boldsymbol{\tau}} = -\dot{\lambda} \frac{\partial \tilde{F}}{\partial \boldsymbol{\tau}}.$$

These relations are obtained from the Legendre transformation between $\hat{\Phi}$ and $\tilde{F}$ in connection with Equation (18). As can be seen, when $\partial \tilde{F}/\partial \boldsymbol{\sigma}$ and $\partial \tilde{F}/\partial \boldsymbol{\tau}$ are not zero, Equation (21) clearly exhibits a natural feature of nonassociated flow rule. This is an obvious advantage of thermomechanical approach for addressing geomaterial behaviour.

3. Derivation of a class of $J2$-flow theory of strain gradient plasticity for isotropic incompressible geomaterials

In this section, the thermomechanical formulations obtained in last section are employed to derive a class of strain gradient plasticity models that may be regarded as extensions of J_2-flow theory for typical pressure-insensitive materials. Further implications of the formulations, including features suitable for describing pressure-sensitive geomaterials, will be addressed in the next section. Following [Fleck and Hutchinson 1993], it is assumed that couple stresses exist in the elastic-plastic body, with elastic strain, plastic strain and Cauchy stress being ε_{ij}^e, ε_{ij}^p and σ_{ij}. The strain gradient effects are manifested in a curvature tensor in its elastic and plastic parts:

$$\chi_{ij}^e = e_{ilm}\eta_{mjl}^e \quad \text{and} \quad \chi_{ij}^p = e_{ilm}\eta_{mjl}^p,$$

where e_{ilm} is the permutation tensor. The couple stress is denoted by m_{ij}, which may be related to the higher-order stress τ_{ijk} defined in last section as: $m_{ij} = e_{ilm}\tau_{mjl}$. In a conventional J_2 flow theory, it is generally regarded that the plastic strain is incompressible, such that the yield condition may be defined through the von Mises effective stress by the deviatoric part of the Cauchy stress s_{ij}: $\sigma_e = \sqrt{(3s_{ij}s_{ij})/2}$. In the presence of strain gradients and couple stresses, the plastic work conjugate forces for ε_{ij}^p and χ_{ij}^p are s_{ij} and m_{ij} respectively. In addressing incompressibility in the presence of strain gradients, Smyshlyaev and Fleck [1996] (see also [Fleck and Hutchinson 2001]) have made an orthogonal decomposition of the strain gradient tensor and obtained an acute incompressible tensor for strain gradients, in which the incompressibility condition for strain gradients is expressed as: $\eta_{kjj} = 0$ for $k = 1, 2, 3$. To avoid the excessive complexity of accounting for the general case of strain gradients, however, here in this paper

we assume the curvature tensor in conjunction with its work conjugate couple stress adequately fulfills the incompressibility condition for gradient effects, at least for a particular range of materials (such as Cosserat materials). In fact, due to the symmetrical nature of the strain gradient tensor, $\eta_{ijk} = \eta_{jik}$, the curvature tensor $\chi_{ii} = 0$, which implies the incompressibility of strain gradients.

Free energy function and elastic relations. We first suppose that the material can be characterized by the linear Gibbs free-energy function

$$\Omega = \tfrac{1}{2}C_{ijkl}\sigma_{ij}\sigma_{kl} + \tfrac{1}{2}M_{ijkl}l^{-2}m_{ji}m_{lk}, \tag{22}$$

where

$$C_{ijkl} = \frac{1+\upsilon}{2E}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) - \frac{\upsilon}{E}\delta_{ij}\delta_{kl}, \qquad M_{ijkl} = \frac{1+\upsilon}{E}\left(\frac{l}{l_e}\right)^2 \delta_{ik}\delta_{jl}, \tag{23}$$

and E and υ denote the Young's modulus and Poisson's ratio. The length scale l is interpreted as the free slip distance between statistically stored dislocations, while l_e is introduced to divide the curvature tensor χ_{ij} into elastic and plastic parts. It is assumed $l_e \ll l$ such that the dominant size effect in the material is associated with plastic rather than elastic strain gradients. It is easy to find that this definition of the free-energy function in Equations (22) and (23) results in a dependence of the instantaneous elastic modulus on the elastic state variables only, which corresponds to the special case of decoupled materials, where the shift stresses contributed from free energy are zero and the dissipative stresses coincide with the true stresses. In this case, the elastic behavior of the material may be totally determined by the Gibbs free-energy function, while the plastic behavior may be completely determined by a dissipative rate function. In view of Equation (6), we obtain from (22) the elastic incremental relations

$$\dot{\varepsilon}^e_{ij} = C_{ijkl}\dot{\sigma}_{kl}, \quad \dot{\chi}^e_{ij} = M_{ijkl}l^{-2}\dot{m}_{lk}. \tag{24}$$

Dissipation function and plastic relations. In defining a function for the rate of plastic dissipation, it is further assumed that the plastic parts of the strains and strain gradients ε^p_{ij} and χ^p_{ij}, may be regarded to be the internal variables. The following general form for the rate-dissipation function is assumed to be capable of appropriately describing the dissipative processes for a class of strain-gradient-dependent materials:

$$\hat{\Phi} = B\big(A_1(\dot{\varepsilon}^p_{ij}\dot{\varepsilon}^p_{ij}) + A_2(l^2\dot{\chi}^p_{ij}\dot{\chi}^p_{ij})\big)^{1/2}, \tag{25}$$

where the coefficients A_1, and A_2 are dimensionless constants. B is a scalar with a dimension of stress, with the possibility of being further extended to a complex function depending on plastic strains and strain gradients. Equation (25) is a generalization of the dissipation function for von Mises materials in [Ziegler and Wehrli 1987, Equation 6.2, p. 208], where the dissipation function is assumed to depend on the second invariant of the plastic strain rate only. In view of Equation (9), the dissipation stresses may be obtained by differentiating the dissipation function (25) with respect to the plastic strain increments and the curvature increments respectively, such that

$$\sigma^d_{ij} = \frac{\partial \hat{\Phi}}{\partial \dot{\varepsilon}^p_{ij}} = \frac{A_1 B^2}{\hat{\Phi}}\dot{\varepsilon}^p_{ij}, \qquad m^d_{ji} = \frac{\partial \hat{\Phi}}{\partial \dot{\chi}^p_{ij}} = \frac{A_2 B^2 l^2}{\hat{\Phi}}\dot{\chi}^p_{ij}. \tag{26}$$

The rate of dissipation function can always be written as

$$\hat{\Phi} = \sigma^d_{ij}\dot{\varepsilon}^p_{ij} + m^d_{ji}\dot{\chi}^p_{ij}. \tag{27}$$

Eliminating the rates of strain and strain gradient from Equations (26) and (27), we obtain for the yield function in dissipative stress space

$$\tilde{F} = \frac{(\sigma_{ij}^d \sigma_{ij}^d)}{A_1} + \frac{(l^{-2} m_{ij}^d m_{ij}^d)}{A_2} - B^2 = 0. \tag{28}$$

We will demonstrate here that a variety of strain gradient plasticity models proposed in the literature may be recovered by selecting different values for the coefficients A_1, A_2 and B for the dissipation function and, in conjunction, the elastic relations presented in (24), We identify several cases:

Case 1. $A_1 = A_2 = 2/3$, B is a scalar and $B \geq 0$.

In this case, the dissipative stresses are identical with the true stresses: $\sigma_{ij}^d = s_{ij}$, $m_{ij}^d = m_{ij}$. Manipulation of Equation (28) easily leads to the following yield surface in the true stress space:

$$F = \Sigma_e - B = 0, \tag{29}$$

where Σ_e is a generalized effective stress defined by

$$\Sigma_e = \sqrt{\tfrac{3}{2} s_{ij} s_{ij} + \tfrac{3}{2} l^{-2} m_{ij} m_{ij}}. \tag{30}$$

In conjunction with (26), the normality of the plastic flow rule of (29) results in

$$\dot{\varepsilon}_{ij}^p = \dot{\lambda} \frac{\partial F}{\partial s_{ij}} = \dot{\lambda} \frac{\partial F}{\partial \Sigma_e} \frac{\partial \Sigma_e}{\partial s_{ij}}, \qquad \dot{\chi}_{ij}^p = \dot{\lambda} \frac{\partial F}{\partial m_{ij}} = \dot{\lambda} \frac{\partial F}{\partial \Sigma_e} \frac{\partial \Sigma_e}{\partial m_{ij}},$$

where $\dot{\lambda} = \dot{\Sigma}_e / h$ and the instantaneous hardening rate $h = h(\Sigma_e)$ are chosen so that the uniaxial tensile response may be reproduced. We thus obtain, for the incremental plastic strains and strain gradients,

$$\dot{\varepsilon}_{ij}^p = \frac{3}{2h} \frac{s_{ij}}{\Sigma_e} \dot{\Sigma}_e, \qquad \dot{\chi}_{ij}^p = \frac{3}{2h} \frac{l^{-1} m_{ji}}{\Sigma_e} \dot{\Sigma}_e. \tag{31}$$

In view of the elastic incremental relations defined in Equation (24), the yield surface defined in (29), and the plastic incremental parts defined in (31), one may readily find that the J_2 flow version of the [Fleck and Hutchinson 1993] strain gradient theory is fully recovered (see Equations (52), (53), (52a), (52b), (52c) and (45) therein). This case can actually be termed an *elastic-perfectly plastic gradient material*, as noted in [Ziegler and Wehrli 1987].

Suppose the principal deviator stresses and principal couple stresses are s_1, s_2, s_3 and m_1, m_2, m_3 respectively. Define

$$\begin{aligned}
S_e^{(I)} &= \sqrt{(s_1^2 + l^{-2} m_1^2)} \operatorname{sgn}(s_1), \\
S_e^{(II)} &= \sqrt{(s_2^2 + l^{-2} m_2^2)} \operatorname{sgn}(s_2), \\
S_e^{(III)} &= \sqrt{(s_3^2 + l^{-2} m_3^2)} \operatorname{sgn}(s_3).
\end{aligned} \tag{32}$$

Obviously the yield locus defined by (29) is a circle with a radius $\sqrt{6}B/3$ in the π-plane defined by $S_e^{(I)}$, $S_e^{(II)}$ and $S_e^{(III)}$, as is shown in Figure 1. The scalar B serves here as a generalized yield stress in tension, which is $\sqrt{3}$ times the generalized yield stress in pure shear. For the case being considered

here, the dissipation stress space coincides with the true stress space so that the yield surfaces in the two stress spaces are identical. Based on similar formulations for the above J_2-flow theories with generalized von Mises yield criteria, other yielding criteria such as Tresca and maximum tension failure can also be formulated with ease, by defining different dissipation rate functions. For example, if we defined the rate of dissipation function as

$$\hat{\Phi} = B\left(\left|\dot{\gamma}_1^P\right| + \left|\dot{\gamma}_2^P\right| + \left|\dot{\gamma}_3^P\right|\right),$$

where $\dot{\gamma}_1^P = \dot{e}_1^P + l\dot{\omega}_1^P$, $\dot{\gamma}_2^P = \dot{e}_2^P + l\dot{\omega}_2^P$, $\dot{\gamma}_3^P = \dot{e}_3^P + l\dot{\omega}_3^P$, and $\dot{e}_i^P$ and $\dot{\omega}_i^P$ $(i = 1, 2, 3)$ are the principal distortion strains and deviatoric curvatures. Denoting the principal deviator stresses and principal couple stresses by s_1, s_2, s_3 and m_1, m_2, m_3, we have

$$s_i = \frac{\partial\hat{\Phi}}{\partial\dot{e}_i^P} = \frac{\partial\hat{\Phi}}{\partial\dot{\gamma}_i^P}\frac{\partial\dot{\gamma}_i^P}{\partial\dot{e}_i^P} = B\,\mathrm{sgn}\left(\dot{\gamma}_i^P\right),$$

$$m_i = \frac{\partial\hat{\Phi}}{\partial\dot{\omega}_i^P} = \frac{\partial\hat{\Phi}}{\partial\dot{\gamma}_i^P}\frac{\partial\dot{\gamma}_i^P}{\partial\dot{\omega}_i^P} = Bl\,\mathrm{sgn}\left(\dot{\gamma}_i^P\right).$$

Adopting the same definition as in (32) and noting that $\mathrm{sgn}(s_i) = \mathrm{sgn}(\dot{\gamma}_i^P)$, we have

$$S_e^{(I)} - S_e^{(II)} = B\left(\mathrm{sgn}(\dot{\gamma}_1^P) - \mathrm{sgn}(\dot{\gamma}_2^P)\right),$$

$$S_e^{(II)} - S_e^{(III)} = B\left(\mathrm{sgn}(\dot{\gamma}_2^P) - \mathrm{sgn}(\dot{\gamma}_3^P)\right),$$

$$S_e^{(I)} - S_e^{(III)} = B\left(\mathrm{sgn}(\dot{\gamma}_1^P) - \mathrm{sgn}(\dot{\gamma}_3^P)\right).$$

The yield surface thus defined is obviously a generalized Tresca Hexagon. If B again serves as a generalized yield stress in tension, in the π-plane defined by $S_e^{(I)}$, $S_e^{(II)}$ and $S_e^{(III)}$, this hexagon circumscribes the generalized von Mises circle as shown in Figure 1. In this case B is twice the generalized yield stress in pure shear, just as in the conventional Tresca condition.

Given the J_2 flow version of strain gradient theory as listed above, the deformation version of strain gradient theory of Fleck and Hutchinson [1993] may also be easily attained. In fact, if we assume that the equivalent plastic strain has the form

$$\bar{\varepsilon}_e = \sqrt{\tfrac{2}{3}\varepsilon_{ij}^P\varepsilon_{ij}^P + \tfrac{2}{3}l_2\chi_{ij}^P\chi_{ij}^P}, \tag{33}$$

the rate of dissipation function has the expression

$$\hat{\Phi} = B\dot{\bar{\varepsilon}}_e.$$

One can easily verify that

$$s_{ij} = \frac{\partial\hat{\Phi}}{\partial\dot{\varepsilon}_{ij}^P} = \frac{\partial\hat{\Phi}}{\partial\dot{\bar{\varepsilon}}_e}\frac{\partial\bar{\varepsilon}_e}{\partial\varepsilon_{ij}^P} = \frac{3}{2}\frac{B}{\bar{\varepsilon}_e}\varepsilon_{ij}^P,$$

$$m_{ij} = \frac{\partial\hat{\Phi}}{\partial\dot{\chi}_{ij}^P} = \frac{\partial\hat{\Phi}}{\partial\dot{\bar{\varepsilon}}_e}\frac{\partial\bar{\varepsilon}_e}{\partial\chi_{ij}^P} = \frac{3}{2}\frac{B}{\bar{\varepsilon}_e}l^2\chi_{ij}^P,$$
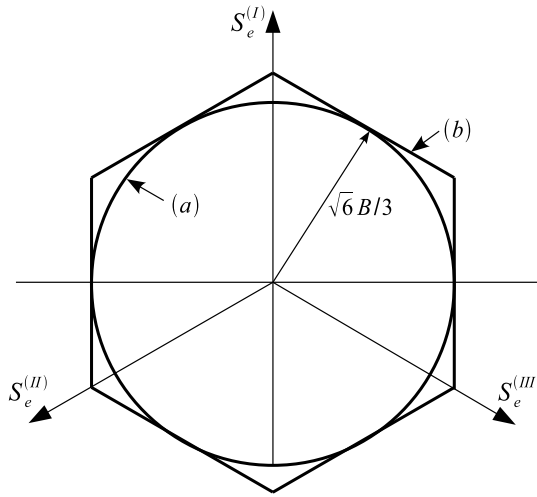
Figure 1. Yield loci of (*a*) the generalized von Mises and (*b*) generalized Tresca gradient-dependent materials in the π-plane.

where the coefficient B is actually equal to the generalized effective stress defined in Equation (30). Thus the deformation version of strain gradient theory proposed by Fleck and Hutchinson [1993] is again recovered.

For the J_2 flow version of strain gradient theory just obtained, a generalized form of Drucker's stability postulate [Drucker 1951] has been suggested by Fleck and Hutchinson when $h > 0$:

$$\dot{\sigma}_{ij}\dot{\varepsilon}_{ij}^p + \dot{m}_{ij}\dot{\chi}_{ij}^p \geq 0.$$

This result was formerly developed by Koiter [1960] for phenomenological plasticity theories with multiple yield functions, and by Hill [1966] in a more general form for a metal crystal deforming in multislip. From a thermomechanical view, the above inequality actually implies the change of plastic dissipative work is always pointing to the dissipating direction and this dissipative process is irreversible. While the widely used normality law is justified by Drucker's postulate, a more general stability postulate that may cover the nonnormality case has been proposed by Ziegler and Wehrli [1987], which, if extended to strain gradient theory, has the formulation

$$\dot{\sigma}_{ij}\dot{\varepsilon}_{ij}^i + \dot{\tau}_{ijk}\dot{\eta}_{ijk}^i \geq 0, \tag{34}$$

where

$$\dot{\varepsilon}_{ij}^i = \left(\frac{\partial^2 \mathbf{\Omega}}{\partial \sigma_{ij} \partial \varepsilon_{kl}^p}\right)\dot{\varepsilon}_{kl}^p,$$

$$\dot{\eta}_{ijk}^i = \left(\frac{\partial^2 \mathbf{\Omega}}{\partial \tau_{ijk} \partial \eta_{lmn}^p}\right)\dot{\eta}_{lmn}^p$$

may be termed as the irreversible strain rate and the irreversible strain gradient rate, respectively. Equation (34) applies to coupled materials as well as decoupled materials.

Case 2. $B = 0$.

As an extreme instance of Case 1, this case implies zero dissipation in the material. Consequently, no plastic strains and strain gradients occur in the deformation, and

$$l = l_e, \qquad \varepsilon_{ij} = \varepsilon_{ij}^e, \qquad \chi_{ij} = \chi_{ij}^e.$$

The total constitutive behavior may be determined by the elastic relations in (24) only. This is similar to the linear elastic strain gradient theory of [Mindlin 1964], where a higher-order elastic compliance tensor of the sixth order, M_{ijklmn}, was defined to relate the higher-order elastic stress τ_{lmn} with elastic strain gradient η_{ijk}, in the form $\eta_{ijk} = M_{ijklmn}\tau_{lmn}$. Equation (24) in this condition may be regarded as a special case of couple stress theory. For a general isotropic elastic case, the tensor M_{ijklmn} is dependent on five material constants and has general symmetry of $M_{ijklmn} = M_{lmnijk}$ [Mindlin 1964].

Case 3. $A_1 = 2$, $A_2 = 0$, and B is a nonnegative function to be specified.

In this case, ε_{ij}^p is regarded as the only internal variable. Incompressible solids are considered to have only the deviatoric stress s_{ij} contributing to the plastic work. χ_{ij}^p does not explicitly enter the rate of dissipation function as an independent internal variable. Instead, it is assumed to enter the coefficient function B in accounting for the accumulated plastic strain gradient effects. In other words, though the strain gradient may appear in the final constitutive formulations, no work-conjugate force is specified for it. Consequently, the terms containing gradients in the free energy function will also be dropped and the free-energy function in Equation (23) turns out to be

$$\Omega = \tfrac{1}{2} C_{ijkl}\sigma_{ij}\sigma_{kl}.$$

Only the first equation in (24) remains valid, and it can be recast as

$$\dot{\sigma}_{kl} = C_{ijkl_{ijkl}}^{-1}\dot{\varepsilon}_{ij}^e.$$

The rate of dissipation function defined in (25) becomes $\hat{\Phi} = B\big(A_1(\dot{\varepsilon}_{ij}^p\dot{\varepsilon}_{ij}^p)\big)^{1/2}$. And the dissipative stress conjugate with $\dot{\varepsilon}_{ij}^p$ is

$$\sigma_{ij}^d = \frac{\partial\hat{\Phi}}{\partial\dot{\varepsilon}_{ij}^p} = \frac{A_1 B^2}{\hat{\Phi}}\dot{\varepsilon}_{ij}^p. \tag{35}$$

Hence the yield function in the dissipation stress space becomes

$$F = \sqrt{(\sigma_{ij}^d\sigma_{ij}^d)/A_1} - B = 0.$$

Here we have assumed that B is a nonnegative function related with the history of plastic strain and plastic strain gradient. However, no rates of plastic strain and strain gradient are involved in B, so that Equation (35) always holds true.

In this case, the dissipative stress σ_{ij}^d coincides with the deviatoric stress s_{ij}. If the effective stress is defined as

$$\tau = \sqrt{s_{ij}s_{ij}/2},$$

the yield function in the true stress presents the form

$$F = \tau - B = 0. \tag{36}$$

Here B is further assumed to be of the form

$$B = \tau_0 \big(f(\gamma^P)^{r_1} + g(\ell^n \eta_n)^{r_2} + k(\gamma_0)^{r_3} \big)^{1/mr_4}, \tag{37}$$

where τ_0 is a measure of the hardening modulus in simple shear. f is a function of the effective plastic strain γ^P, and g is the measure of the effective plastic strain gradient η of any order, while k denotes a function of the initial yield strain in shear γ_0. The power n relates to the order of the gradient used to represent η. If the first order gradient is used then $n = 1$. r_1, r_2, r_3 and r_4 are assumed as phenomenological material constants (or interaction coefficients). This expression in (37) actually corresponds to an extending form of the hardening law proposed by Voyiadjis and Abu Al-Rub [2005], where a function for the initial yield stress is expressed in terms of the initial yield shear strain, which enters the bracket as a function k. These authors have demonstrated this hardening law is not phenomenological but physically based and is derived from a set of dislocation mechanics-based considerations. Variable internal length scales instead of constant ones have also been used in the same reference in developing a strain gradient plasticity model. However, most of the current strain gradient theories still retain the method of using constant length scales, due to the simplicity and convenience that it affords.

Consequently, if we choose

$$r_1 = r_2 = r_3 = 1, \quad r_4 = 1/m, \quad f(\gamma^P) = c_0(\gamma^P)^{1/m},$$
$$g(\ell^n \eta_n) = c_1 \ell \eta + c_2 \ell^2 \eta_2, \quad \text{with } n = 2,$$

where c_0 is a constant but c_1 and c_2 may depend on γ^P, the gradient terms are expressed as

$$c_1 \ell \eta = c_1 \ell \big| \nabla \gamma^P \big| = c_1 \ell \sqrt{\nabla \gamma^P \cdot \nabla \gamma^P} \quad \text{and} \quad c_2 \ell^2 \eta_2 = c_2 \ell^2 \nabla^2 \gamma^P,$$

where ∇ and ∇^2 are the forward gradient and Laplacian operators respectively. As a result, the yield function of (36) can be rewritten as

$$F = \tau - \tau_0 \big(h(\gamma_0) + c_0(\gamma^P)^{1/m} + c_1 \ell \big| \nabla \gamma^P \big| + c_2 \ell^2 \nabla^2 \gamma^P \big) = 0. \tag{38}$$

It is easy to prove that a yield function presenting a form of (38) may cover a family of Aifantis and his coworkers' gradient plasticity models, by appropriately simplifying the coefficients c_0, c_1 and c_2 [Aifantis 1984; 1987; Zbib and Aifantis 1988b; 1988a; Mühlhaus and Aifantis 1991 and references therein].

Otherwise, if we adopt the parameters

$$r_1 = r_2 = r_3 = 2, \quad n = 1, \quad r_4 = 2/(3m), \quad f(\gamma^P) = c_0(\gamma^P)^{1/m},$$
$$g(\ell^n \eta_n) = c_1 \ell \alpha, \quad h(\gamma_0) = \gamma_0,$$

where $\alpha = \sqrt{2\chi_{ij}^P \chi_{ij}^P} = \sqrt{2}\eta$, and c_0 and c_1 are constants, we obtained the yield surface

$$F = \tau - \tau_0 \big((\gamma_0)^2 + c_1(\gamma^P)^2 + c_1(\ell\alpha)^2 \big)^{3/2} = 0. \tag{39}$$

If the differential of strain gradients over γ^p can be neglected, the consistency condition of Equation (39) results in the instantaneous hardening rate:

$$h = \frac{\dot{\tau}}{\dot{\gamma}^p} = 3\tau_0 c_0 \gamma^p \left((\gamma_0)^2 + c_0 (\gamma^p)^2 + c_1 (\ell\alpha)^2\right)^{1/2}. \tag{40}$$

Or alternatively

$$h = \frac{\dot{\tau}}{\dot{\gamma}^p} = G\xi \left(1 + \frac{c_1 (\ell\alpha/\gamma_0)^2}{1 + c_0 (\gamma^p/\gamma_0)^2}\right)^{1/2}, \tag{41}$$

where $G = \tau_0/\gamma_0$ is the elastic shear modulus, and

$$\xi(\gamma^p) = 3c_0 \gamma^p \sqrt{1 + c_0 (\gamma^p/\gamma_0)^2}.$$

The hardening rate presented in Equation (41) shares a similarity with that proposed in [Acharya and Bassani 1996; 2001, (13)]. Given the function $\xi(\gamma^p)$ in (41) is further generalized such that the power-law based hardening rate in the latter reference may be included as a special case, in conjunction with the elastic relations in (34b), the J_2-flow theory of strain gradient plasticity formulated in both references may be immediately recovered. Note that for crystal metals undergoing incompatible lattice distortion, as stated in [Bassani 2001], the physical implications of the hardening rate present in (41) may be interpreted as follows: the term in the numerator of the square bracket may be regarded to reflect the increase in hardening due to incompatibility induced by Geometrically Necessary Dislocations (GNDs), while the term in the denominator gives rise to a greater effect of incompatibility in the early stages of deformation and approximately account for a significant initial density of GNDs.

A common difference between the gradient plasticity theories pioneered by Aifantis and coworkers and other gradient theories such as those based on micromorphic, micropolar or second gradient continua is that the description of the kinematics of the deformation at a given material point by the latter theories is generally enriched by the introduction of the microdeformations along with a work-conjugate thermodynamic quantity playing the role of a higher-order stress tensor. The introduction of these extra terms into the equilibrium equations requires additional boundary conditions to be specified in the solution of practical boundary value problems. While in the models based on gradient plasticity of Aifantis and coworkers, the yield function, the flow rule and/or the dilatancy condition are assumed to depend on the Laplacian of some suitable scalar measure of the accumulated plastic strain or a set of internal variables. No higher-order stress tensors are introduced into the constitutive description. Comparisons of these theories have been addressed in [Chambon et al. 2004].

4. Derivations of constitutive models for isotropic pressure-sensitive gradient-dependent geomaterials

4.1. *Yield functions in the dissipative stress space.* As is well known, geomaterials are generally sensitive to hydrostatic pressure. The failure modes of these materials are quite different for small stress level and high mean compressive stresses. Under high hydrostatic pressure, they can yield and flow like metal in such a way that if gradient effects are considered, the aforementioned generalized J_2-flow theories are capable of addressing them. However, in the low and intermediate compressive stress stage, the failure criteria of geomaterials are sensitive to hydrostatic states of stress. The yield criteria described

above are no longer adequate to address the plastic deformation process in this case, and thus pressure-dependent failure models are required. Typical pressure-dependent conventional plasticity models are the Mohr–Coulomb, Drucker–Prager and the critical-state Cam–Clay failure criteria. In considering pressure-dependence, it is appropriate to let pressure-terms (typically the first invariants of stresses) enter the dissipation function. As for gradient-dependent materials, here we only consider the first invariants of Cauchy stress, neglecting the first invariant of higher-order stresses. This assumption is generally acceptable and convenient for most geomaterials since, in most cases, the hydrostatic part for the Cauchy stress is easily physically interpreted while that of the higher-order stress is not. In gradient theories, the higher-order stress is frequently related with higher-order boundary traction. However, except in such special cases as pure bending where only momentum force boundary conditions are imposed (in this case the first invariant of couple stress is zero), physical meanings of other higher-order boundary tractions are yet to be comprehended. Thus in the following formulations, only the first invariant of Cauchy stress is considered whenever the explicit dependence of dissipation function on the true stresses is addressed. We note that in this subsection, the influence of intermediate stress on the failure behavior or, alternatively, on the Lode-dependence, is not considered, but will be addressed in next subsection.

In view of these assumptions, we consider the following general form for the pressure-dependent dissipation function rate for gradient materials:

$$\hat{\Phi} = (B_0 + B_1 I_1 + B_2 I_1^2)^{1/2} \big(A_1(\dot{\varepsilon}_{ij}^p \dot{\varepsilon}_{ij}^p) + A_2(l^2 \dot{\eta}_{ijk}^p \dot{\eta}_{ijk}^p)\big)^{1/2}, \tag{42}$$

where $I_1 = \sigma_{kk}$. A_1, A_2, B_0, B_1 and B_2 are constants, where

$$A_1 \geq 0, \quad A_2 \geq 0, \quad B_0 \geq 0, \quad B_1 \leq 0, \quad B_2 \geq 0, \quad |B_1| \geq 2\sqrt{B_0 B_2}.$$

It should also be noted that if $A_2 = 0$ in (42), it is equivalent to the expression proposed by [Ziegler and Wehrli 1987, Equation 7.11] for conventional pressure-sensitive materials without gradient effects. The dissipative stresses may thus be obtained from (42) by

$$\sigma_{ij}^d = \frac{\partial \hat{\Phi}}{\partial \dot{\varepsilon}_{ij}^p} = \frac{A_1(B_0 + B_1 I_1 + B_2 I_1^2)}{\hat{\Phi}} \dot{\varepsilon}_{ij}^p,$$

$$\tau_{ijk}^d = \frac{\partial \hat{\Phi}}{\partial \dot{\eta}_{ijk}^p} = \frac{A_2(B_0 + B_1 I_1 + B_2 I_1^2)l^2}{\hat{\Phi}} \dot{\eta}_{ijk}^p. \tag{43}$$

Consequently, the following yield function in the dissipative stress space is obtained

$$\tilde{F} = \frac{(\sigma_{ij}^d \sigma_{ij}^d)}{A_1} + \frac{(l^{-2}\tau_{ijk}^d \tau_{ijk}^d)}{A_2} - (B_0 + B_1 I_1 + B_2 I_1^2) = 0. \tag{44}$$

For convenience, we assume further that $A_1 = A_2 = 2/3$, and we define the following generalized J_2 and principal invariants:

$$\tilde{J}_2 = \tfrac{3}{2}(\sigma_{ij}^d \sigma_{ij}^d + l^{-2}\tau_{ijk}^d \tau_{ijk}^d), \qquad t_{II} = \sqrt{\tfrac{3}{2}\big((\sigma_{II}^d)^2 + l^{-2}\tau_{II}^2\big)},$$

$$t_I = \sqrt{\tfrac{3}{2}\big((\sigma_I^d)^2 + l^{-2}\tau_I^2\big)}, \qquad t_{III} = \sqrt{\tfrac{3}{2}\big((\sigma_{III}^d)^2 + l^{-2}\tau_{III}^2\big)}.$$

The following relations hold for the most general case.

$$I_1 \leq \frac{-B_1 - \sqrt{B_1^2 - 4B_0 B_2}}{2B_2}. \tag{45}$$

In the compound dissipative stress, Equation (44) describes a hyperboloid shape, whose axis is

$$t_I = t_{II} = t_{III}, \qquad I_1 = \frac{-\sqrt{B_1^2 - 4B_0 B_2} - B_1}{2B_2}. \tag{46}$$

The intersection of this cone with the π-plane defined by t_I, t_{II} and t_{III} is a circle with radius

$$\sqrt{B_0 + B_1 I_1 + B_2 I_1^2},$$

as shown in Figure 2, left. The longitudinal section of this hyperboloid cone, defined by $I_1/\sqrt{3}$ and $\sqrt{2\tilde{J}_2}$, by analogy to the traditional meridian plane for conventional models, is illustrated in Figure 2, right. As a special case, if

$$B_1 = -2\sqrt{B_0 B_2},$$

the yield surface (44) may be further rewritten as

$$\tilde{F} = \frac{3(\sigma_{ij}^d \sigma_{ij}^d)}{2} + \frac{3(l^{-2} \tau_{ijk}^d \tau_{ijk}^d)}{2} - \left(\sqrt{B_0} - \sqrt{B_2} I_1\right)^2 = 0, \tag{47}$$

with $I_1 \leq \sqrt{B_0/B_2}$. Equation (47) is actually a generalized Drucker–Prager criterion for the pressure-sensitive gradient-dependent materials. It denotes a conoid-shaped cone in the compound dissipative stress space, with a circle of radius

$$\sqrt{B_0} - \sqrt{B_1} I_1$$

in the intersected π-plane defined by t_I, t_{II} and t_{III}. The section in the meridian plane defined by $I_1/\sqrt{3}$ and $\sqrt{2\tilde{J}_2}$ is two lines (Figure 2, right), which are intersected with the curves defined by (46) at $\tilde{J}_2 = B_0$.

Based on the generalized Drucker–Prager criterion in Equation (47), if it is further assumed that $B_2 = 0$, the generalized von Mises case as summarized in Section 3 will be recovered. This is a natural consequence, since in conventional plasticity theories, the Drucker–Prager criterion is an extension of von Mises generalized to the pressure-sensitive case. Based on (47), further simplifications may be made by considering that $B_0 = 0$, so the yield surface becomes

$$\tilde{F} = \sqrt{\tilde{J}_2} \pm \sqrt{B_2} I_1 = 0. \tag{48}$$

As a result, a generalized form of the cohesionless Coulomb criterion that cannot sustain tension is obtained, which has a similar cone in the π-plane defined by t_I, t_{II} and t_{III} as the generalized Drucker–Prager criterion and also two lines in the meridian plane, except that the corner of the two lines is located at O' for (48) (Figure 2). Here $\sqrt{B_2}$ acts as a frictional coefficient and may be regarded as $\sqrt{B_2} = \tan\phi_0$, where ϕ_0 is the frictional angle of the material.
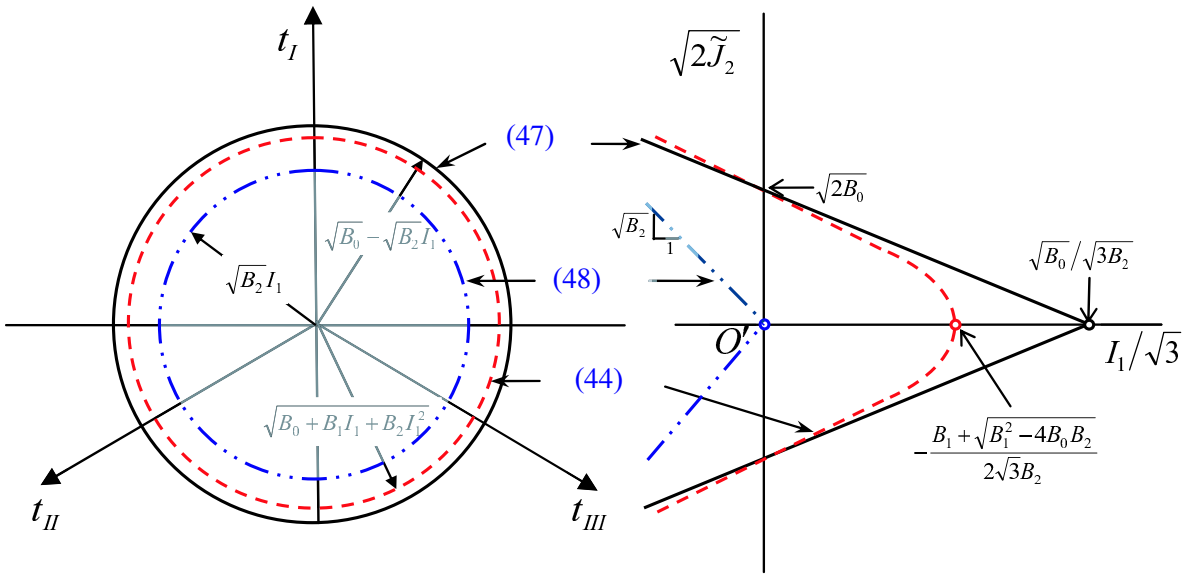
Figure 2. Yield loci for the generalized pressure-sensitive gradient materials defined by
(47) (solid line), (48) (double dashed line) and (44) (dashed line) respectively. Left: view
in the π-plane defined by t_I, t_{II} and t_{III}. Right: view in the meridian plane.

4.2. *Lode-angle-dependence of isotropic pressure-sensitive gradient materials.* Experimental investigations indicate that the third stress invariant of Cauchy stress affects significantly the behavior of pressure sensitive materials, such that the mechanical behavior of geomaterials may be better described by a Lode-angle-dependence relation. In conventional soil mechanics, this is attained by adding a function of the Lode angle into the yield functions. To achieve this aim for pressure-sensitive gradient materials, we will demonstrate in this section that Lode-angle-dependence may be easily manipulated by including the Lode angle in the expression for the rate of dissipation function. Bardet [1990] has tried to generalize the lode dependence for pressure-sensitive materials without gradient effects. In conventional soil mechanics, the Lode angle ϑ is defined by

$$\vartheta = -\frac{1}{3}\arcsin\left(\frac{3\sqrt{3}}{2}\frac{s^3}{J^3}\right), \tag{49}$$

where $s = \left(s_{ij}s_{jk}s_{ki}/3\right)^{1/3}$, $J = \left(s_{ij}s_{ij}/2\right)^{1/2}$. ϑ varies between $-\pi/6$ and $\pi/6$. In line with the experimental origin of Lode angle, including the influence of the deviatoric parts of higher-order stresses (e.g., incorporating the second invariant of higher-order stresses) into the definition of ϑ will be somewhat misleading. Thus even in the presence of strain gradients, the definition for ϑ will not be modified, and remains in form of (49). Accordingly, we use the general function of ϑ proposed by Eekelen [1980]

$$g(\vartheta) = a(1 + b\sin 3\vartheta)^{-c},$$

where a, b and c are constants. By appropriately selecting values for a, b and c, a large variety of failure criteria such as von Mises, Tresca, Mohr–Coulomb, Matsuoka and Nakai [1974] and Lade and Duncan [1975] may be approximated.

Starting from (42), it is assumed the rate of dissipation function is

$$\hat{\Phi} = (B_0 + B_1 I_1 + B_2 I_1^2)^{1/2} g(\vartheta) \big(A_1 (\dot{\varepsilon}_{ij}^p \dot{\varepsilon}_{ij}^p) + A_2 (l^2 \dot{\eta}_{ijk}^p \dot{\eta}_{ijk}^p) \big)^{1/2}. \tag{50}$$

Repeating the derivation in last subsection one readily obtains various Lode-dependence criteria. As an illustrative example, when $B_1 = -2\sqrt{B_0 B_2}$, a Lode-dependence generalized Drucker–Prager yield surface in the compound dissipative stress space may be obtained, yielding

$$\tilde{F} = \sqrt{\tilde{J}_2} - \big(\sqrt{B_0} - \sqrt{B_2} I_1 \big) g(\vartheta) = 0. \tag{51}$$

In the π-plane defined by t_I, t_{II} and t_{III}, the deviatoric curve for Equation (51) will depend on the selection of $g(\vartheta)$. For example, when

$$a = \sin \phi_0 \quad \text{(for } \phi_0 \text{ the frictional angle)},$$
$$b = 4a/3\sqrt{3},$$
$$c = 0.25,$$

a generalized deviatoric curve like that of [Lade and Duncan 1975] is obtained. If, instead,

$$a = \sin \phi_0,$$
$$b = 2a(3 - a^2)/3\sqrt{3},$$
$$c = 0.25,$$

for Equation (45), we get a generalized deviatoric curve [Matsuoka and Nakai 1974]. When $g(\vartheta) = 1$, the generalized compression Drucker–Prager curve is obtained. These curves are compared in Figure 3.
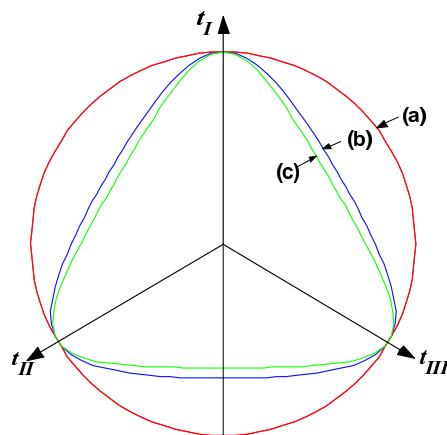

Figure 3. Deviatoric curves for various generalized criteria for pressure-sensitive gradient geomaterials in the π-plane defined by t_I, t_{II} and t_{III}. (a) Generalized compression Drucker–Prager criterion. (b) Generalized Lade–Duncan criterion. (c) Generalized Matsuoka and Nakai criterion.

4.3. *Example of a constitutive relation derivation for a particular case.* To derive the constitutive relations, yield function, and flow rule in the true stress space for pressure-sensitive gradient geomaterials, the free energy function must be specified. Without loss of generality, it is assumed here that the free-energy function may be decoupled into an elastic part and a plastic part as in Equation (9). In particular, we assume for the two parts the expressions

$$\Psi^e = \frac{1}{2}D^e_{ijkl}\varepsilon^e_{ij}\varepsilon^e_{kl} + \frac{1}{2}N^e_{ijklmn}l^{-2}\eta^e_{ijm}\eta^e_{lkn},$$

$$\Psi^p = \frac{1}{2}D^p_{ijkl}\varepsilon^p_{ij}\varepsilon^p_{kl} + \frac{1}{2}N^p_{ijklmn}l^{-2}\eta^p_{ijm}\eta^p_{lkn},$$

where

$$D^e_{ijkl} \quad \text{and} \quad D^p_{ijkl}$$

are elastic and plasticity stiffness, while

$$N^e_{ijklmn} \quad \text{and} \quad N^p_{ijklmn}$$

are higher-order elastic and plasticity stiffness. The rate of dissipation function uses the special case of (50) when $B_1 = -2\sqrt{B_0 B_2}$ so that the Lode-dependent generalized Drucker–Prager surface in the dissipative stress space in (51) may be attained. In view of (11), it is readily found that the elastic behavior of the material is governed by

$$\sigma_{ij} = \frac{\partial \Psi^e}{\partial \varepsilon^e_{ij}} = D^e_{ijkl}\varepsilon^e_{kl}, \qquad \tau_{ijk} = \frac{\partial \Psi^e}{\partial \eta^e_{ijk}} = N^e_{ijmlkn}l^{-2}\eta^e_{lmn},$$

while the back stresses defined in (13) are

$$\rho_{ij} = \frac{\partial \Psi^p}{\partial \varepsilon^p_{ij}} = D^p_{ijkl}\varepsilon^p_{kl}, \qquad \pi_{ijk} = \frac{\partial \Psi^p}{\partial \eta^p_{ijk}} = N^p_{ijmlkn}l^{-2}\eta^p_{lkn}.$$

In connection with Equations (12) and (51), the yield function in the true compound stress space of σ_{ij} and τ_{ijk} is

$$F = \sqrt{\bar{J}_2} - \left(\sqrt{B_0} - \sqrt{B_2}I_1\right)g(\vartheta) = 0,$$

where

$$\bar{J}_2 = \frac{3}{2}\left((\sigma_{ij} - \rho_{ij})(\sigma_{ij} - \rho_{ij}) + l^{-2}(\tau_{ijk} - \pi_{ijk})(\tau_{ijk} - \pi_{ijk})\right).$$

By analogy with Equation (43), normality relations hold in the dissipative stress space:

$$\sigma^d_{ij} = \frac{A_1\left(\sqrt{B_0} - \sqrt{B_2}I_1\right)^2\left(g(\vartheta)\right)^2}{\hat{\Phi}}\dot{\varepsilon}^p_{ij},$$

$$\tau^d_{ijk} = \frac{A_2\left(\sqrt{B_0} - \sqrt{B_2}I_1\right)^2\left(g(\vartheta)\right)^2 l^2}{\hat{\Phi}}\dot{\eta}^p_{ijk},$$

where $A_1 = A_2 = 2/3$. Therefore, in considering (13), the plastic strain and plastic strain gradient increments satisfy

$$\dot{\varepsilon}^p_{ij} = \hat{\Phi} \frac{\sigma_{ij} - \rho_{ij}}{A_1 \left(\sqrt{B_0} - \sqrt{B_2} I_1\right)^2 \left(g(\vartheta)\right)^2},$$

$$\dot{\eta}^p_{ijk} = \hat{\Phi} \frac{l^{-2}(\tau_{ijk} - \pi_{ijk})}{A_2 \left(\sqrt{B_0} - \sqrt{B_2} I_1\right)^2 \left(g(\vartheta)\right)^2}.$$

In combination with (21), the plastic multiplier may be determined, and thus the plastic incremental relations are obtained. While differentiating the yield surface equation, attention should be paid to the dependence of the Lode angle function $g(\vartheta)$ on the stresses. Thus far, the total constitutive relations for the Lode-angle-dependence pressure-sensitive gradient material are derived from the free-energy function and the rate of dissipation function.

5. Conclusions

Strain gradient plasticity constitutive models for gradient-dependent geomaterials can be constructed systematically from two thermodynamical potentials via the thermomechanical approach. In this way, the first and second thermodynamic laws are satisfied simultaneously, and yield surface in the dissipative stress space as well as the true stress space may be attained without difficulty. Both associated and nonassociated flow rule can be achieved. Appropriate selection of free-energy function and dissipation rate function makes it possible to recover a large range of strain gradient plasticity models proposed in the literature. In the framework of strain gradient theory, the flexibility of defining dissipation rate functions makes it easy to account for special features associated with geomaterials, such as pressure-sensitivity and Lode-angle dependency. Various failure criteria in geomechanics, such as those of von Mises, Tresca, Drucker–Prager, Coulomb, Lade–Duncan and Matsuoka–Nakai, have been generalized to include the strain gradient effects. Further extension of the obtained formulation to other criteria, such as critical state models, may also be obtained with relative ease. Further investigation will be directed towards incorporating the coupling between plasticity and damage for a range of geomaterials, like quasibrittle rocks, within the framework of strain gradient theory and thermomechanics. It is also noted that, regarding the numerical implementations of the gradient-dependent models developed from the current framework, some of the existing finite element methods proposed for strain gradient plasticity can be used and there is no need to develop new algorithms [Shu et al. 1999; Zervos et al. 2001; Abu Al-Rub and Voyiadjis 2005].

Acknowledgments

References

[Abu Al-Rub and Voyiadjis 2005] R. K. Abu Al-Rub and G. Z. Voyiadjis, "A direct finite element implementation of the gradient-dependent theory", *Int. J. Numer. Methods Eng.* **63**:4 (2005), 603–629.

[Acharya and Bassani 1996] A. Acharya and J. L. Bassani, "On non-local flow theories that preserve the classical structure incremental boundary value problems", pp. 3–9 in *IUTAM Symposium on Micromechanics of Plasticity and Damage of Multi-phase Materials* (Sèvres, 1995), edited by A. Pineau and A. Zaoui, Solid mechanics and its applications, Kluwer, Dordrecht, 1996.

[Aifantis 1984] E. C. Aifantis, "On the microstructural origin of certain inelastic models", *J. Eng. Mater. Technol. (Trans. ASME)* **106** (1984), 326–330.

[Aifantis 1987] E. C. Aifantis, "The physics of plastic deformation", *Int. J. Plast.* **3**:3 (1987), 211–247.

[Aifantis et al. 1999] E. C. Aifantis, F. Oka, A. Yashima, and T. Adachi, "Instability of gradient dependent elasto-viscoplasticity for clay", *Int. J. Numer. Anal. Methods Geomech.* **23**:10 (1999), 973–994.

[Bardet 1990] J. P. Bardet, "Lode dependences for isotropic pressure-sensitive elastoplastic materials", *J. Appl. Mech. (Trans. ASME)* **57** (1990), 498–506.

[Bassani 2001] J. L. Bassani, "Incompatibility and a simple gradient theory of plasticity", *J. Mech. Phys. Solids* **49**:9 (2001), 1983–1996.

[Benzerga et al. 2003] A. A. Benzerga, A. Needleman, Y. Brechet, and E. Van der Giessen, *Discrete dislocation prediction of single crystal hardening: tension and bending*, Kluwer, Dordrecht, 2003.

[Chambon et al. 1996] R. Chambon, D. Caillerie, and N. El Hassan, "Etude de la localisation unidimensionelle á l'aide d'un modéle de second gradient", *C. R. Acad. Sci.* **323**:IIb (1996), 231–238.

[Chambon et al. 1998] R. Chambon, D. Caillerie, and N. El Hassan, "One-dimensional localization studied with a second grade model", *Eur. J. Mech. A:Solids* **17**:4 (1998), 637–656.

[Chambon et al. 2001] R. Chambon, D. Caillerie, and T. Matsuchima, "Plastic continuum with microstructure, local second gradient theories for geomaterials: localization studies", *Int. J. Solids Struct.* **38**:46-47 (2001), 8503–8527.

[Chambon et al. 2004] R. Chambon, D. Caillerie, and C. Tamagnini, "A strain space gradient plasticity theory for finite strain", *Comput. Methods Appl. Mech. Eng.* **193**:27-29 (2004), 2797–2826.

[Collins 2005] I. F. Collins, "Elastic/plastic models for soils and sands", *Int. J. Mech. Sci.* **47**:4-5 (2005), 493–508.

[Collins and Houlsby 1997] I. F. Collins and G. T. Houlsby, "Application of thermomechanical principles to the modelling of geotechnical materials", *P. Roy. Soc. Lond. A Mat.* **453**:1964 (1997), 1975–2001.

[Collins and Kelly 2002] I. F. Collins and P. A. Kelly, "A thermomechanical analysis of a family of soil models", *Géotechnique* **52**:7 (2002), 507–518.

[Cosserat and Cosserat 1909] E. Cosserat and F. Cosserat, *Théorie des corps deformables*, Hermann, Paris, 1909.

[Coussy 1995] O. Coussy, *Mechanics of porous continua*, John Wiley, New York, 1995.

[Drucker 1951] D. C. Drucker, "A more fundamental approach to plastic stress-strain relations", pp. 487–491 in *Proceedings of the First U.S. National Congress of Applied Mechanics* (Chicago), ASME, New York, 1951.

[Eekelen 1980] H. A. M. Eekelen, "Isotropic yield surfaces in three dimensions for use in soil mechanics", *Int. J. Numer. Anal. Methods Geomech.* **4**:1 (1980), 89–101.

[Eringen and Kafadar 1976] A. C. Eringen and C. B. Kafadar, "Polar field theories", pp. 205–267 in *Continuum physics*, vol. III/3, edited by A. Eringen, Springer, Berlin, 1976.

[Fleck and Hutchinson 1993] N. A. Fleck and J. W. Hutchinson, "A phenomenological theory for strain gradient effects in plasticity", *J. Mech. Phys. Solids* **41**:12 (1993), 1825–1857.

[Fleck and Hutchinson 1997] N. A. Fleck and J. W. Hutchinson, "Strain gradient plasticity", pp. 295–361 in *Advances in applied mechanics*, vol. 33, edited by J. W. Hutchinson and T. Y. Wu, Academic Press, New York, 1997.

[Fleck and Hutchinson 2001] N. A. Fleck and J. W. Hutchinson, "A reformulation of strain gradient plasticity", *J. Mech. Phys. Solids* **49**:10 (2001), 2245–2271.

[Fleck et al. 1994] N. A. Fleck, G. M. Muller, M. F. Ashby, and J. W. Hutchinson, "Strain gradient plasticity: theory and experiment", *Acta Metall. Mater.* **42**:2 (1994), 475–487.

[Fremond and Nedjar 1996] M. Fremond and B. Nedjar, "Damage, gradient of damage and principle of virtual power", *Int. J. Solids Struct.* **33**:8 (1996), 1083–1103.

[Germain 1973] P. Germain, "The method of virtual power in continuum mechanics. Part 2: microstructure", *SIAM J. Appl. Math.* **25**:3 (1973), 556–575.

[Hill 1966] R. Hill, "Generalized constitutive relations for incremental deformation of metal crystals by multislip", *J. Mech. Phys. Solids* **14**:2 (1966), 95–102.

[Houlsby 1981] G. T. Houlsby, *A study of plasticity theories and their applicability to soils*, Ph.D. thesis, Cambridge University, 1981.

[Houlsby 1982] G. T. Houlsby, "A derivation of the small-strain incremental theory of plasticity from thermo-mechanics", pp. 109–118 in *Deformation and failure of granular materials* (Delft, 1982), edited by P. A. Vermeer and H. J. Luger, Balkema, Rotterdam, 1982.

[Houlsby and Puzrin 2000] G. T. Houlsby and A. M. Puzrin, "A thermomechanical framework for constitutive models for rate-independent dissipative materials", *Int. J. Plast.* **16**:9 (2000), 1017–1047.

[Koiter 1960] W. T. Koiter, "General theorems for elastic-plastic solids", pp. 167–221 in *Progress in solid mechanics*, vol. I, edited by I. N. Sneddon and R. Hill, North-Holland, Amsterdam, 1960.

[Lade 1988] P. V. Lade, "Effects of voids and volume changes on the behaviour of frictional materials", *Int. J. Numer. Anal. Methods Geomech.* **12**:4 (1988), 351–370.

[Lade and Duncan 1975] P. V. Lade and J. M. Duncan, "Elastoplastic stress-strain theory for cohesionless soil", *J. Geotech. Geoenviron. Eng.* **101** (1975), 1037–1053.

[Lubliner 1972] J. Lubliner, "On the thermodynamic foundations of non-linear solid mechanics", *Int. J. Non-Linear Mech.* **7**:3 (1972), 237–254.

[Lubliner 1990] J. Lubliner, *Plasticity theory*, Macmillian, New York, 1990.

[Matsuoka and Nakai 1974] H. Matsuoka and T. Nakai, "Stress-deformation and strength characteristic of soil under three different principal stresses", *Proc. JSCE* **232** (1974), 59–70.

[Matsushima et al. 2000] T. Matsushima, R. Chambon, and D. Caillerie, "Second gradient models as a particular case of microstructured models: a large strain finite elements analysis", *C. R. Acad. Sci.* **328**:IIb (2000), 179–186.

[Matsushima et al. 2002] T. Matsushima, R. Chambon, and D. Caillerie, "Large strain finite element analysis of a local second gradient model: application to localization", *Int. J. Numer. Methods Eng.* **54**:4 (2002), 499–521.

[Maugin 1992] G. A. Maugin, *The thermomechanics of plasticity and fracture*, Cambridge University Press, Cambridge, 1992.

[Mindlin 1964] R. D. Mindlin, "Micro-structure in linear elasticity", *Arch. Ration. Mech. An.* **16**:1 (1964), 51–78.

[Mindlin 1965] R. D. Mindlin, "Second gradient of strain and surface tension in linear elasticity", *Int. J. Solids Struct.* **1**:4 (1965), 417–438.

[Mroz 1973] Z. Mroz, *Mathematical models of inelastic material behaviour*, University of Waterloo Press, 1973.

[Mühlhaus 1986] H. B. Mühlhaus, "Shear band analysis for granular materials within the framework of Cosserat theory", *Ing. Arch.* **56** (1986), 389–399.

[Mühlhaus and Aifantis 1991] H. B. Mühlhaus and E. C. Aifantis, "A variational principle for gradient plasticity", *Int. J. Solids Struct.* **28**:7 (1991), 845–857.

[Mühlhaus and Vardoulakis 1987] H. B. Mühlhaus and I. Vardoulakis, "The thickness of shear bands in granular materials", *Géotechnique* **37** (1987), 217–283.

[Nix and Gao 1998] W. D. Nix and H. Gao, "Indentation size effects in crystalline materials: a law for strain gradient plasticity", *J. Mech. Phys. Solids* **46**:3 (1998), 411–425.

[Oda and Iwashita 1999] M. Oda and K. Iwashita, *An introduction to granular materials*, Balkema, Rotterdam, 1999.

[Oka et al. 2002] F. Oka, Y. Higo, and S. Kimoto, "Effect of dilatancy on the strain localization of water-saturated elasto-viscoplastic soil", *Int. J. Solids Struct.* **39**:13-14 (2002), 3625–3647.

[Olsson 1999] W. A. Olsson, "Theoretical and experimental investigation of compaction bands in porous rock", *J. Geophys. Res.* **104**:B4 (1999), 7219–7228.

[Pijaudier-Cabot and Bazant 1987] G. Pijaudier-Cabot and Z. P. Bazant, "Nonlocal damage theory", *J. Eng. Mech.-ASCE* **113** (1987), 1512–1533.

[di Prisco et al. 2002] C. di Prisco, S. Imposimato, and E. C. Aifantis, "A visco-plastic constitutive model for granular soils modified according to non-local and gradient approaches", *Int. J. Numer. Anal. Methods Geomech.* **26**:2 (2002), 121–138.

[Puzrin and Houlsby 2001] A. M. Puzrin and G. T. Houlsby, "A thermomechanical framework for rate-independent dissipative materials with internal functions", *Int. J. Plast.* **17**:8 (2001), 1147–1165.

[Rajagopal and Srinivasa 1998a] K. R. Rajagopal and A. R. Srinivasa, "Mechanics of the inelastic behaviour of materials, I: Theoretical underpinnings", *Int. J. Plast.* **14**:10–11 (1998), 945–967.

[Rajagopal and Srinivasa 1998b] K. R. Rajagopal and A. R. Srinivasa, "Mechanics of the inelastic behaviour of materials, II: inelastic response", *Int. J. Plast.* **14**:10–11 (1998), 969–995.

[Rice 1971] J. R. Rice, "Inelastic constitutive relations for solids: an internal variable theory and its application to metal plasticity", *J. Mech. Phys. Solids* **19**:6 (1971), 433–455.

[Shu et al. 1999] J. Y. Shu, W. E. King, and N. A. Fleck, "Finite elements for materials with strain gradient effects", *Int. J. Numer. Methods Eng.* **44**:3 (1999), 373–391.

[Smyshlyaev and Fleck 1996] V. P. Smyshlyaev and N. A. Fleck, "The role of strain gradients in the grain size effect for polycrystals", *J. Mech. Phys. Solids* **44**:4 (1996), 465–495.

[Toupin 1962] R. A. Toupin, "Elastic materials with couple stresses", *Arch. Ration. Mech. An.* **11**:1 (1962), 385–414.

[Tsagrakis and Aifantis 2002] I. Tsagrakis and E. C. Aifantis, "Recent developments in gradient plasticity. Part I: formulation and size effects", *J. Eng. Mater. Technol. (Trans. ASME)* **124**:3 (2002), 352–357.

[Valanis 1996] K. C. Valanis, "A gradient theory of internal variables", *Acta Mech.* **116**:1-4 (1996), 1–14.

[Vardoulakis and Aifantis 1989] I. Vardoulakis and E. C. Aifantis, "Gradient dependent dilatancy and its implications in shear banding and liquefaction", *Ing. Arch.* **59**:3 (1989), 197–208.

[Vardoulakis and Aifantis 1991] I. Vardoulakis and E. C. Aifantis, "A gradient flow theory of plasticity for granular materials", *Acta Mech.* **87**:3-4 (1991), 197–217.

[Vardoulakis and Aifantis 1994] I. Vardoulakis and E. C. Aifantis, "On the role of microstructure in the behaviour of soils: effects of higher order gradients and internal inertia", *Mech. Mater.* **18**:2 (1994), 151–158.

[Vardoulakis and Graf 1985] I. Vardoulakis and B. Graf, "Calibration of constitutive models for granular materials using data from biaxial experiments", *Géotechnique* **35** (1985), 299–317.

[Vardoulakis and Sulem 1995] I. Vardoulakis and J. Sulem, *Bifurcation analysis in geomechanics*, Chapman & Hall, London, 1995.

[Voyiadjis and Abu Al-Rub 2005] G. Z. Voyiadjis and R. K. Abu Al-Rub, "Gradient plasticity theory with a variable length scale parameter", *Int. J. Solids Struct.* **42**:14 (2005), 3998–4029.

[Walsh and Tordesillas 2004] S. D. C. Walsh and A. Tordesillas, "A thermomechanical approach to the development of micropolar constitutive models of granular media", *Acta Mech.* **167**:3–4 (2004), 145–169.

[Zbib and Aifantis 1988a] H. M. Zbib and E. C. Aifantis, "On the localization and post-localization behavior of plastic deformation, I–III", *Res Mech.* **23** (1988), 261–305.

[Zbib and Aifantis 1988b] H. M. Zbib and E. C. Aifantis, "On the structure and width of shear bands", *Scripta Metall.* **22**:5 (1988), 703–708.

[Zervos et al. 2001] A. Zervos, P. Papanastasiou, and I. Vardoulakis, "A finite element displacement formulation for gradient elastoplasticity", *Int. J. Numer. Methods Eng.* **50**:6 (2001), 1369–1388.

[Zhao and Sheng 2006] J. D. Zhao and D. C. Sheng, "Strain gradient plasticity by internal-variable approach with normality structure", *Int. J. Solids Struct.* **43**:18-19 (2006), 5836–5850.

[Zhao et al. 2005] J. D. Zhao, D. C. Sheng, and W. Y. Zhou, "Shear banding analysis of geomaterials by strain gradient enhanced damage model", *Int. J. Solids Struct.* **42**:20 (2005), 5335–5355.

[Zhou et al. 2002] W. Y. Zhou, J. D. Zhao, Y. G. Liu, and Q. Yang, "Simulation of localization failure with strain-gradient-enhanced damage mechanics", *Int. J. Numer. Anal. Methods Geomech.* **26**:8 (2002), 793–813.

[Ziegler 1975] H. Ziegler, "Non-linearity in thermomechanics", *Int. J. Non-Linear Mech.* **10**:3-4 (1975), 145–154.

[Ziegler 1983] H. Ziegler, *An introduction to thermomechanics*, 2nd ed., North-Holland, Amsterdam, 1983.

[Ziegler and Wehrli 1987] H. Ziegler and C. Wehrli, "On a principle of maximal rate of entropy production", *J. Non-Equil. Thermodyn.* **12** (1987), 229–243.

JIDONG ZHAO: Jidong.zhao@newcastle.edu.au
School of Engineering, The University of Newcastle, Callaghan, NSW 2308, Australia

DAICHAO SHENG: Daichao.sheng@newcastle.edu.au
School of Engineering, The University of Newcastle, Callaghan, NSW 2308, Australia

IAN F. COLLINS: i.collins@auckland.ac.nz
Department of Engineering Science, School of Engineering, The University of Auckland, Auckland, New Zealand

THE EFFECT OF CONTACT CONDITIONS AND MATERIAL PROPERTIES ON ELASTIC-PLASTIC SPHERICAL CONTACT

VICTOR BRIZMER, YUVAL ZAIT, YURI KLIGERMAN AND IZHAK ETSION

The elastic-plastic contact between a sphere and rigid flat is analyzed under perfect slip and full stick conditions for a wide range of the sphere mechanical properties. The analysis provides comparison of the contact load, contact area and distribution of the contact pressure for these two contact conditions. It is found that the contact conditions and mechanical properties have little effect on the global contact parameters. However, the evolution of the plastic zone with increasing interference is substantially different for contacts under perfect slip or full stick conditions.

Nomenclature

a = contact area radius

a_c = critical contact area radius: $\sqrt{R\omega_c}$ in slip, $\sqrt{R\delta_c}$ in stick

A = contact area

A_c = critical contact area: $\pi R\omega_c$ in slip, $\pi R\delta_c$ in stick

A^* = $A/\pi R\omega_c$ in slip, $A/\pi R\delta_c$ in stick

C_ν = $p_{mc}/Y_0 = 1.234 + 1.256\nu$

E = Young's modulus of the sphere

E_T = tangent modulus of the sphere

L_c = critical load in stick

p = contact pressure

p_{av} = average contact pressure, P/A

p_{avc} = critical average pressure in slip, $(2/3)p_{mc}$

p_{mc} = critical maximum pressure in slip

P = load

P_c = critical load in slip

P^* = P/P_c in slip, P/L_c in stick

R = radius of the sphere

Y_0 = virgin yield strength of the sphere

z_0 = location of yielding inception

δ_c = critical interference in stick

ν = Poisson's ratio of the sphere

ω = interference

ω_c = critical interference in slip

ω^* = ω/ω_c in slip, ω/δ_c in stick

ω_t = transition interference, at which the plastic region first reaches the sphere surface

ω_t^* = ω_t/ω_c in slip, ω_t/δ_c in stick

ζ_0 = dimensionless yielding inception depth, z_0/a

ζ_0^* = yielding inception depth ratio (stick over slip).

1. Introduction

The elastic-plastic contact of a deformable sphere and a rigid flat is a primary problem in contact mechanics with important scientific and technological aspects. The subject stems from the classical work of Hertz in 1881, who derived an analytical solution for the frictionless (that is, perfect slip) elastic contact of two spheres [Johnson 1985]. The stress field associated with elastic spherical contact was calculated in detail by Huber in 1904 [Fischer-Cripps 2000]. These pioneering works were extended in the following years to expand the frictionless contact into the elastic plastic regime, and to include frictional contact.

Chang et al. [1987] developed a model (CEB model) for the elastic-plastic contact between rough surfaces, which was based on analyzing the contact between a single sphere and a rigid flat under perfect slip condition. According to this model the sphere remains elastic, and hence the theory of Hertz holds until a certain critical interference at which yielding inception in the sphere is reached. Above the critical interference volume conservation of the sphere summit is imposed while the contact pressure distribution is assumed uniform and equal to the maximum pressure at the yielding inception. This simplified model resulted in a discontinuity in the contact load and in the derivative of the contact area with respect to the interference at the transition from elastic to elastic-plastic regime. Several modifications of the original CEB model were suggested [Evseev et al. 1991; Chang 1997; Zhao et al. 2000] in order to smooth the transition from the elastic to plastic regime, However, all these modifications are purely mathematical manipulations without any physical basis. Mesarovic and Fleck [2000] analyzed the problem of a normal frictionless contact between two dissimilar spheres. They studied the effect of material strain-hardening and the geometrical and mechanical dissimilarity of the spheres on the contact parameters and on the regime of deformation. Kogut and Etsion [2002] used a finite element method to study numerically the evolution of the plastic zone in elastic-plastic contact between a sphere and a rigid flat under perfect slip condition. They provided convenient dimensionless expressions for the contact load, contact area and mean contact pressure, covering a wide range of interferences for a single value of the Poisson's ratio ($\nu = 0.3$). Similar results were then obtained by Quicksall et al. [2004] and Jackson and Green [2005]. Etsion et al. [2005] studied the process of loading-unloading of an elastic-plastic loaded sphere in contact with a rigid flat under perfect slip condition. They calculated the contact load, stresses and deformations in the sphere during both loading and unloading, for a wide range of interferences and several combinations of material properties.

The first analytical solution of the elastic spherical contact problem under full stick condition is by Goodman [1962]. Goodman found a simplified solution for the tangential stress distribution over the contact area of two dissimilar elastic spheres. The effect of these tangential stresses on the normal displacements was neglected, so that the pressure distribution over the contact area was assumed to be Hertzian. A more exact analysis of the elastic spherical contact under full stick contact condition was performed by Spence [1968] who solved simultaneously the dual integral equations for shear stresses and pressure distribution over the contact area and calculated the total compressive load. It follows from Spence's results that for small values of Poisson's ratio, the influence of the shear stresses on the contact load is appreciable. Spence [1975] extended his previous analysis to the case of partial stick using a certain value of friction coefficient. Hills and Sackfield [1987] presented a complete picture of the stress distribution in the elastic spherical contact under full and partial stick contact conditions, using the assumptions of [Goodman 1962]. Kosior et al. [1999] analyzed an elastic spherical contact under partial slip condition (with a finite Coulomb friction) using a domain decomposition method coupled with a boundary element method. They calculated stress distribution, contact radius and displacement as functions of the sphere mechanical properties. Their results were in good agreement with the analytical solution of Spence [1975]. Mesarovic and Fleck [1999] studied numerically the problem of spherical indentation of a rigid sphere into an elastic-plastic half-space under perfect slip and full stick conditions. They investigated the effect of contact conditions and material strain-hardening on the contact parameters and on the regime of deformation. Brizmer et al. [2006] studied the influence of the two different contact conditions (full stick vs. perfect slip) and material properties on the elasticity terminus of a contact between a smooth elastic sphere and rigid flat. The plastic yield inception of ductile materials and the failure inception of brittle materials were studied separately.

From the above introduction it can be seen that most of the existing literature on elastic-plastic spherical contact concerns perfect slip condition while full stick condition is mostly limited to elastic contact. Moreover, an accurate study of the effect of the sphere mechanical properties on the contact parameters, such as contact load and contact area, and on the evolution of the plastic zone with increasing interference, is still missing. Since a realistic elastic-plastic contact of a sphere and flat may be far from the ideal assumption of perfect slip, it seems appropriate to analyze the problem of spherical contact under stick condition for different material properties, and to compare the results with the corresponding solutions of perfect slip.

2. Theoretical background

Figure 1 presents a deformable sphere in contact with a rigid flat. The solid and dashed lines show the contours of the contacting bodies after and before the loading. The inner solid contour of the deformed sphere shown in Figure 1 corresponds to relatively low values of Poisson's ratio (that is, a compressible material), when the volume of the sphere diminishes under compressive loading. For high values of Poisson's ratio (that is, a nearly incompressible material) the volume of the sphere remains almost constant during the deformation. For conservation of volume, the points of the sphere will tend to move outwards rather than inwards in the radial direction (the outer solid contour in Figure 1). Two different types of the contact conditions are analyzed in the present work: perfect slip and full stick. The former case assumes no tangential stresses in the contact area. The latter case implies that corresponding
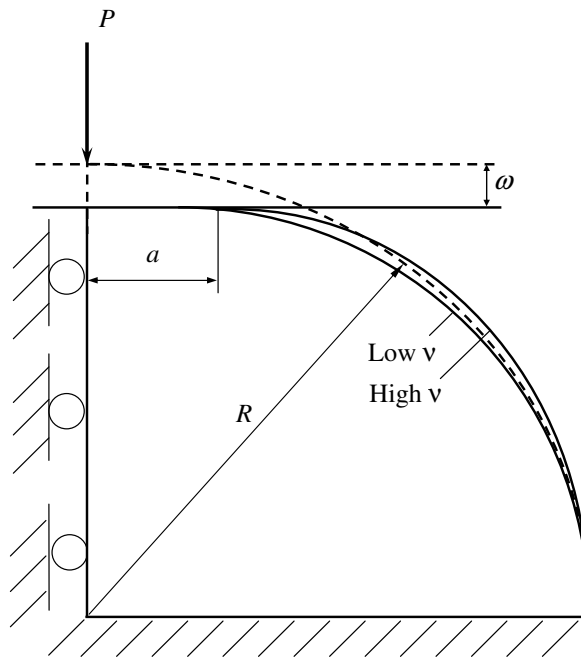
Figure 1. Model of a deformable sphere in contact with a rigid flat.

points of the sphere and the flat (which initially laid outside the contact area and were free to acquire a relative displacement) that are overtaken by the expanding contact zone, are prevented from further relative displacement [Johnson et al. 1973]. The perfect slip assumption, although not describing a general realistic contact condition, allows a relatively simple analytical solution in the elastic regime and can be considered as a limiting case. The full stick assumption, on the other hand, is more realistic when considering the junction formed in the interface of contacting bodies [Tabor 1959]. In this case the stress field and yielding criterion at the junction interface obey material constitutive laws. This approach is different from previous frictional contact solutions (for example, [Spence 1975]) where partial slip is obtained when a certain local Coulomb friction law is assumed.

The contact area, its radius a, and interference ω (see Equation (1)) correspond to a contact load P. The material of the sphere is assumed to be elastic-isotropic linear hardening [Etsion et al. 2005].

Since the problem is axisymmetric, it is sufficient to consider only half of the axisymmetric hemisphere section, as shown in Figure 1. The boundary conditions consist of constraints in the vertical and radial directions at the bottom of the hemisphere and in the radial direction at the axis of symmetry (see Equation (1)). The surface of the sphere is free elsewhere except for tractions imposed by the contacting rigid flat.

The critical interference ω_c or δ_c at yielding inception in perfect slip or in full stick, respectively, and their corresponding values of the critical loads, P_c and L_c, were given by Brizmer et al. [2006] in the

form:

$$\omega_c = \left(C_\nu \frac{\pi(1-\nu^2)}{2} \left(\frac{Y_0}{E} \right) \right)^2 R \tag{1}$$

$$\delta_c/\omega_c = 6.82\nu - 7.83(\nu^2 + 0.0586) \tag{2}$$

$$P_c = \frac{\pi^3 Y_0}{6} C_\nu^3 \left(R(1-\nu^2) \left(\frac{Y_0}{E} \right) \right)^2 \tag{3}$$

$$L_c/P_c = 8.88\nu - 10.13(\nu^2 + 0.089). \tag{4}$$

In Equations (1) and (3), $C_\nu = 1.234 + 1.256\,\nu$ is the maximum dimensionless contact pressure at yielding inception in slip, p_{mc}/Y_0. The parameters Y_0, E, and ν are the virgin yield stress, the Young modulus, and the Poisson's ratio of the sphere material, respectively. Equations (1) and (3) for the perfect slip were obtained analytically, using the Hertz solution [Johnson 1985] and applying the von Mises yield criterion. Equations (2) and (4), corresponding to the full stick condition, were derived numerically.

The critical value of the contact area, corresponding to yield inception for slip contact condition follows from Hertz solution [Johnson 1985]:

$$A_c = \pi \omega_c R \tag{5}$$

and for stick, as was found in [Brizmer et al. 2006]:

$$A_c = \pi \delta_c R. \tag{6}$$

The dimensionless yielding inception depth in perfect slip, ζ_0, and the ratio of yielding inception depth in stick over that in slip, ζ_0^*, are [Brizmer et al. 2006]:

$$\zeta_0 = z_0/a = 0.381 + \nu/3 \tag{7}$$

$$\zeta_0^* = 0, \nu \leq 0.26;$$
$$\zeta_0^* = 1.54(\nu - 0.26)^{0.294}, 0.26 < \nu \leq 0.5; \tag{8}$$

where z_0 is the yielding inception depth in perfect slip.

3. Finite elements model

The loading of an elastic-plastic contact is a complicated problem. To avoid oversimplifications, this contact problem was solved numerically, by a finite element method using the commercial package AN-SYS 8.0. The mesh consisted of 8,800 six-node triangular axisymmetric elements (Plane2) comprising a total of 29,277 nodes. The sphere was divided into three different mesh density zones (see Figure 2), where zones I and II were within $0.015R$ and $0.1R$, respectively, from the sphere summit, and zone III outside the $0.1R$ distance. Zone I had the finest mesh and the other zones had a gradually coarser mesh at increasing distance from the sphere summit. The sphere surface consisted of contact elements (Conta172) that matched the size of the elements in each zone. The rigid flat was modeled by a single nonflexible element (Targe169). The material of the sphere was assumed elastic linear isotropic hardening with a tangent modulus, E_T, that was selected as 2% of the Young modulus E, which is an upper limit of many
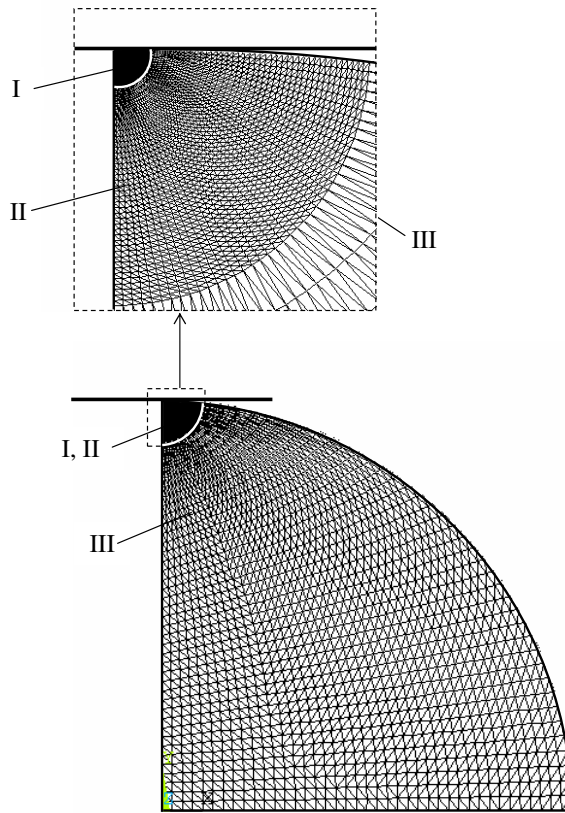
Figure 2. The finite elements model.

practical materials, see, for example, [Carmichael 1955]. This linear isotropic hardening significantly improves convergence compared to an elastic perfectly plastic material without causing much change (less than 2.5%) in the results. Hence, this hardening enables comparison with existing results for perfect slip condition, such as those in [Kogut and Etsion 2002], yet provides a better convergence of the numerical solution. The von Mises yielding criterion was used to detect local transition from elastic to plastic deformation, and Hooke's and the Prandtl–Reuss' constitutive laws governed the stress-strain relations in the elastic and plastic zones, respectively. A nonlinear finite deformation definition was used to allow large interferences (up to $\omega^* = 110$).

To verify the accuracy of the finite element model, results for purely elastic contact under perfect slip condition were tested against the Hertz solution. The correlations of the contact load and contact area were within 1% and 3%, respectively. Another check was done at high interferences in full stick and in perfect slip by increasing the mesh density and ensuring convergence of the results within a small pre-defined tolerance. Typical computation times on a 1.6 GHz PC were about 2–3 minutes for small, and 10–15 minutes for large interferences.

4. Results and discussion

The effects of the contact conditions (full stick vs. perfect slip), and of the material properties on the contact load and contact area were investigated in the present work over a wide range of dimensionless interferences from elastic to full plastic contact $0.25 \leq \omega^* \leq 110$ (here ω^* is ω/ω_c in slip and ω/δ_c in stick). Three different values of Poisson's ratio typical for ductile metals, 0.25, 0.35 and 0.45, were selected in the present analysis along with three values of the ratio E/Y_0: 500, 1000 and 2000. Each combination of the material properties was examined over the full range of interferences.

The results for the dimensionless contact load and contact area in slip as functions of ω^* in the elastic regime, $\omega^* < 1$, were found identical to the Hertz solution, as would be expected. The results of the contact area in stick were found to be identical to those in slip. From the Hertz solution $A = \pi R \omega$, hence, from the definition of the critical areas in slip and stick in Equations (5) and (6), respectively, it is clear that $A^* = A/A_c = \omega^*$ in both stick and slip regardless of the Poisson's ratio. The results of the contact load, P^*, in stick (P/L_c) were found very similar to these in slip (P/P_c) particularly at an increasing Poisson's ratio. The diminishing difference between the contact load P^* in stick and slip with an increasing Poisson's ratio is attributed to the increasing material incompressibility. This can be seen, at least for the elastic regime, from the Hertz solution for the radial displacement, u_r, of the sphere points at the contact area:

$$u_r(r, v) = f(r) \cdot (1 - 2v)(1 + v), \tag{9}$$

where $f(r)$ is a certain function of the radial coordinate, the radius of the contact area and the Young's modulus of the sphere. From Equation (9) it is clear that a higher Poisson's ratio results in lower relative displacements between the sphere and the flat in slip, making the slip condition closer to the zero relative displacements in stick condition.

In the elastic-plastic regime, $\omega^* > 1$, the numerical results for the dimensionless contact area A^* and contact load P^* as functions of ω^* were again very similar in stick and in slip. These results were best fitted and have the following forms:

$$A^* = \omega^* \left(1 + \exp \left(\frac{1}{1 - (\omega^*)^\alpha} \right) \right), \tag{10}$$

$$P^* = (\omega^*)^{3/2} \left(1 - \exp \left(\frac{1}{1 - (\omega^*)^\beta} \right) \right), \tag{11}$$

where α and β are linear functions of the Poisson's ratio:

$$\alpha = 0.25 + 0.125v; \qquad \beta = 0.174 + 0.08v.$$

The approximate expressions in Equations (10) and (11) for A^* and P^* describe the numerical FEM results with an average difference of 4% in A^* and 2% in P^*.

From Equation (10) it can be seen that for ω^* slightly larger than one (that is, just after the yielding inception), A^* approaches ω^*, in agreement with the elastic solution of Hertz. On the other hand, for large values of ω^*, A^* approaches $2\omega^*$, i.e. the approximate fully plastic solution of [Abbott and Firestone 1933]. Equation (11) gives $P^* \approx (\omega^*)^{3/2}$ for ω^* slightly larger than 1, which is in agreement with Hertz solution. These results compare favorably with the relevant dimensionless expressions presented
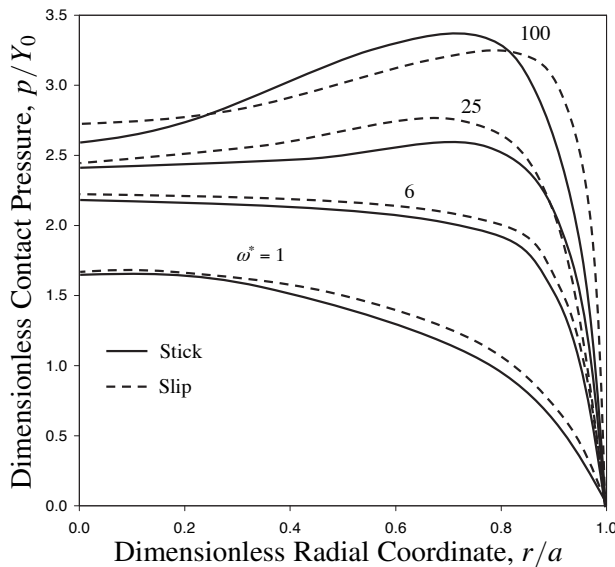
Figure 3. The effect of dimensionless interference on the dimensionless contact pressure distribution, p/Y_0, in slip and in stick, for a typical case of $\nu = 0.3$.

in [Kogut and Etsion 2002] (the KE model) and in [Etsion et al. 2005] for $\nu = 0.3$, and slip condition for elastic perfectly plastic and for elastic linear hardening materials, respectively.

Equations (10) and (11) suggest that normalizing the results for the contact area and contact load by their appropriate critical values at yielding inception for either slip or stick (see Equations (1) to (6)) as given by [Brizmer et al. 2006], provides, for ductile materials, a powerful universal solution for the elastic-plastic contact problem in both stick and slip where an analytical solution is impossible. It extends the classical Hertz solution into the elastic-plastic regime while providing a physical insight of the main dimensionless parameters affecting this contact problem in both slip and stick.

It can be seen from Equations (10) and (11) that the ratio E/Y_0 has no effect on the results for a given dimensionless interference ω^* in both stick and slip, while the Poisson's ratio has little effect on the dimensionless contact area and contact load. It is also evident that isotropic linear hardening has negligible effect on the results compared to elastic perfectly plastic behavior.

The effect of the contact conditions on the contact pressure distribution, p/Y_0, at the contact area for a typical case $\nu = 0.3$ is shown in Figure 3. The solid and dashed lines correspond to full stick and perfect slip, respectively. It can be seen that the two contact conditions yield very similar results for the full range of interferences with pressures in slip slightly higher than in stick, at least for $\omega^* \leq 25$. This seemingly counterintuitive result is due to the fact that the critical interference in stick δ_c is significantly less than ω_c in slip (e.g. $\delta_c/\omega_c \approx 0.88$ for $\nu = 0.3$, see Equation (2)). Hence, the pressure distributions in slip correspond to dimensional interference, ω, that is 12% higher than the dimensional interference in stick for a given value of ω^*. In other words, the pressure in slip would be somewhat lower than that in stick for the same dimensional interference. Recalling that the contact area depends on the interference (same contact area for same ω in both slip and stick), the results in Figure 3 are in agreement with the
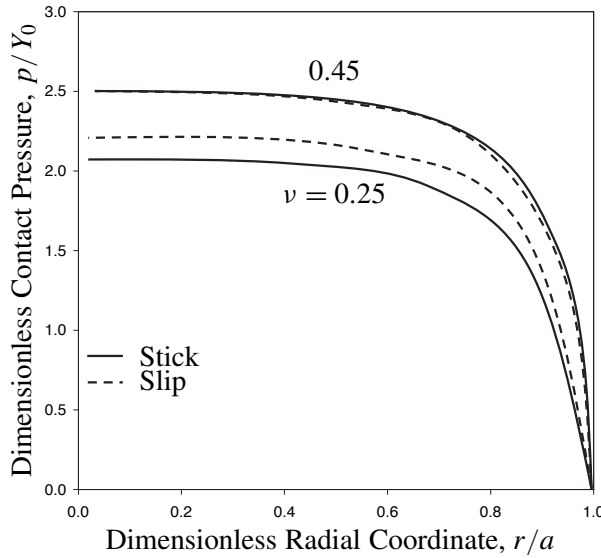
Figure 4. The effect of the Poisson's ratio on the dimensionless contact pressure distribution in slip and in stick for the case of $\omega^* = 6$.

observation (see, for example, [Johnson 1985, p. 123]) that "friction can increase the total load required to produce a contact of given size by at most 5% compared with Hertz".

The effect of the Poisson's ratio on contact pressure distribution in slip and in stick is shown in Figure 4 for the case of $\omega^* = 6$. It can be seen that a higher Poisson's ratio results in higher contact pressure for this dimensionless interference in both slip and stick. This is attributed to the fact that at a higher Poisson's ratio, the material is less compressible and so higher pressure is required to deform it. For the case of $\nu = 0.45$ where the critical interferences in stick and slip are almost identical, as in Equation (2) (ω values in stick and in slip are equal), the solid and dashed lines coincide, that is, the pressure distributions in slip and in stick are almost identical. On the other hand, for $\nu = 0.25$ where $\delta_c/\omega_c \approx 0.76$ (see Equation (2)) the dimensional interference, ω, in slip for a given value of ω^* is larger than that in stick causing the pressure distribution in slip to be slightly higher than that in stick.

The dimensionless average contact pressure, p_{av}/Y_0, vs. the dimensionless interference, ω^*, in slip has the form (see Equations (1), (3), and (5) and the definition of C_ν):

$$\left(\frac{p_{av}}{Y_0} \right)_{\text{slip}} = \frac{P}{AY_0} = \frac{P^*}{A^*} \cdot \frac{P_c}{\pi \omega_c R Y_0} = \frac{P^*}{A^*} \cdot \frac{p_{avc}}{Y_0}, \tag{12}$$

while in stick:

$$\left(\frac{p_{av}}{Y_0} \right)_{\text{stick}} = \frac{P^*}{A^*} \cdot \frac{L_c}{\pi \delta_c R Y_0} = \frac{P^*}{A^*} \cdot \frac{p_{avc}}{Y_0} \cdot \frac{(L_c/P_c)}{(\delta_c/\omega_c)}, \tag{12a}$$
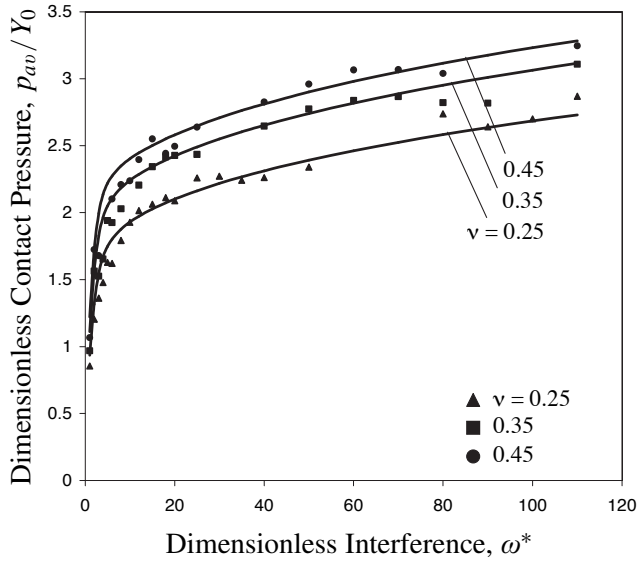
Figure 5. The dimensionless average contact pressure p_{av}/Y_0 vs. the dimensionless interference.

where the dimensionless parameters: p_{avc}/Y_0, L_c/P_c and δ_c/ω_c are known functions of ν (see Nomenclature and Equations (2) and (4)). From Equations (12) and (12a) we have:

$$\frac{(p_{av})_{\text{slip}}}{(p_{av})_{\text{stick}}} = \frac{\delta_c/\omega_c}{L_c/P_c}. \tag{13}$$

By using Equations (2) and (4) it can easily be shown that for $\nu = 0.25$, the ratio of the average contact pressures, slip over stick, in Equation (13) is 1.1, while for $\nu \geq 0.3$ it is very close to a unity. Hence, like the dimensionless contact area and contact load (Equations (10) and (11)), the average contact pressures in stick and in slip are very similar and can also be properly normalized by their critical values.

Substituting A^* and P^* from Equations (10) and (11), respectively, into Equation (12) or (12a) yields the dimensionless average pressure in stick or in slip as a function of ω^* and ν. The results are presented in Figure 5 along with the numerical FEM results showing a very good agreement with an average difference of 4%. It was found that Equations (12) or (12a) can also be well approximated by the simpler form:

$$\frac{p_{av}}{Y_0} = (\omega^*)^{1/2} \cdot \tanh\left(\frac{1}{2((\omega^*)^\gamma - 1)}\right) \cdot \frac{p_{avc}}{Y_0}, \tag{14}$$

where

$$\gamma = 0.2 + 0.06\nu,$$

without any loss of accuracy. It can be easily seen that for ω^* slightly larger than 1, $p_{av}/p_{avc} = (\omega^*)^{1/2}$, which is in agreement with the Hertz solution.

From Figure 5 it can be seen that the average contact pressure increases sharply up to an interference that is about $10\omega^*$. From there on the rate of increase of the average pressure diminishes, and at about $110\omega^*$ the average pressure reaches a value that varies from $2.73Y_0$ for $\nu = 0.25$ to $3.28Y_0$ for $\nu = 0.45$.
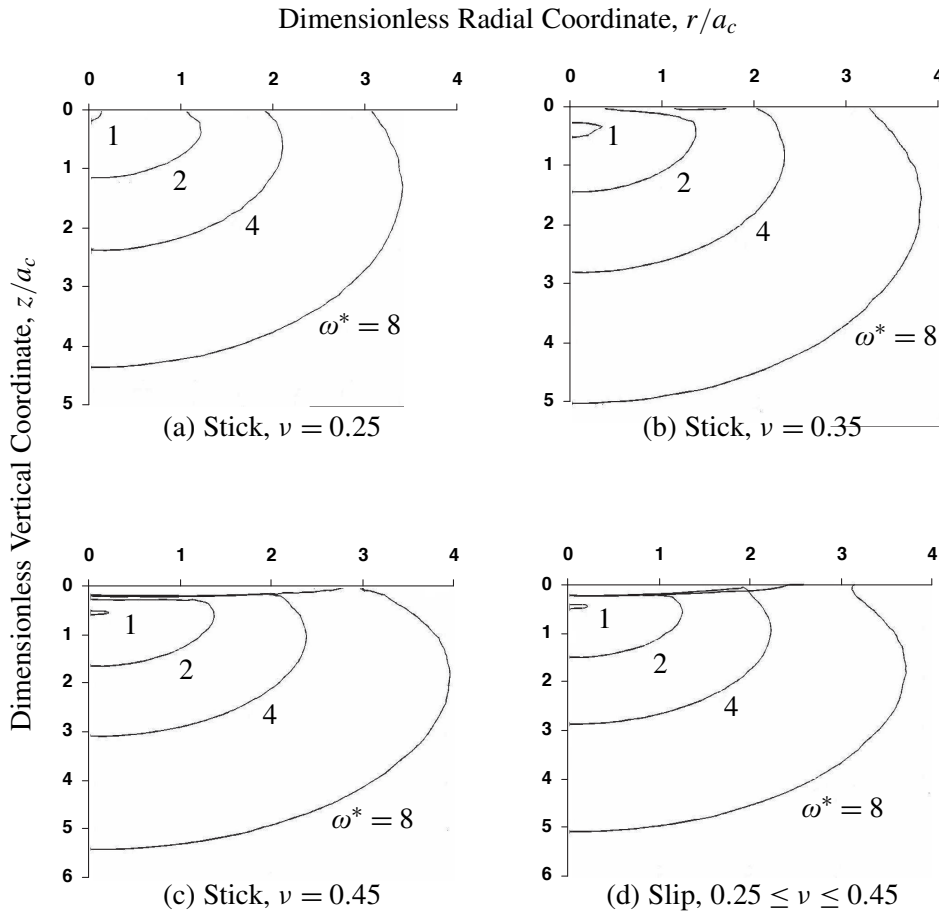
Dimensionless Radial Coordinate, r/a_c

Figure 6. The early evolution of the plastic region for small values of the dimensionless interference in stick at different values of the Poisson's ratio (a)–(c) in comparison with that in slip for $0.25 \le \nu \le 0.45$ (d).

The higher average contact pressure at larger Poisson's ratios is due to the reduced compressibility as shown in Figure 4. The results in Figure 5 are in good agreement (within 5%) with [Kogut and Etsion 2002] where a value of $2.8Y_0$ was found for $\nu = 0.3$. A similar effect of the Poisson's ratio on the hardness was found by Kogut and Komvopoulos [2004] for indentation and Jackson and Green [2005]. In these two references a reduction in p_{av}/Y_0 was observed for very large interferences corresponding to a/R values in access of 0.1. Our analysis is limited to $a/R < 0.05$.

The early evolution of the plastic region for small dimensionless interferences ($\omega^* \le 8$) within the sphere tip is presented in Figure 6. As we see, the Poisson's ratio has a large effect on the early evolution of the plastic region under full stick condition (see Figure 6a–c). On the other hand it was found that under slip condition the Poisson's ratio effect is negligible. Hence, for slip condition just one typical case for the range $0.25 \le \nu \le 0.45$ is presented in Figure 6d. The differences in the early evolution of the
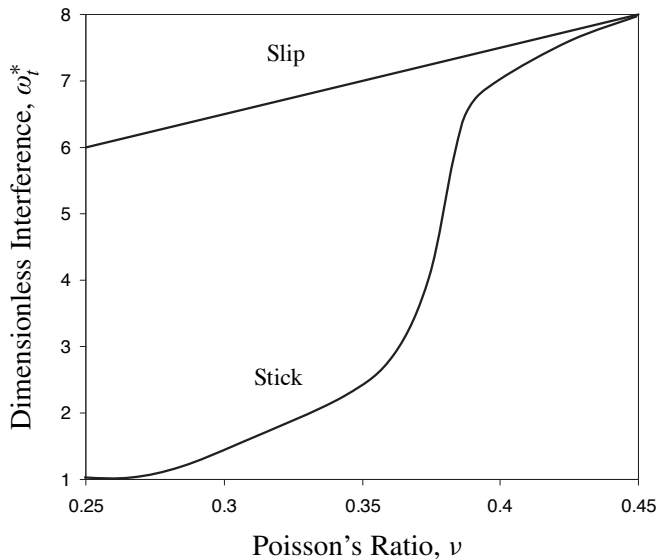
Figure 7. The dimensionless transition interference ω_t^*, for stick and for slip as a function of the Poisson's ratio.

plastic region between stick and slip are in good agreement with the results of [Brizmer et al. 2006] for the yielding inception depth that was found highly dependent on the Poisson's ratio in stick, but almost unaffected by it in slip (see Equations (7) and (8). Another interesting difference between stick and slip concerns the existence of an elastic core close to the contact interface. This core is completely missing in Figure 6a, it has an annular shape in Figure 6b, and is continuous over a central portion of the contact area in Figures 6c and 6d. Further numerical simulations in the range $0.3 \leq \nu \leq 0.4$ revealed that this elastic core (that is, the elastic zone at the very top of the sphere surrounded by the expending plastic region), which persists up to relatively high values of ω^* in slip [Kogut and Etsion 2002], is completely missing in stick when $\nu < 0.35$. As can further be seen from Figure 6c, d the early evolution of the plastic region in stick for a Poisson's ratio of 0.45 is similar to that in slip.

The transition interference, ω_t^*, at which the evolving plastic zone first reaches the surface of the sphere is also very different in stick and in slip. This can be seen from Figures 6a–d, and more clearly from Figure 7. From this figure we see that in slip, ω_t^* varies linearly from 6 to 8 over the entire range of the Poisson's ratio. In stick, on the other hand, when $\nu \leq 0.26$ the plastic zone starts at the contact interface (see Equation (8) hence, $\omega_t^* = 1$. As the Poisson's ratio increases the transition interference, ω_t^*, in stick increases too, first moderately until $\nu = 0.35$, then very rapidly until $\nu = 0.39$, and finally as ν approaches 0.45 the transition interference in stick approaches that in slip. The higher values of ω_t^* in slip compared to stick, as shown in Figure 7, are due to the deeper location of yielding inception under the former contact condition (see Equations (7) and (8)), which requires more deformation for the plastic zone to reach the sphere surface. Also, at Poisson's ratios below 0.35 in stick, the plastic region first reaches the sphere surface at the center of the contact area even when incepting slightly below it (see Figures 6a and b). In slip this event starts close to the circumference of the contact area (see also [Kogut and Etsion 2002]) at a much longer distance from the point of yield inception and thus requires
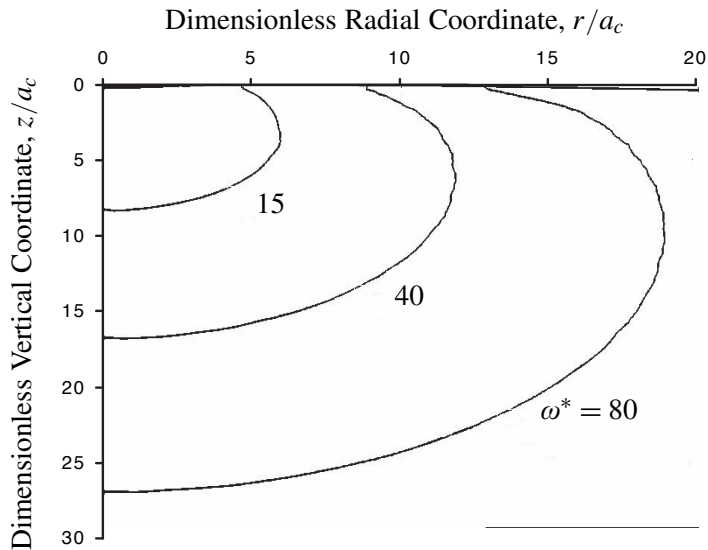
Figure 8. Typical evolution of the plastic zone for high values of the dimensionless interference for $0.25 \leq \nu \leq 0.45$, showing identical behavior in stick and slip.

larger interferences. The sharp increase of ω_t^* in stick observed above $\nu = 0.35$ can be explained by the appearance of the annular elastic core (see Figure 6b) that extends towards the center of the contact area with further increases of the Poisson's ratio (Figure 6c), blocking the upward advancement of the plastic zone front and forcing it radially out, similarly to the behavior in slip.

The evolution of the plastic zone for dimensionless interferences higher than 15 is shown in Figure 8. As can be seen the differences between stick and slip conditions disappear and the plastic zone becomes identical regardless of the Poisson's ratio. The reason for this identical behavior is that at higher interferences with increasing size of the plastic zone, the different historical location of the yielding inception and the development of the elastic core in stick and slip become insignificant.

5. Conclusion

The effects of contact condition (perfect slip or full stick) and material properties on elastic-plastic normally loaded spherical contact were investigated for a large range of interferences. The results for the dimensionless contact area, contact load and average contact pressure were found to be almost insensitive to the contact condition, independent of the ratios E/Y_0 and E_T/E and of the sphere radius, and slightly affected by the Poisson's ratio. The numerical results were approximated by simple empirical expressions based on realistic physical behavior. It was shown that normalizing the various parameters by their corresponding critical values at yielding inception provides a powerful general solution for the contact problem that is valid under both perfect slip and full stick conditions. Unlike the global contact parameters, the early evolution of the plastic zone in the contact region is very much affected by Poisson's ratio mainly under stick contact condition. Up to about $\nu = 0.4$ this evolution in stick is also very different

from the evolution in slip. At higher values of the Poisson's ratio the behavior in stick becomes very similar to that in slip. As the dimensionless interference increases and the plastic zone continues to expand, the differences between stick and slip diminish for the full range of Poisson's ratio.

References

[Abbott and Firestone 1933] E. J. Abbott and F. A. Firestone, "Specifying surface quantity – a method based on accurate measurement and comparison", *Mech. Eng.* **55** (1933), 569–572.

[Brizmer et al. 2006] V. Brizmer, Y. Kligerman, and I. Etsion, "The effect of contact conditions and material properties on the elasticity terminus of a spherical contact", *Int. J. Solids Struct.* **43**:18-19 (2006), 5736–5749.

[Carmichael 1955] C. Carmichael, *Kent's mechanical engineers' handbook: Design and production volume*, 12th ed., edited by C. Carmichael, Wiley, New York, 1955.

[Chang 1997] W. R. Chang, "An elastic-plastic contact model for a rough surface with an ion-plated soft metallic coating", *Wear* **212**:2 (1997), 229–237.

[Chang et al. 1987] W. R. Chang, I. Etsion, and D. B. Bogy, "An elastic-plastic model for the contact of rough surfaces", *J. Tribology* (*Trans. ASME*) **109** (1987), 257–263.

[Etsion et al. 2005] I. Etsion, Y. Kligerman, and Y. Kadin, "Unloading of an elastic-plastic loaded spherical contact", *Int. J. Solids Struct.* **42**:13 (2005), 3716–3729.

[Evseev et al. 1991] D. G. Evseev, B. M. Medvedev, and G. G. Grigoriyan, "Modification of the elastic-plastic model for the contact of rough surfaces", *Wear* **150**:1-2 (1991), 79–88.

[Fischer-Cripps 2000] A. C. Fischer-Cripps, *Introduction to contact mechanics*, Springer, New York, 2000.

[Goodman 1962] L. E. Goodman, "Contact stress analysis of normally loaded rough spheres", *J. Appl. Mech.* (*Trans. ASME*) **29** (1962), 515–522.

[Hills and Sackfield 1987] D. A. Hills and A. Sackfield, "The stress field induced by normal contact between dissimilar spheres", *J. Appl. Mech.* (*Trans. ASME*) **54** (1987), 8–14.

[Jackson and Green 2005] R. L. Jackson and I. Green, "A finite element study of elasto-plastic hemispherical contact against a rigid flat", *J. Tribology* (*Trans. ASME*) **127**:2 (2005), 343–354.

[Johnson 1985] K. L. Johnson, *Contact mechanics*, Cambridge University Press, Cambridge, MA, 1985.

[Johnson et al. 1973] K. L. Johnson, J. J. O'Connor, and A. C. Woodward, "The effect of the indenter elasticity on the hertzian fracture of brittle materials", *P. Roy. Soc. Lond. A Mat.* **A334** (1973), 95–117.

[Kogut and Etsion 2002] L. Kogut and I. Etsion, "Elastic-plastic contact analysis of a sphere and a rigid flat", *J. Appl. Mech.* (*Trans. ASME*) **69**:5 (2002), 657–662.

[Kogut and Komvopoulos 2004] L. Kogut and K. Komvopoulos, "Analysis of the spherical indentation cycle for elastic-perfectly plastic solids", *J. Mater. Res.* **19**:12 (2004), 3641–3653.

[Kosior et al. 1999] F. Kosior, N. Guyot, and G. Maurice, "Analysis of frictional contact problem using boundary element method and domain decomposition method", *Int. J. Numer. Methods. Eng.* **46**:1 (1999), 65–82.

[Mesarovic and Fleck 1999] S. D. Mesarovic and N. A. Fleck, "Spherical indentation of elastic-plastic solids", *P. Roy. Soc. Lond. A Mat.* **A455**:1987 (1999), 2707–2728.

[Mesarovic and Fleck 2000] S. D. Mesarovic and N. A. Fleck, "Frictionless indentation of dissimilar elastic-plastic spheres", *Int. J. Solids Struct.* **37**:46-47 (2000), 7071–7091.

[Quicksall et al. 2004] J. J. Quicksall, R. L. Jackson, and I. Green, "Elasto-plastic hemispherical contact models for various mechanical properties", *Proc. Inst. Mech. Eng.* **218**:4 (2004), 313–322. Part J: J. Eng. Tribol.

[Spence 1968] D. A. Spence, "Self-similar solutions to adhesive contact problems with incremental loading", *P. Roy. Soc. Lond. A Mat.* **A305**:1480 (1968), 55–80.

[Spence 1975] D. A. Spence, "The Hertz contact problem with finite friction", *J. Elasticity* **5**:3-4 (1975), 297–319.

[Tabor 1959] D. Tabor, "Junction growth in metallic friction: the role of combined stresses and surface contamination", *P. Roy. Soc. Lond. A Mat.* **A251**:1266 (1959), 378–393.

[Zhao et al. 2000] Y. Zhao, D. M. Maietta, and L. Chang, "An asperity microcontact model incorporating the transition from elastic deformation to fully plastic flow", *J. Tribology* (*Trans. ASME*) **122**:1 (2000), 86–93.

VICTOR BRIZMER: brizmer@technion.ac.il
Dept. of Mechanical Engineering, Technion, Haifa 32000, Israel

YUVAL ZAIT: syuvalza@t2.technion.ac.il
Dept. of Mechanical Engineering, Technion, Haifa 32000, Israel

YURI KLIGERMAN: mermdyk@technion.ac.il
Dept. of Mechanical Engineering, Technion, Haifa 32000, Israel

IZHAK ETSION: etsion@technion.ac.il
Dept. of Mechanical Engineering, Technion, Haifa 32000, Israel

ASYMPTOTIC FIELDS AT FRICTIONLESS AND FRICTIONAL COHESIVE CRACK TIPS IN QUASIBRITTLE MATERIALS

QiZhi Xiao and Bhushan Lal Karihaloo

The lack of any work on the asymptotic fields at the tips of cohesive cracks belies the widespread use of cohesive crack models. This study is concerned with the solution of asymptotic fields at cohesive crack tips in quasibrittle materials. Only normal cohesive separation is considered, but the effect of Coulomb friction on the cohesive crack faces is studied. The special case of a pure mode I cohesive crack is fully investigated. The solution is valid for any separation law that can be expressed in a special polynomial form. It is shown that many commonly used separation laws of quasibrittle materials, for example, rectangular, linear, bilinear, and exponential, can be easily expressed in this form. The asymptotic fields obtained can be used as enrichment functions in the extended/generalized finite element method at the tip of long cohesive cracks, as well as short branches/kinks.

1. Introduction

Cohesive zone (or crack) models, which were introduced by Barenblatt [1962] and Dugdale [1960] for elastoplastic fracture in ductile metals, and by Hillerborg et al. [1976] for quasibrittle materials (who called them fictitious crack models), have become an important tool for describing localization and failure in engineering materials and structures. In a cohesive zone model, the nonlinear fracture process zone — due to degrading mechanisms such as plastic microvoiding or microcracking — in front of the actual crack tip is lumped into a discrete line (two-dimensional) or plane (three-dimensional) and represented by stress-displacement relationships across this line or plane. The cohesive crack model is a constitutive assumption in the sense that a cohesive crack can develop anywhere in a specimen or a structure, and not only ahead of a preexisting crack tip. For ductile fracture, the most important parameters of the cohesive zone model are the tensile strength f_t and the work of separation or fracture energy G_c (see, [Hutchinson and Evans 2000]), which is the work needed to create a unit area of a fully developed crack. For quasibrittle fracture, the decohesion law stems from microcracking as in concrete or ceramics; the shape of the stress-separation relation (called the softening or stress-crack opening or tension-softening curve) plays a much bigger role and is sometimes even more important than the value of the tensile strength f_t (see [Chandra et al. 2002]).

Elices et al. [2002] have reviewed the background of cohesive crack models and discussed the determination of the tension-softening function by inverse analysis procedures. They also illustrated the predictive capability of the cohesive zone model for concrete, glassy polymer and steel. Recently, Karihaloo et al. [2003] have proposed a simple method for determining the true specific fracture energy of concrete and also a method [Abdalla and Karihaloo 2004] for constructing a softening curve corresponding to this energy. In the most widely used standard formulation of the cohesive crack model for quasibrittle

materials, it is assumed that the stress-strain behavior is isotropic linear elastic, and that the crack is initiated at a point where the maximum principal stress σ_I reaches the tensile strength f_t, and that the crack is oriented normal to the direction of σ_I. An evolution law is also postulated for the monotonic mode I loading so that the cohesive stress is a unique function of the crack opening which, for concrete, decreases monotonically along the cohesive zone. Although this standard formulation of the cohesive crack model is highly simplified, it is able to capture the essence of fracture of concrete specimens and structures.

A concise overview of the various ways to implement numerically the cohesive zone methodology is given in [de Borst et al. 2004]. The recently developed extended/generalized finite element method (XFEM/GFEM) (see [Moës et al. 1999; Strouboulis et al. 2001; Babuška et al. 2003; Karihaloo and Xiao 2003b; Xiao and Karihaloo 2005]) provides a proper representation of the discrete character of cohesive zone formulations avoiding any mesh bias. The XFEM/GFEM enriches the standard local FE approximations with known information about the problem, such as a displacement discontinuity across a crack, the asymptotic solution at a crack tip, or a strain discontinuity across an interface, with the use of the partition of unity (PU). It avoids meshes conforming with the discontinuity and adaptive remeshing as the discontinuity grows as is the case with the FEM. Wells and Sluys [2001], Moës and Belytschko [2002] and Hansbo and Hansbo [2004] analyzed a continuous cohesive crack that runs through an existing FE mesh. Remmers et al. [2003] further studied the possibility of defining cohesive segments that can arise at arbitrary locations and in arbitrary directions and thus allow for the resolution of complex crack patterns including crack nucleation at multiple locations, followed by growth and coalescence. In these existing XFEM/GFEM implementations of the cohesive zone, the enrichment function used at the cohesive crack tip is usually a jump function (cohesive crack tip touches the element boundary) [Wells and Sluys 2001] or a branch function [Moës and Belytschko 2002] which does not represent the true asymptotic nature of the displacement/stress field there. (Hansbo and Hansbo [2004] considered an element traversed by a discontinuity as a double element with each being used for the interpolation of one side of the discontinuity). Recently, Xiao and Karihaloo [2005] have demonstrated that, for a crack with traction-free faces, when the crack tip asymptotic field is available and used as an enrichment function, the XFEM/GFEM not only avoids using a mesh conforming with the crack but is also more accurate than FEM. However, it is necessary to ensure that the unknown coefficients of the crack tip field at all the enriched nodes are equal to one another. Hence XFEM/GFEM can use a much coarser mesh around the crack tip. However, when the enrichment function does not represent the true asymptotic nature of the crack tip field, the mesh needs to be refined in the same manner as in the FEM. Thus it is advantageous to know the true asymptotic fields around a cohesive crack tip.

Planas et al. [2003] discussed possible generalizations of the cohesive crack model to mixed mode. Many other studies on mixed mode cohesive cracks can also be found in the literature, for example, [Valente 1991; Cocchetti et al. 2002], but there is doubt about the accuracy of the cohesion-sliding relation because it is difficult to isolate it from frictional forces between the rough cohesive crack faces in quasibrittle materials such as concrete.

Coulomb friction along the contacting crack faces has been considered by many researchers. Deng [1994] studied the plane strain/stress asymptotic crack tip fields of stationary and steadily moving cracks along bimaterial interfaces and in homogeneous solids. He considered both anisotropic and isotropic solids. Leblond and Frelat [2004] studied crack kinking from an initially closed, homogeneous or

interface crack, in the presence of Coulomb friction. Bialas and Mróz [2005] analyzed progressive inter-face delamination failure in antiplane shear of an elastic plate bonded to a rigid substrate under monotonic loading by normal compressive stress and varying shear stress using the cohesive crack model. Mróz and Bialas [2005] considered a rigid softening interface (critical stress softening) under both monotonic and cyclic loadings.

In the cohesive cracks, the friction is considered for a finite opening. In this sense frictional cohesive cracks are different from the previously mentioned works on frictional contact of crack faces, where the crack faces are in contact and not open. However, in cohesive cracks, although the crack faces are not in contact because of the applied cohesive stresses, frictional forces can come into play between the faces when there is relative sliding.

Hong and Kim [2003] studied plane elastic eigenfunction expansions of the cohesive crack tip field due to the closing tractions and the separation-gradients at the cohesive zone ahead of a semiinfinite crack in an inverse manner, similar to that used by Karihaloo [1999]. In these works, the softening curve is not defined a priori but is obtained parametrically from the analysis. This often leads to softening diagrams that are not representative of real materials [Planas et al. 2001].

The lack of any work on the asymptotic fields at the tips of cohesive cracks belies the widespread use of cohesive crack models. In this study, we will solve the asymptotic fields at the tips of cohesive cracks in quasibrittle materials. The material outside the fracture process (that is, cohesive) zone is isotropic linear elastic. This is true of quasibrittle materials. We will consider frictionless as well as frictional cohesive cracks. We will use the eigenfunction expansion method of Williams [1957] and combine it with the complex function formalism of Muskhelishvili [1953] in the spirit of Sih and Liebowitz [1968]. The cohesive and frictional laws on the crack faces are imposed through appropriate boundary conditions.

This paper is organized as follows: Section 2 discusses a cohesive law of concrete suitable for the asymptotic analysis of cohesive cracks; Section 3 discusses the mathematical formulation and local symmetry and/or boundary conditions; Section 4 gives asymptotic fields for several cases; Section 5 discusses the applicability of the results obtained in Section 4 to other cohesive laws; the implementation of the asymptotic fields in XFEM/GFEM is illustrated in Section 6 with examples of mode I cohesive crack tip fields; and finally conclusions and discussion are presented in Section 7.

2. Reformulation of a tension-softening diagram for quasibrittle materials

Cornelissen et al. [1986] introduced the following exponential relation to fit their results from uniaxial tests on double edge notched normal and lightweight concrete panels:

$$\frac{\sigma}{f_t} = f\left(\frac{w}{w_c}\right) - \frac{w}{w_c} f(1), \qquad f\left(\frac{w}{w_c}\right) = \left[1 + \left(C_1 \frac{w}{w_c}\right)^3\right] e^{-C_2 w/w_c}. \qquad (1)$$

It fits their experimental results with a high degree of accuracy. In Equation (1), σ and f_t are the stress normal to the cohesive crack face and the uniaxial tensile strength, respectively; w and w_c are the opening displacement of the cohesive crack faces, and the critical opening displacement of the preexisting macrocrack tip at which the cohesive crack tip begins to grow; and C_1 and C_2 are fitting parameters. Details of the test set up as well as the cohesive relation Equation (1) can be found in [Karihaloo 1995].
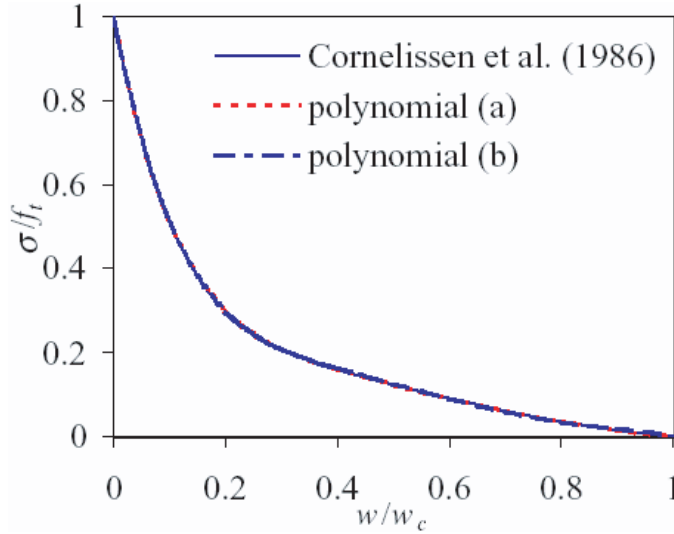
Figure 1. A comparison of formula Equation (1) with polynomials (2) and (3) for NC.

The following polynomial, denoted as polynomial (a), fits the results of Cornelissen et al. [1986] just as well as Equation (1)

$$\frac{\sigma}{f_t} = 1 + \sum_{i=1}^{5} a_i \left(\frac{w}{w_c}\right)^i - \left(1 + \sum_{i=1}^{5} a_i\right)\left(\frac{w}{w_c}\right)^6, \tag{2}$$

where the a_i, $i = 1, \ldots, 5$, are fitting parameters.

To simplify the derivation of the cohesive crack tip asymptotic fields, we will represent the relation Equation (1) by the polynomial, denoted polynomial (b),

$$\frac{\sigma}{f_t} = 1 + \sum_{i=1}^{5} \alpha_i \left(\frac{w}{w_c}\right)^{(2/3)i} - \left(1 + \sum_{i=1}^{5} \alpha_i\right)\left(\frac{w}{w_c}\right)^4, \tag{3}$$

where the α_i, $i = 1, \ldots, 5$, are fitting parameters.

Obviously, all three formulae (1)–(3) satisfy the following requirements: at the tip of the cohesive crack, $w/w_c = 0$ and $\sigma/f_t = 1$; at the tip of the preexisting traction-free macrocrack: if $w/w_c = 1$ then $\sigma/f_t = 0$.

Note that although there are five unknown parameters in (2) and (3) to be determined by regression, this is easier than the determination of the two parameters in the exponential relation (1).

For a normal concrete (NC) with density $2370\,\text{kg/m}^3$, compressive strength $f_c = 47\,\text{MPa}$, Young modulus $E = 39\,\text{GPa}$, $f_t = 3.2\,\text{MPa}$, $w_c = 160\,\mu\text{m}$, and specific fracture energy $G_F = 100\,\text{J/m}^2$ (area under the tension-softening curve), Cornelissen et al. [1986] fitted their experimental results by Equation (1) with $C_1 = 3$ and $C_2 = 6.93$. Their tension-softening diagram can be fitted by Equation (2) with $a_1 = -7.04$, $a_2 = 26.456$, $a_3 = -55.233$, $a_4 = 63.741$, and $a_5 = -38.305$. The correlation coefficient is 1. This diagram can also be fitted by Equation (3) with $\alpha_1 = -0.872$, $\alpha_2 = -16.729$, $\alpha_3 = 67.818$,
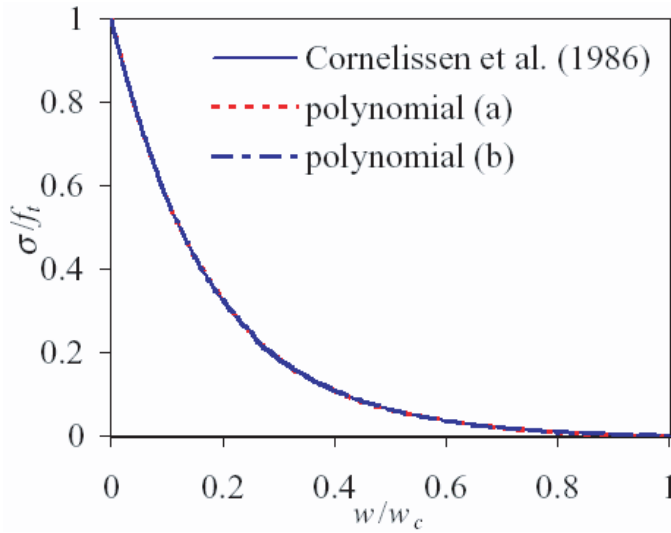
Figure 2. A comparison of formula Equation (1) with polynomials (2) and (3) for LC.

$\alpha_4 = -110.462$, and $\alpha_5 = 83.158$. The correlation coefficient is again 1. These three fittings are compared in Figure 1, and cannot be distinguished on the scale of the figure.

For a lightweight concrete (LC) with density 1865 kg/m^3, $f_c = 49$ MPa, $E = 22.4$ GPa, $f_t = 2.43$ MPa, $w_c = 140\,\mu$m, and $G_F = 61$ J/m^2, Cornelissen et al. [1986] fitted their experimental results by Equation (1) with $C_1 = 1$ and $C_2 = 5.64$. Their tension-softening diagram can be fitted by Equation (2) with $a_1 = -5.618$, $a_2 = 15.36$, $a_3 = -25.378$, $a_4 = 25.659$ and $a_5 = -14.525$. The correlation coefficient is 1. This diagram can also be fitted by polynomial Equation (3) with $\alpha_1 = -0.753$, $\alpha_2 = -12.335$, $\alpha_3 = 41.08$, $\alpha_4 = -57.205$, and $\alpha_5 = 38.412$. The correlation coefficient is again 1. These three fittings are compared in Figure 2, and again cannot be distinguished on the scale of the figure.

3. Mathematical formulation

Muskhelishvili [1953] showed that, for plane problems, the stresses and displacements in the Cartesian coordinate system (see Figure 3) can be expressed in terms of two analytic functions $\phi(z)$ and $\chi(z)$ of the complex variable $z = re^{i\theta}$

$$
\begin{aligned}
\sigma_x + \sigma_y &= 2\big[\phi'(z) + \overline{\phi'(z)}\,\big], \\
\sigma_y - \sigma_x + 2i\,\tau_{xy} &= 2\big[\bar{z}\phi''(z) + \chi''(z)\big], \\
2\mu(u+iv) &= \kappa\phi(z) - z\overline{\phi'(z)} - \overline{\chi'(z)},
\end{aligned}
\tag{4}
$$

where a prime denotes differentiation with respect to z, and an overbar denotes a complex conjugate. In Equation (4), $\mu = E/[2(1+\nu)]$ is the shear modulus; the Kolosov constant is $\kappa = 3 - 4\nu$ for plane strain or $\kappa = (3-\nu)/(1+\nu)$ for plane stress; E and ν are Young's modulus and the Poisson ratio.

For a general plane mixed mode I + II problem, the complex functions $\phi(z)$ and $\chi(z)$ can be chosen as series of complex eigenvalue Goursat functions [Sih and Liebowitz 1968; Owen and Fawkes 1983;
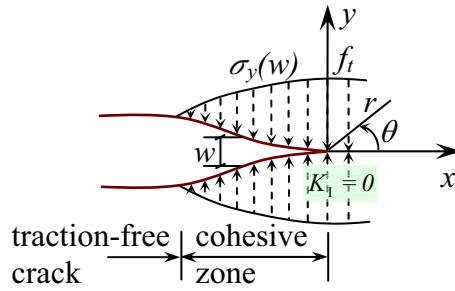
Figure 3. A real traction-free crack terminating in a fracture process (cohesive) zone (FPZ) with residual stress transfer capacity $\sigma_y(w)$ whose faces close smoothly near its tip ($K_1 = 0$). The material outside the FPZ is linear elastic, but within the FPZ is softening.

Karihaloo and Xiao 2003a]

$$
\begin{aligned}
\phi(z) &= \sum_{n=0} A_n z^{\lambda_n} \quad\;\; = \sum_{n=0} A_n r^{\lambda_n} e^{i\lambda_n\theta}, \\
\chi(z) &= \sum_{n=0} B_n z^{\lambda_n+1} = \sum_{n=0} B_n r^{\lambda_n+1} e^{i(\lambda_n+1)\theta},
\end{aligned}
\tag{5}
$$

where the complex coefficients are $A_n = a_{1n} + ia_{2n}$ and $B_n = b_{1n} + ib_{2n}$. The eigenvalues λ_n and coefficients a_{1n}, a_{2n}, b_{1n} and b_{2n} are real.

If we substitute the complex functions Equation (5) into (4), the complete series expansions of the displacements and stresses near the tip of the crack become

$$
2\mu u = \sum_{n=0} r^{\lambda_n} \Big\{ \kappa(a_{1n}\cos\lambda_n\theta - a_{2n}\sin\lambda_n\theta) + \lambda_n\big[-a_{1n}\cos(\lambda_n-2)\theta + a_{2n}\sin(\lambda_n-2)\theta\big] \\
+ (\lambda_n+1)(-b_{1n}\cos\lambda_n\theta + b_{2n}\sin\lambda_n\theta) \Big\},
\tag{6}
$$

$$
2\mu v = \sum_{n=0} r^{\lambda_n} \Big\{ \kappa(a_{1n}\sin\lambda_n\theta + a_{2n}\cos\lambda_n\theta) + \lambda_n\big[a_{1n}\sin(\lambda_n-2)\theta + a_{2n}\cos(\lambda_n-2)\theta\big] \\
+ (\lambda_n+1)(b_{1n}\sin\lambda_n\theta + b_{2n}\cos\lambda_n\theta) \Big\},
\tag{7}
$$

$$
\sigma_x = \sum_{n=0} r^{\lambda_n-1} \Big\{ 2\lambda_n\big[a_{1n}\cos(\lambda_n-1)\theta - a_{2n}\sin(\lambda_n-1)\theta\big] \\
- \lambda_n(\lambda_n-1)\big[a_{1n}\cos(\lambda_n-3)\theta - a_{2n}\sin(\lambda_n-3)\theta\big] \\
- (\lambda_n+1)\lambda_n\big[b_{1n}\cos(\lambda_n-1)\theta - b_{2n}\sin(\lambda_n-1)\theta\big] \Big\},
\tag{8}
$$

$$
\sigma_y = \sum_{n=0} r^{\lambda_n-1} \Big\{ 2\lambda_n\big[a_{1n}\cos(\lambda_n-1)\theta - a_{2n}\sin(\lambda_n-1)\theta\big] \\
+ \lambda_n(\lambda_n-1)\big[a_{1n}\cos(\lambda_n-3)\theta - a_{2n}\sin(\lambda_n-3)\theta\big] \\
+ (\lambda_n+1)\lambda_n\big[b_{1n}\cos(\lambda_n-1)\theta - b_{2n}\sin(\lambda_n-1)\theta\big] \Big\},
\tag{9}
$$

$$\tau_{xy} = \sum_{n=0} r^{\lambda_n-1} \Big\{ \lambda_n(\lambda_n - 1)\big[a_{1n}\sin(\lambda_n - 3)\theta + a_{2n}\cos(\lambda_n - 3)\theta\big]$$
$$+ (\lambda_n + 1)\lambda_n\big[b_{1n}\sin(\lambda_n - 1)\theta + b_{2n}\cos(\lambda_n - 1)\theta\big]\Big\}. \quad (10)$$

The opening displacement (COD) behind the cohesive zone tip is

$$w = v\big|_{\theta=\pi} - v\big|_{\theta=-\pi} = \sum_{n=0} \frac{r^{\lambda_n}}{\mu}\big[(\kappa + \lambda_n)a_{1n} + (\lambda_n + 1)b_{1n}\big]\sin\lambda_n\pi \quad (11)$$

and the sliding displacement of the crack faces is

$$\delta = u\big|_{\theta=\pi} - u\big|_{\theta=-\pi} = \sum_{n=0} \frac{r^{\lambda_n}}{\mu}\big[(\lambda_n - \kappa)a_{2n} + (\lambda_n + 1)b_{2n}\big]\sin\lambda_n\pi. \quad (12)$$

To complete the asymptotic analysis of the crack tip fields, solutions need to satisfy the proper symmetry conditions along the line of extension of the cohesive crack, and boundary conditions on the cohesive crack faces.

If the crack faces are traction-free (that is, there is no cohesive zone), then

$$\sigma_y\big|_{\theta=\pi} = \sigma_y\big|_{\theta=-\pi} = 0, \qquad \tau_{xy}\big|_{\theta=\pi} = \tau_{xy}\big|_{\theta=-\pi} = 0. \quad (13)$$

If normal cohesive separation applies to the crack faces, relationship Equation (3) needs to be satisfied over the cohesive zone. The stresses at the cohesive crack tip are nonsingular (because the stress intensity factor $K_I = 0$). Moreover, the following conditions need to be satisfied:

(a) If the cohesive crack faces are frictionless, we have

$$\sigma_y\big|_{\theta=\pi} = \sigma_y\big|_{\theta=-\pi} \neq 0, \qquad \tau_{xy}\big|_{\theta=\pi} = \tau_{xy}\big|_{\theta=-\pi} = 0, \quad (14)$$

(b) If the Coulomb friction is considered, we have

$$\sigma_y\big|_{\theta=\pi} = \sigma_y\big|_{\theta=-\pi} \neq 0, \qquad \tau_{xy}\big|_{\theta=\pi} = \tau_{xy}\big|_{\theta=-\pi} = -\mu_f\sigma_y\big|_{\theta=\pm\pi} \neq 0, \quad (15)$$

where μ_f equals the positive or negative value of the coefficient of kinetic friction, which is assumed to be constant, depending on the relative sliding direction of the two crack faces. Specifically, $\mu_f > 0$ when $\delta > 0$ and $\mu_f < 0$ when $\delta < 0$.

(c) If the cohesive crack faces are in pure mode I condition, we have

$$\sigma_y\big|_{\theta=\pi} = \sigma_y\big|_{\theta=-\pi} \neq 0, \quad \tau_{xy}\big|_{\theta=\pi} = \tau_{xy}\big|_{\theta=-\pi} = 0, \quad \tau_{xy}\big|_{\theta=0} = 0, \quad v\big|_{\theta=0} = 0. \quad (16)$$

In all three situations, the length of the process (cohesive) zone is either prescribed (that is, an initial cohesive zone exists before the loading is applied, and does not propagate under the present loading) or is determined by the condition $w = w_c$ in the normal cohesion-separation relation Equation (3) at the instant of growth of the preexisting traction-free crack.

4. Asymptotic crack tip fields

For completeness and later use, we will first deduce the Williams expansion for a traction-free crack without a cohesive zone.

4.1. *Williams expansions for cracks with traction-free faces.* After enforcing the traction-free conditions Equation (13) on the crack faces, we have

$$\lambda_n = \tfrac{1}{2}n, \ n = 0, 1, 2, \ldots \tag{17}$$

and

$$-b_{1n} = \frac{(n/2) + (-1)^n}{(n/2) + 1} \, a_{1n}, \qquad -b_{2n} = \frac{(n/2) - (-1)^n}{(n/2) + 1} \, a_{2n}. \tag{18}$$

It is easy to confirm that with the eigenvalues Equation (17) and the coefficient relationships (18) the solutions (6)–(10) reduce to the well-known Williams expansions [Owen and Fawkes 1983; Karihaloo and Xiao 2003a]:

$$u = \sum_{n=0}^{\infty} \frac{r^{n/2}}{2\mu} \left\{ a_{1n} \left[\left(\kappa + \tfrac{1}{2}n + (-1)^n \right) \cos \tfrac{1}{2}n\theta - \tfrac{1}{2}n \cos \left(\tfrac{1}{2}n - 2 \right)\theta \right] \right. \\ \left. - a_{2n} \left[\left(\kappa + \tfrac{1}{2}n - (-1)^n \right) \sin \tfrac{1}{2}n\theta - \tfrac{1}{2}n \sin \left(\tfrac{1}{2}n - 2 \right)\theta \right] \right\}, \tag{19}$$

$$v = \sum_{n=0}^{\infty} \frac{r^{n/2}}{2\mu} \left\{ a_{1n} \left[\left(\kappa - \tfrac{1}{2}n - (-1)^n \right) \sin \tfrac{1}{2}n\theta + \tfrac{1}{2}n \sin \left(\tfrac{1}{2}n - 2 \right)\theta \right] \right. \\ \left. + a_{2n} \left[\left(\kappa - \tfrac{1}{2}n + (-1)^n \right) \cos \tfrac{1}{2}n\theta + \tfrac{1}{2}n \cos \left(\tfrac{1}{2}n - 2 \right)\theta \right] \right\}, \tag{20}$$

$$\sigma_x = \sum_{n=1}^{\infty} \tfrac{1}{2}n \, r^{(n/2)-1} \left\{ a_{1n} \left[\left(2 + \tfrac{1}{2}n + (-1)^n \right) \cos \left(\tfrac{1}{2}n - 1 \right)\theta - \left(\tfrac{1}{2}n - 1 \right) \cos \left(\tfrac{1}{2}n - 3 \right)\theta \right] \right. \\ \left. - a_{2n} \left[\left(2 + \tfrac{1}{2}n - (-1)^n \right) \sin \left(\tfrac{1}{2}n - 1 \right)\theta - \left(\tfrac{1}{2}n - 1 \right) \sin \left(\tfrac{1}{2}n - 3 \right)\theta \right] \right\}, \tag{21}$$

$$\sigma_y = \sum_{n=1}^{\infty} \tfrac{1}{2}n \, r^{(n/2)-1} \left\{ a_{1n} \left[\left(2 - \tfrac{1}{2}n - (-1)^n \right) \cos \left(\tfrac{1}{2}n - 1 \right)\theta + \left(\tfrac{1}{2}n - 1 \right) \cos \left(\tfrac{1}{2}n - 3 \right)\theta \right] \right. \\ \left. - a_{2n} \left[\left(2 - \tfrac{1}{2}n + (-1)^n \right) \sin \left(\tfrac{1}{2}n - 1 \right)\theta + \left(\tfrac{1}{2}n - 1 \right) \sin \left(\tfrac{1}{2}n - 3 \right)\theta \right] \right\}, \tag{22}$$

$$\tau_{xy} = \sum_{n=1}^{\infty} \tfrac{1}{2}n \, r^{(n/2)-1} \left\{ a_{1n} \left[\left(\tfrac{1}{2}n - 1 \right) \sin \left(\tfrac{1}{2}n - 3 \right)\theta - \left(\tfrac{1}{2}n + (-1)^n \right) \sin \left(\tfrac{1}{2}n - 1 \right)\theta \right] \right. \\ \left. + a_{2n} \left[\left(\tfrac{1}{2}n - 1 \right) \cos \left(\tfrac{1}{2}n - 3 \right)\theta - \left(\tfrac{1}{2}n - (-1)^n \right) \cos \left(\tfrac{1}{2}n - 1 \right)\theta \right] \right\}. \tag{23}$$

The displacements corresponding to $n = 0$

$$u_0 = \frac{\kappa + 1}{2\mu} \, a_{10}, \qquad v_0 = \frac{\kappa + 1}{2\mu} \, a_{20} \tag{24}$$

are rigid body translations at the crack tip. The displacements corresponding to a_{22}

$$\begin{aligned} \hat{u}_2 &= -\frac{\kappa + 1}{2\mu} \, a_{22} r \sin\theta = -\frac{\kappa + 1}{2\mu} \, a_{22} y, \\ \hat{v}_2 &= \ \ \frac{\kappa + 1}{2\mu} \, a_{22} r \cos\theta = \frac{\kappa + 1}{2\mu} \, a_{22} x \end{aligned} \tag{25}$$

represent the rigid body rotation $\theta_0 = -(\kappa + 1)a_{22}/(2\mu)$ with respect to the crack tip, with $x = r\cos\theta$ and $y = r\sin\theta$. Terms involving a_{10}, a_{20} and a_{22} do not contribute to the strains or stresses. Terms involving coefficients a_{1n} (a_{2n}), $n \geq 1$, correspond to pure mode I (II) expansions. The corresponding opening and sliding displacements of the crack faces are

$$w = \sum_{n=1,3,5,\dots} \frac{r^{n/2}}{\mu} a_{1n}(\kappa + 1)\sin(n\pi/2) \tag{26}$$

$$\delta = \sum_{n=1,3,5,\dots} -\frac{r^{n/2}}{\mu} a_{2n}(\kappa + 1)\sin(n\pi/2). \tag{27}$$

4.2. *Frictionless cohesive crack with normal cohesive separation.* The relationship Equation (3) between cohesion and normal separation will be discussed below after considering conditions Equation (14) on the crack faces.

Imposition of the left side of Equation (14) on (9) gives

$$(a_{2n} + b_{2n})\sin(\lambda_n - 1)\pi = 0 \implies \begin{cases} \lambda_n = n + 1, & \text{for } n = 0, 1, 2, \dots \quad \text{(a)} \\ \text{or} \\ b_{2n} = -a_{2n} & \text{(b)} \end{cases}$$

and imposition of the right side of Equation (14) on (10) gives

$$\left[(\lambda_n - 1)a_{1n} + (\lambda_n + 1)b_{1n}\right]\sin(\lambda_n - 1)\pi = 0$$
$$\left[(\lambda_n - 1)a_{2n} + (\lambda_n + 1)b_{2n}\right]\cos(\lambda_n - 1)\pi = 0.$$

We thus have for case (a)

$$b_{2n} = -\frac{\lambda_n - 1}{\lambda_n + 1} a_{2n},$$

and for case (b)

$$\cos(\lambda_n - 1)\pi = 0, \quad \text{or equivalently,} \quad \lambda_n = \frac{2n + 1}{2} + 1 \text{ for } n = 0, 1, 2, \dots$$

and

$$b_{1n} = -\frac{\lambda_n - 1}{\lambda_n + 1} a_{1n}.$$

Taken together, the solutions are composed of two parts. The first part corresponds to integer eigenvalues

$$\text{(a)} \quad \lambda_n = n + 1, \qquad b_{2n} = -\frac{n}{n + 2}a_{2n}, \qquad n = 0, 1, 2, \dots, \tag{28}$$

giving

$$\sigma_y\big|_{\theta=\pm\pi} = \sum_{n=0}(n + 2)(n + 1)r^n(a_{1n} + b_{1n})\cos n\pi$$

or

$$\hat{\sigma}_y = \frac{\sigma_y|_{\theta=\pm\pi}}{f_t} = \sum_{n=0} c_n r^n = 1 + \sum_{n=1} c_n r^n, \tag{29}$$

where

$$c_n = \frac{(n+2)(n+1)(a_{1n}+b_{1n})\cos n\pi}{f_t} \quad \text{and} \quad c_0 = \frac{2(a_{10}+b_{10})}{f_t} = 1, \tag{30}$$

since $\sigma_y|_{\theta=\pm\pi} = f_t$ when $r \to 0$.

The opening and sliding displacements of the cohesive crack faces vanish for integer eigenvalues

$$w = 0 \qquad \text{and} \qquad \delta = 0. \tag{31}$$

The second part of the asymptotic solutions corresponds to noninteger eigenvalues

$$\text{(b) } \lambda_n = \frac{2n+3}{2}, \qquad b_{1n} = -\frac{2n+1}{2n+5}a_{1n}, \qquad b_{2n} = -a_{2n}, \ n = 0, 1, 2, \ldots, \tag{32}$$

giving

$$\sigma_y|_{\theta=\pm\pi} = 0,$$

$$w = \sum_{n=0}^{\infty} \frac{r^{(2n+3)/2}}{\mu}\left[\left(\kappa + \frac{2n+3}{2}\right)a_{1n} + \frac{2n+5}{2}b_{1n}\right]\sin\frac{2n+3}{2}\pi \tag{33}$$

or

$$\hat{w} = \frac{w}{w_c} = \sum_{n=0} \bar{d}_n r^{(2n+3)/2},$$

$$\bar{d}_n = \frac{\left[(\kappa + (2n+3)/2)a_{1n} + ((2n+5)/2)b_{1n}\right]\sin\left((2n+3)/2\right)\pi}{\mu w_c}, \tag{34}$$

$$\delta = \sum_{n=0} \frac{r^{(2n+3)/2}}{\mu}\left[\left(\frac{2n+3}{2} - \kappa\right)a_{2n} + \frac{2n+5}{2}b_{2n}\right]\sin\frac{2n+3}{2}\pi. \tag{35}$$

The cohesive separation relationship Equation (3) is rewritten in normalized form using (29) and (34):

$$\hat{\sigma}_y = 1 + \sum_{i=1}^{5}\alpha_i\hat{w}^{(2/3)i} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)\hat{w}^4. \tag{36}$$

Consider the truncated $N+1$ terms of $\hat{w}$ Equation (34), and set $d_0 = \bar{d}_0$, $d_n = \bar{d}_n/d_0$ for $n > 1$. We have

$$\hat{w} = d_0 r^{3/2}\left(1 + \sum_{n=1}^{N}d_n r^n\right). \tag{37}$$

The expansion of $\hat{w}$ raised to the power $(2/3)i$ is also truncated to $N+1$ terms, since these terms include only the truncated $N+1$ terms of $\hat{w}$. Hence

$$\hat{w}^{(2/3)i} = d_0^{(2/3)i} r^i\left(1 + \sum_{n=1}^{N}\beta_{in} r^n\right) \tag{38}$$

with

$$\beta_{in} = \frac{f_i^{(n)}(0)}{n!}, \qquad f_i(r) = \left(1 + \sum_{n=1}^{N}d_n r^n\right)^{(2/3)i}, \tag{39}$$

where $f_i^{(n)}(0)$ denotes the n-th derivative at $r = 0$.

The first five derivatives of $f_i(r)$ Equation (39) are given in the Appendix, and the corresponding five coefficients β_{in} are

$$\beta_{i1} = \tfrac{2}{3}id_1,$$

$$\beta_{i2} = \tfrac{1}{3}i\left(\tfrac{2}{3}i - 1\right)d_1^2 + \tfrac{2}{3}id_2,$$

$$\beta_{i3} = \tfrac{1}{9}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)d_1^3 + \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)d_1d_2 + \tfrac{2}{3}id_3,$$

$$\beta_{i4} = \tfrac{1}{36}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)\left(\tfrac{2}{3}i - 3\right)d_1^4 + \tfrac{1}{3}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)d_1^2d_2$$
$$\quad + \tfrac{1}{3}i\left(\tfrac{2}{3}i - 1\right)d_2^2 + \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)d_1d_3 + \tfrac{2}{3}id_4, \tag{40}$$

$$\beta_{i5} = \tfrac{1}{180}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)\left(\tfrac{2}{3}i - 3\right)\left(\tfrac{2}{3}i - 4\right)d_1^5,$$
$$\quad + \tfrac{1}{9}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)\left(\tfrac{2}{3}i - 3\right)d_1^3d_2$$
$$\quad + \tfrac{1}{3}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)d_1d_2^2 + \tfrac{1}{3}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)d_1^2d_3$$
$$\quad + \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)d_2d_3 + \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)d_1d_4 + \tfrac{2}{3}id_5.$$

With the use of Equation (38), the right hand side of the cohesive relationship (36) becomes

$$1 + \sum_{i=1}^{5} \alpha_i d_0^{(2/3)i} r^i \left(1 + \sum_{n=1}^{N} \beta_{in} r^n\right) - \left(1 + \sum_{i=1}^{5} \alpha_i\right)d_0^4 r^6 \left(1 + \sum_{n=1}^{N} \beta_{6n} r^n\right).$$

If we choose $N = 5$, then after satisfying the cohesive relationship Equation (36) we have the following expressions for the coefficients c_n in Equation (29):

$$c_1 = \alpha_1 d_0^{2/3},$$

$$c_2 = \alpha_2 d_0^{4/3} + \alpha_1 d_0^{2/3}\beta_{11},$$

$$c_3 = \alpha_3 d_0^2 + \alpha_1 d_0^{2/3}\beta_{12} + \alpha_2 d_0^{4/3}\beta_{21},$$

$$c_4 = \alpha_4 d_0^{8/3} + \alpha_1 d_0^{2/3}\beta_{13} + \alpha_2 d_0^{4/3}\beta_{22} + \alpha_3 d_0^2\beta_{31},$$

$$c_5 = \alpha_5 d_0^{10/3} + \alpha_1 d_0^{2/3}\beta_{14} + \alpha_2 d_0^{4/3}\beta_{23} + \alpha_3 d_0^2\beta_{32} + \alpha_4 d_0^{8/3}\beta_{41},$$

$$c_6 = \alpha_1 d_0^{2/3}\beta_{15} + \alpha_2 d_0^{4/3}\beta_{24} + \alpha_3 d_0^2\beta_{33} + \alpha_4 d_0^{8/3}\beta_{42} + \alpha_5 d_0^{10/3}\beta_{51} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4,$$

$$c_7 = \alpha_2 d_0^{4/3}\beta_{25} + \alpha_3 d_0^2\beta_{34} + \alpha_4 d_0^{8/3}\beta_{43} + \alpha_5 d_0^{10/3}\beta_{52} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4\beta_{61},$$

$$c_8 = \alpha_3 d_0^2\beta_{35} + \alpha_4 d_0^{8/3}\beta_{44} + \alpha_5 d_0^{10/3}\beta_{53} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4\beta_{62},$$

$$c_9 = \alpha_4 d_0^{8/3}\beta_{45} + \alpha_5 d_0^{10/3}\beta_{54} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4\beta_{63},$$

$$c_{10} = \alpha_5 d_0^{10/3}\beta_{55} - \left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4\beta_{64},$$

$$c_{11} = -\left(1 + \sum_{i=1}^{5}\alpha_i\right)d_0^4\beta_{65}. \tag{41}$$

The asymptotic solution above is not for a pure mode I cohesive crack tip (compare Equation (14) and (16)), since along the line of extension of the crack, $\theta = 0$, the shear stress does not vanish ($\tau_{xy} \neq 0$).

For noninteger eigenvalues Equation (39), the coefficients a_{1n} and a_{2n} may be regarded as independent, so that coefficients b_{1n} are linearly dependent on a_{1n} and b_{2n} on a_{2n}. For integer eigenvalues (28), coefficients a_{1n} and a_{2n} may also be regarded as independent, so that coefficients b_{2n} now depend linearly on a_{2n}. However, the coefficients b_{1n} for integer eigenvalues will depend both linearly on a_{1n} for integer eigenvalues and nonlinearly on a_{1n} for noninteger eigenvalues via (37), (34), (37), (39) and (41). The inherent nonlinear nature of the problem is reflected in these nonlinear relationships between the coefficients of the asymptotic fields.

The displacements (8), (7) corresponding to $\lambda_{-1} = 0$, or $n = -1$ in (35) are rigid body translations at the crack tip

$$2\mu u_{-1} = \kappa a_{1,-1} - b_{1,-1}, \qquad 2\mu v_{-1} = \kappa a_{2,-1} + b_{2,-1}. \tag{42}$$

The displacements corresponding to a_{20} ($n = 0$, $\lambda_0 = 1$ and $b_{20} = 0$ from Equation (28)) represent rigid body rotation with respect to the crack tip

$$2\mu \hat{u}_0 = -r(\kappa + 1)a_{20} \sin\theta, \qquad 2\mu \hat{v}_0 = r(\kappa + 1)a_{20} \cos\theta. \tag{43}$$

4.3. *Coulomb frictional cohesive crack with normal cohesive separation.* In principle, a cohesive relationship can also be considered in the tangential direction for quasibrittle materials. However, this is a contentious issue, since it is difficult to separate the cohesive-sliding relation from the frictional force between the rough cohesive crack faces. Hence, in the following, we consider the Coulomb friction between the crack faces instead of a tangential cohesive relationship. The corresponding boundary conditions are Equation (15).

The complete asymptotic solutions are again composed of two parts. The first part corresponding to integer eigenvalues is similar to case (a) in Section 4.2 but with different constraints on the coefficients

$$\lambda_n = n + 1, \qquad na_{2n} + (n+2)b_{2n} = -\mu_f(n+2)(a_{1n} + b_{1n}), \ n = 0, 1, 2, \ldots. \tag{44}$$

From (44), we have

$$b_{2n} = -\frac{n}{n+2}a_{2n} - \mu_f(a_{1n} + b_{1n}).$$

When $\mu_f = 0$, the cohesive crack faces are frictionless, and Equation (44) reduces to (28). These solutions have nonzero σ_y and τ_{xy} along the cohesive crack faces, but zero crack opening w. The second part of the asymptotic solutions corresponding to noninteger eigenvalues satisfies

$$b_{1n} = -\frac{\lambda_n - 1}{\lambda_n + 1}a_{1n}, \qquad b_{2n} = -a_{2n}, \quad (\mu_f a_{1n} - a_{2n})\cos(\lambda_n - 1)\pi = 0. \tag{45}$$

If we assume that

$$\mu_f a_{1n} - a_{2n} \neq 0, \tag{46}$$

the third equation in Equation (45) gives

$$\cos(\lambda_n - 1)\pi = 0 \tag{47}$$

so that the second part of asymptotic solutions is identical to (b) of Section 4.2 (that is, (33)–(35)).

The remaining solution procedure and final asymptotic solutions as well as the dependence of the coefficients are similar to those in 4.2. Equations (42) and (43) again represent the rigid body modes for the present case.

4.4. *A pure mode I cohesive crack.*

For a pure mode I cohesive crack, the crack faces are frictionless. After satisfying the conditions (16)$_3$ and (16)$_4$, we have

$$(\lambda_n - 1)a_{2n} + (\lambda_n + 1)b_{2n} = 0 \tag{48}$$
$$(\kappa + \lambda_n)a_{2n} + (\lambda_n + 1)b_{2n} = 0$$

and finally $a_{2n} = b_{2n} = 0$. Conditions (16)$_1$ and (16)$_2$ are satisfied if $[(\lambda_n-1)a_{1n}+(\lambda_n+1)b_{1n}]\sin(\lambda_n-1)\pi$ vanishes, which is to say,

$$
\begin{cases}
\lambda_n = n + 1, n = 0, 1, 2, \ldots & \text{(c)} \\
\text{or} \\
b_{1n} = -\dfrac{\lambda_n - 1}{\lambda_n + 1}a_{1n}. & \text{(d)}
\end{cases}
$$

Solution (c) gives the same normal cohesive stress σ_y as Equation (29) along the cohesive crack faces and a nonvanishing σ_y along the line of extension of the crack, but without a jump in the normal displacement w, that is, as in Equation (31).

Solution (d) corresponds to noninteger eigenvalues. Without loss of generality, but to simplify the enforcement of the normal cohesive relationship Equation (36), we can choose such noninteger eigenvalues that result in no tractions on the cohesive crack faces (but a nonvanishing σ_y along the line of extension of the crack) and a displacement discontinuity w in the normal direction. In other words,

$$\sigma_y\big|_{\theta=\pm\pi} = 0 \Rightarrow (a_{1n} + b_{1n})\cos(\lambda_n - 1)\pi = 0 \Rightarrow \lambda_n = \frac{2n + 3}{2}, \ n = 0, 1, 2, \ldots. \tag{49}$$

It is easy to confirm that these solutions are nothing but the nonsingular odd terms in pure mode I Williams expansions Equation (19)–Equation (23) for a traction-free crack:

$$u = \sum_{n=1} \frac{r^{(2n+1)/2}}{2\mu}a_{1n}\left[\left(\kappa + \frac{2n-1}{2}\right)\cos\frac{2n+1}{2}\theta - \frac{2n+1}{2}\cos\frac{2n-3}{2}\theta\right] \tag{50}$$

$$v = \sum_{n=1} \frac{r^{(2n+1)/2}}{2\mu}a_{1n}\left[\left(\kappa - \frac{2n-1}{2}\right)\sin\frac{2n+1}{2}\theta + \frac{2n+1}{2}\sin\frac{2n-3}{2}\theta\right] \tag{51}$$

$$\sigma_x = \sum_{n=1} \frac{2n+1}{2}r^{(2n-1)/2}a_{1n}\left(\frac{2n+3}{2}\cos\frac{2n-1}{2}\theta - \frac{2n-1}{2}\cos\frac{2n-5}{2}\theta\right) \tag{52}$$

$$\sigma_y = \sum_{n=1} \frac{2n+1}{2}r^{(2n-1)/2}a_{1n}\left(\frac{-2n+5}{2}\cos\frac{2n-1}{2}\theta + \frac{2n-1}{2}\cos\frac{2n-5}{2}\theta\right) \tag{53}$$

$$\tau_{xy} = \sum_{n=1} \frac{2n+1}{2}r^{(2n-1)/2}a_{1n}\left(\frac{2n-1}{2}\sin\frac{2n-5}{2}\theta - \frac{2n-1}{2}\sin\frac{2n-1}{2}\theta\right). \tag{54}$$

Solutions (c) and (d) together give the asymptotic solutions for a pure mode I cohesive crack. The normal cohesive relationship is satisfied in the same way as in Section 4.2. The COD corresponding to Equation (51) is

$$
\begin{aligned}
w &= \sum_{n=0} \frac{r^{(2n+3)/2}}{\mu} a_{1n}(\kappa + 1) \sin \frac{2n+3}{2}\pi \\
\hat{w} &= \sum_{n=0} \bar{d}_n r^{(2n+3)/2}, \qquad \bar{d}_n = \frac{(\kappa + 1)}{\mu w_c} a_{1n} \sin \frac{2n+3}{2}\pi.
\end{aligned}
\tag{55}
$$

Now $\hat{w}$ can again be reformulated as Equation (37), but the coefficients d_n are different. For the cohesive law Equation (36), the formal relationships between the coefficients will be the same as Equation (41).

Note that the rigid body modes for the present case are no longer represented by Equation (42) and Equation (43), but are

$$
\begin{aligned}
u_{\text{rigid}} &= u_0 + y\theta_0 = u_0 + \theta_0 r \sin\theta, \\
v_{\text{rigid}} &= v_0 - x\theta_0 = v_0 - \theta_0 r \cos\theta.
\end{aligned}
\tag{56}
$$

These rigid body modes must be appended to the asymptotic fields above to obtain the complete crack tip displacements. However, in the analysis of pure mode I crack problems, it is more efficient to use symmetry conditions along the line of extension of the crack, and set $u_0 = v_0 = \theta_0 = 0$.

5. Applicability of the results to other cohesive-separation diagrams

From the formulation in Section 4, it became clear that the eigenvalues and asymptotic fields are uniquely defined for traction-free cracks (Section 4.1) and frictionless cohesive cracks with normal cohesion between their faces (Section 4.2). However, for a crack with normal cohesion and Coulomb friction (Section 4.3), and a pure mode I cohesive crack (Section 4.4), the eigenvalues and asymptotic fields are not completely unique. Additional assumptions had to be made to ensure uniqueness. Thus, in Section 4.3 we imposed the additional condition Equation (46) and in Section 4.4 the condition (49). These conditions however do not lead to any loss of generality.

The derivations above were for a special form of the normal cohesion-separation relation (3), or equivalently, (36), which made the expansion of a power of $\hat{w}$ in (37) possible, as shown in (38). Since this relation has five free parameters, it is believed to be able to fit a large amount of experimental data on many grades of concrete. It was already shown above to represent almost exactly the exponential relation (1). Below we show that it can equally accurately represent other normal cohesion-separation relations commonly used for quasibrittle materials.

The widely used linear tension-softening law

$$
\hat{\sigma}_y = 1 - \hat{w}
\tag{57}
$$

cannot be used in the previous asymptotic analysis since $\hat{\sigma}_y$ only includes terms corresponding to integer eigenvalues (see, for example, (29)), and $\hat{w}$ only includes terms corresponding to noninteger eigenvalues (see, for example, (34)). However, (57) can be represented by (3) or (36) with nonvanishing coefficients $\alpha_1 = -0.2612$ and $\alpha_2 = -1.0215$, that is,

$$
\hat{\sigma}_y = 1 + \alpha_1 \hat{w}^{\frac{2}{3}} + \alpha_2 \hat{w}^{\frac{4}{3}} - (1 + \alpha_1 + \alpha_2)\hat{w}^2.
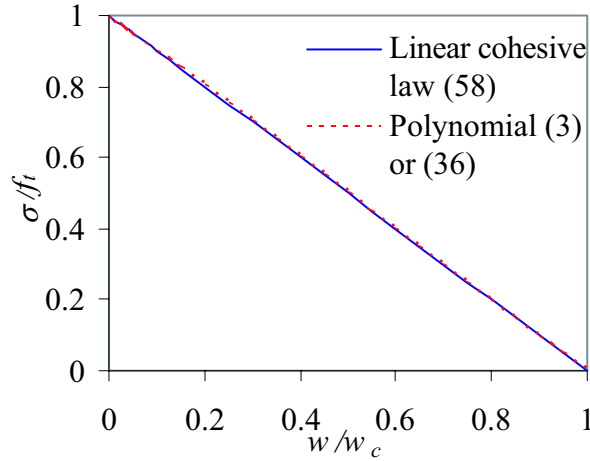$$

Figure 4. Linear tension-softening law.

The correlation coefficient is 1. The linear law Equation (57) is compared with Equation (3) in Figure 4; they cannot be distinguished on the scale of the figure. Hence results obtained in Section 4 can be used directly to the linear tension-softening law.

Similarly, the widely used bilinear tension-softening law (Figure 5)

$$\hat{\sigma} = \begin{cases} 1 - (1 - \hat{f}_1)\dfrac{\hat{w}}{\hat{w}_1}, & 0 \le \hat{\sigma} \le \hat{f}_1 \\[2mm] \dfrac{\hat{f}_1}{1 - \hat{w}_1}(1 - \hat{w}), & \hat{f}_1 < \hat{\sigma} \le 1, \end{cases} \tag{58}$$

(where $\hat{f}_1 = f_1/f_t$ and $\hat{w}_1 = w_1/w_c$) is also not suitable for the asymptotic analysis in Section 4.
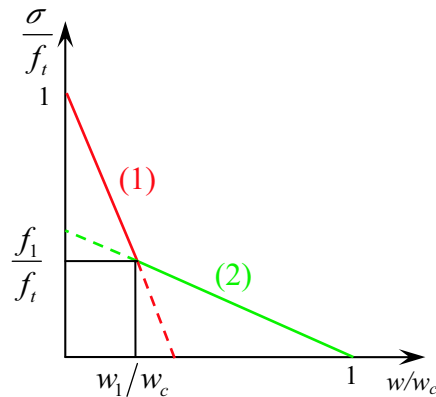
Figure 5. Bilinear tension-softening law.

However, its two linear parts Equations (1) and (2) can be rewritten into two linear laws as shown in 5. The first part can be written into (57) using a new definition of w_c as

$$w_c = \frac{w_1}{1 - \hat{f}_1}. \tag{59}$$

The second part can be written into (57) using a new definition of f_t as

$$f_t = \frac{f_1}{1 - \hat{w}_1}. \tag{60}$$

Then the asymptotic field of the bilinear law can be obtained by the corresponding linear cohesive laws.

Moreover, the procedures are also applicable to cohesion-separation relations in which the power of $\hat{w}$ is $2i$ rather than $2i/3$ as in Equation (3) or (36)

$$\hat{\sigma}_y = 1 + \sum_{i=1}^{5} \alpha_i \hat{w}^{2i} - \left(1 + \sum_{i=1}^{5} \alpha_i\right) \hat{w}^{12}. \tag{61}$$

For this relation, the counterpart of (38) becomes

$$\hat{w}^{2i} = d_0^{2i} r^{3i} \left(1 + \sum_{n=1}^{N} \beta_{in} r^n\right), \tag{62}$$

where

$$\beta_{in} = \frac{f_i^{(n)}(0)}{n!}, \qquad f_i(r) = \left(1 + \sum_{n=1}^{N} d_n r^n\right)^{2i}. \tag{63}$$

Substitution of Equation (62) into the right hand side of (61) gives

$$1 + \sum_{i=1}^{5} \alpha_i d_0^{2i} r^{3i} \left(1 + \sum_{n=1}^{N} \beta_{in} r^n\right) - \left(1 + \sum_{i=1}^{5} \alpha_i\right) d_0^{12} r^{18} \left(1 + \sum_{n=1}^{N} \beta_{6n} r^n\right).$$

Collecting the terms with like powers of r and comparing them with (29) results in relationships between the coefficients similar to (41).

Wecharatana [1990] introduced the softening relationship

$$\hat{\sigma}^m + \hat{w}^{2m} = 1 \tag{64}$$

(where $m = 0.27$ for concrete with compressive strength $f_c = 24$ MPa and $m = 0.2$ for concrete with $f_c = 83$ MPa) to fit his experimental results from uniaxial tests on normal and high strength concrete using dog-bone-shaped specimens with edge notches (see also [Karihaloo 1995]). This relationship cannot be used in the asymptotic analysis above as m is not an integer. However, for $m = 0.27$ in the range of $0 \leq \hat{w} \leq 0.6$, we can fit Equation (64) using (3) or (36) with $\alpha_1 = -6.9495$, $\alpha_2 = 29.9794$, $\alpha_3 = -87.2663$, $\alpha_4 = 148.3647$, and $\alpha_5 = -128.84$. The correlation coefficient is 1. When $\hat{w} = 0.6$, then $\hat{\sigma} = 0.005148$; when $\hat{w} > 0.6$, then $\hat{\sigma}$ is negligibly small. As compared in Figure 6, they cannot be distinguished on the scale of the figure.
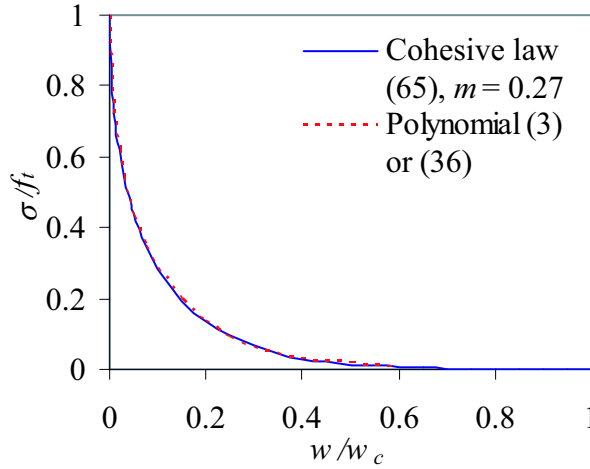
Figure 6. A comparison of cohesive law Equation (64) ($m = 0.27$) with polynomial Equation (3) or (36).

For $m = 0.2$ in the range of $0 \leq \hat{w} \leq 0.3$, Equation (64) can also be fitted using a polynomial in the form of (36) but including three higher order terms

$$\hat{\sigma}_y = 1 + \sum_{i=1}^{8} \alpha_i \hat{w}^{2/3i} - \left(1 + \sum_{i=1}^{8} \alpha_i\right) \hat{w}^6, \tag{65}$$

with

$$\alpha_1 = -25.77925, \qquad \alpha_2 = 459.7579,$$
$$\alpha_3 = -5.14083 \times 10^3, \qquad \alpha_4 = 3.51282 \times 10^4,$$
$$\alpha_5 = -1.48405 \times 10^5, \qquad \alpha_6 = 3.86621 \times 10^5,$$
$$\alpha_7 = -5.99122 \times 10^5, \qquad \alpha_8 = 4.99809 \times 10^5.$$

The correlation coefficient is also 1. When $\hat{w} = 0.3$, then $\hat{\sigma} = 0.008155$; when $\hat{w} > 0.3$, then $\hat{\sigma}$ is negligibly small. They are compared in Figure 7, and cannot be distinguished on the scale of the figure.

The simplest rectangular cohesive law in which the cohesive stress is constant and identical to the strength of the material f_t in the cohesive zone has also been used by some researchers. This law can be approximated by

$$\hat{\sigma}_y = 1 - \hat{w}^{2n} \tag{66}$$

as illustrated in Figure 8. Obviously, the cohesive law Equation (66) is a simplified form of (61) with coefficients $\alpha_i = 0$. The procedures and results for the cohesive law (61) are therefore correct for the rectangular law. Alternatively, the rectangular law can be enforced directly by assuming $\hat{\sigma}_y$ in (29) to be a constant, that is, all coefficients c_n $(n > 0)$ vanish and there are no further constraints on the coefficients.
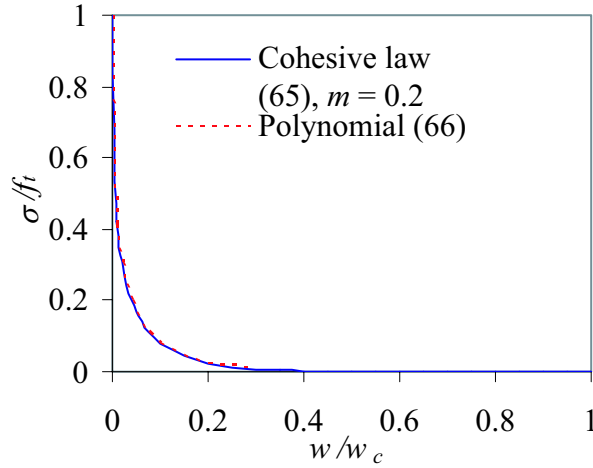
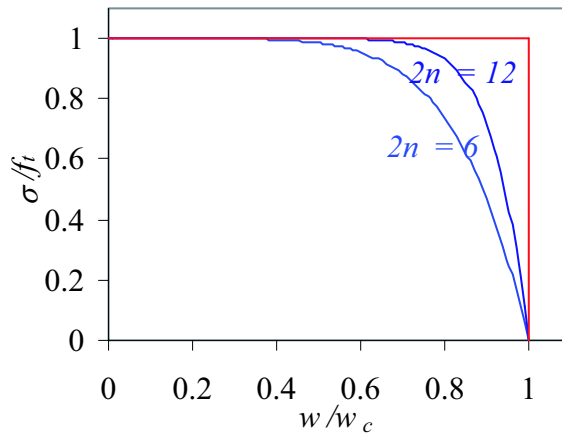**Figure 7.** A comparison of the cohesive law Equation (64) ($m = 0.2$) with polynomial (65).

Figure 8. Comparison of the rectangular cohesive law with (66).

6. Implementation of the asymptotic fields in XFEM/GFEM and examples of mode I cohesive crack tip fields

In the context of the implementation of the cohesive crack asymptotic fields as enrichment functions in the XFEM/GFEM, if not only the first term but also the higher order terms are used as in [Liu et al. 2004], the linear dependence of the coefficients can be enforced in advance, while the nonlinear dependence of the coefficients can be enforced as constraints in the solution process. It is more convenient to use only the leading term of the displacement asymptotic field at the tip of a cohesive crack (which ensures a displacement discontinuity normal to the cohesive crack face) as the enrichment function, as in most implementations of the XFEM in the literature. The complete implementation with several examples can be found in [Xiao et al. 2006, in press].
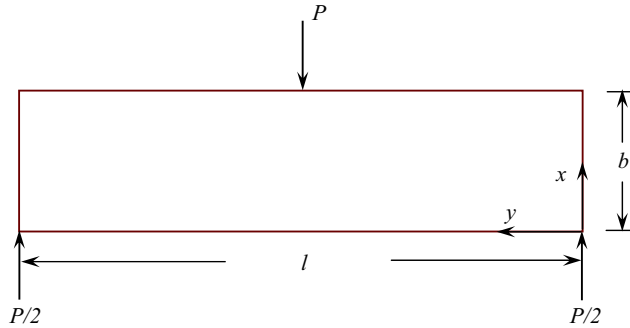
Figure 9. An unnotched three-point bend beam (TPB).

In the following, a typical mode I cohesive cracking problem of quasibrittle materials — a three point bend beam without any initial crack (Figure 9) made of a quasibrittle material with the linear softening law (57) — is analyzed. A state of plane strain is considered. The geometrical parameters are

$$b = 150 \text{ mm}, \qquad l = 4b, \qquad t = b,$$

where t is the specimen thickness in the out-of-plane direction. The material properties are

$$E = 36.5 \text{ GPa}, \qquad \nu = 0.1, \qquad f_t = 3.19 \text{ MPa}, \qquad G_F = 50 \text{ Nm}^{-1},$$

where E is Young's modulus, ν the Poisson ratio, and G_F the specific fracture energy. The dimensions for force and length are N and mm, respectively.

The details of simulation are the same as in [Xiao et al. 2006, in press]. Two meshes, as shown in Figure 10, are used in the analysis. The coarser mesh consists of $50 \times 100 = 5000$ rectangular elements, giving a total of 5151 nodes. The finer mesh consists of $150 \times 120 = 18000$ rectangular elements, giving a total of 18271 nodes. Both meshes are uniformly divided in the x-direction. For the coarser mesh, the central 50 layers of elements have an identical height (y-direction) of 3 mm; the remaining elements have an identical height of 9 mm. Therefore, elements in the central zone are $3 \times 3 \text{ mm}^2$ squares. For the finer mesh, the central 60 layers of elements have an identical height of 1 mm; the remaining elements have an identical height of 9 mm. Therefore elements in the central zone are $1 \times 1 \text{ mm}^2$ squares. The intention of using two meshes is to study the mesh size sensitivity of the global responses. The conventional 4-node bilinear isoparametric Q4 elements are also used as background elements. The first layer of nodes surrounding the cohesive crack tip (the elements that include the crack tip k are defined as the first layer elements of the crack tip with enriched nodes; the nodes in the first layer elements are called the first layer enriched nodes) are enriched with the first term of the asymptotic displacement field (50)–(54) at the tip of a cohesive crack corresponding to a noninteger eigenvalue that gives a normal displacement discontinuity over the cohesive-crack faces

$$u = \frac{r^{3/2}}{2\mu} a_{11} \left[\left(\kappa + \tfrac{1}{2} \right) \cos \tfrac{3}{2}\theta - \tfrac{3}{2} \cos \tfrac{1}{2}\theta \right], \tag{67}$$
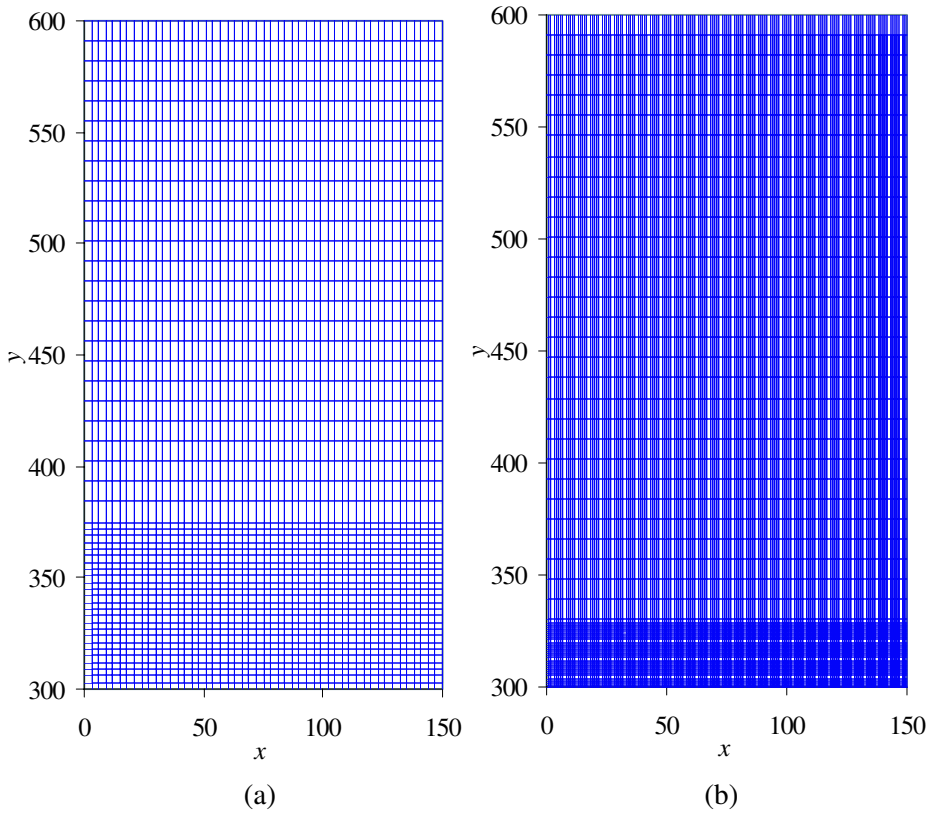
Figure 10. Coarse (a) and fine (b) mesh for half of the specimen.

$$v = \frac{r^{3/2}}{2\mu} a_{11} \left[\left(\kappa - \tfrac{1}{2} \right) \sin \tfrac{3}{2}\theta - \tfrac{3}{2} \sin \tfrac{1}{2}\theta \right]. \tag{68}$$

The potential fracture locus coincides with the specimen's axis of symmetry. The crack is modelled by enriching the nodes on the crack faces with jump and branch functions without the double nodes that are used in the traditional FEM.

As in [Moës and Belytschko 2002], the x-direction of nodes with coordinates $(0, 0)$ and $(0, 600\,\text{mm})$ and the y-direction of the node with coordinates $(150\,\text{mm}, 300\,\text{mm})$ are constrained; the load is distributed over a length of 6 mm in the coarse mesh (Figure 10a) and 2 mm (2 elements) in the fine mesh (10b). Since a low Poisson ratio of 0.1 is used, the results are believed to be close to those of Carpinteri and Colombo [1989], where a plane stress condition is assumed and a concentrated load was considered.

The nondimensional load-midspan deflection curves are shown in 11(a). They agree very well with the results of Carpinteri and Colombo [1989]. The evolution of the cohesive zone size as the cohesive tip travels through the beam is shown in Figure 11(b), which agrees very well with [Moës and Belytschko 2002]. It is clear that the results are insensitive to the mesh size.

In order to visualize the cohesive crack tip fields derived in Section 4, it is necessary to determine the unknown coefficients by fitting the numerically computed crack tip fields with the theoretically obtained fields. This requires a sophisticated optimization scheme. For the present purpose of illustration, we
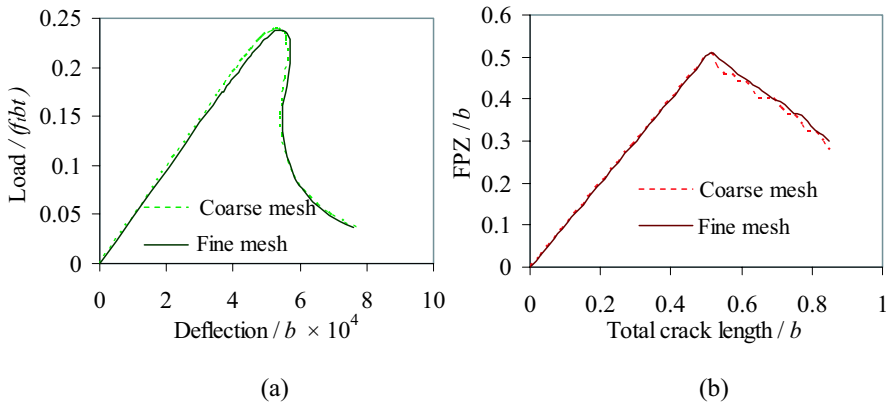
Figure 11. (a) Nondimensional load-midspan deflection curves of the three-point bend beam ($G_F = 50\,\mathrm{Nm}^{-1}$); (b) evolution of the cohesive zone size as the cohesive tip travels through the beam.

will solve a mode I cohesive crack problem shown in Figure 12 using the obtained asymptotic fields and the numerically computed opening profile of the cohesive crack. The dimensions of the displacement and length parameters are in millimeters and those of the stresses in megapascals, unless mentioned otherwise.

We consider the subdomain bounded by broken lines in Figure 12, and assume the length of the cohesive crack is 10.5 mm. The opening profile of the cohesive crack (Figure 13) adopts the computed results above at the loading stage with the total cohesive crack = 31.5 mm and load/($f_t bt$) = 0.228. It can be represented by the expansion Equation (55) corresponding to noninteger eigenvalues with nonvanishing coefficients

$$a_{10} = -0.192, \qquad a_{11} = -5.708 \cdot 10^{-3}, \qquad a_{12} = 1.2339 \cdot 10^{-5}.$$

From these coefficients, we can obtain parameters β Equation (40) and c (41). The cohesive stress (29) corresponding to these coefficients c is compared in Figure 14 with the results obtained by the linear tension-softening relationship Equation (57). The agreement is excellent, with a maximum error less than 0.5%.
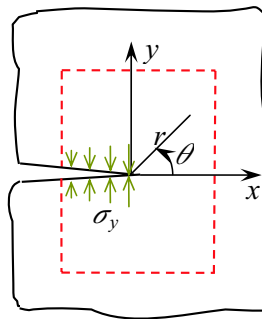

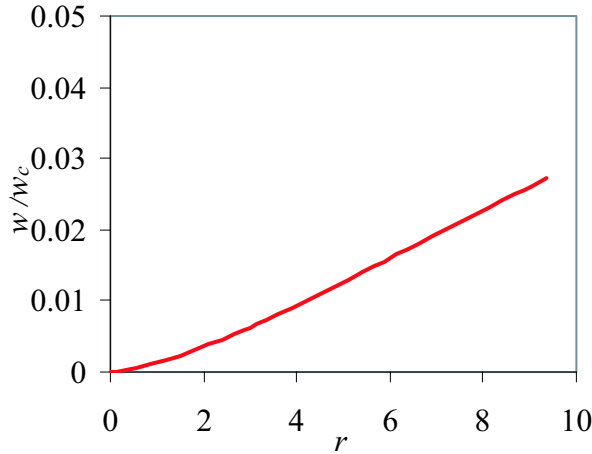**Figure 12.** Illustration of the cohesive crack problem.
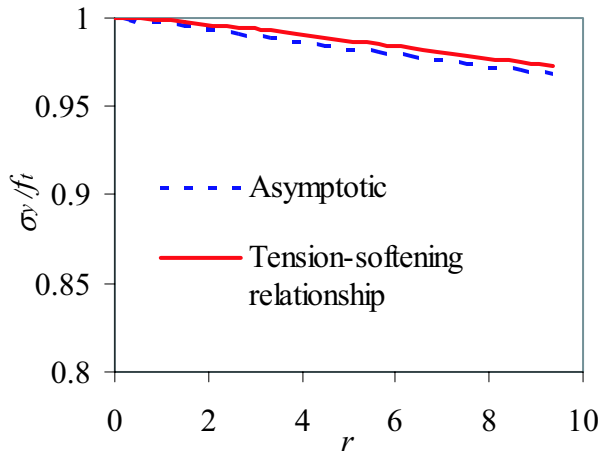
Figure 13. Opening profile of the cohesive crack.

Figure 14. Comparison of the cohesive stress obtained by the tension-softening relationship Equation (57) and the asymptotic analysis.

The c coefficients also provide relationships between coefficients a_{1n} and b_{1n} for each integer eigenvalue as in Equation (30). In other words, the complete asymptotic fields are known except the coefficients a or b corresponding to integer eigenvalues. They are determined from the applied boundary conditions. Since we have assumed the crack opening profile, we cannot assume the prescribed boundary conditions again to avoid inconsistency. Therefore some weaker constraints are used instead.

We assume the boundary $x = -10.5$ is, or is very close to, a traction-free surface. Then the coefficients a_{1n} or b_{1n} for integer eigenvalues can be determined by assuming the boundary $x = -10.5$ to be nearly traction-free. More precisely, we divide the segment $x = -10.5$, $0 \le y \le 10.5$ into 100 identical segments, and minimize the value of σ_x and τ_{xy} at these 101 locations (including corner nodes) using the *Minimize* function of Mathcad 11. The convergence and constraint tolerances are chosen as 10^{-8}. The values
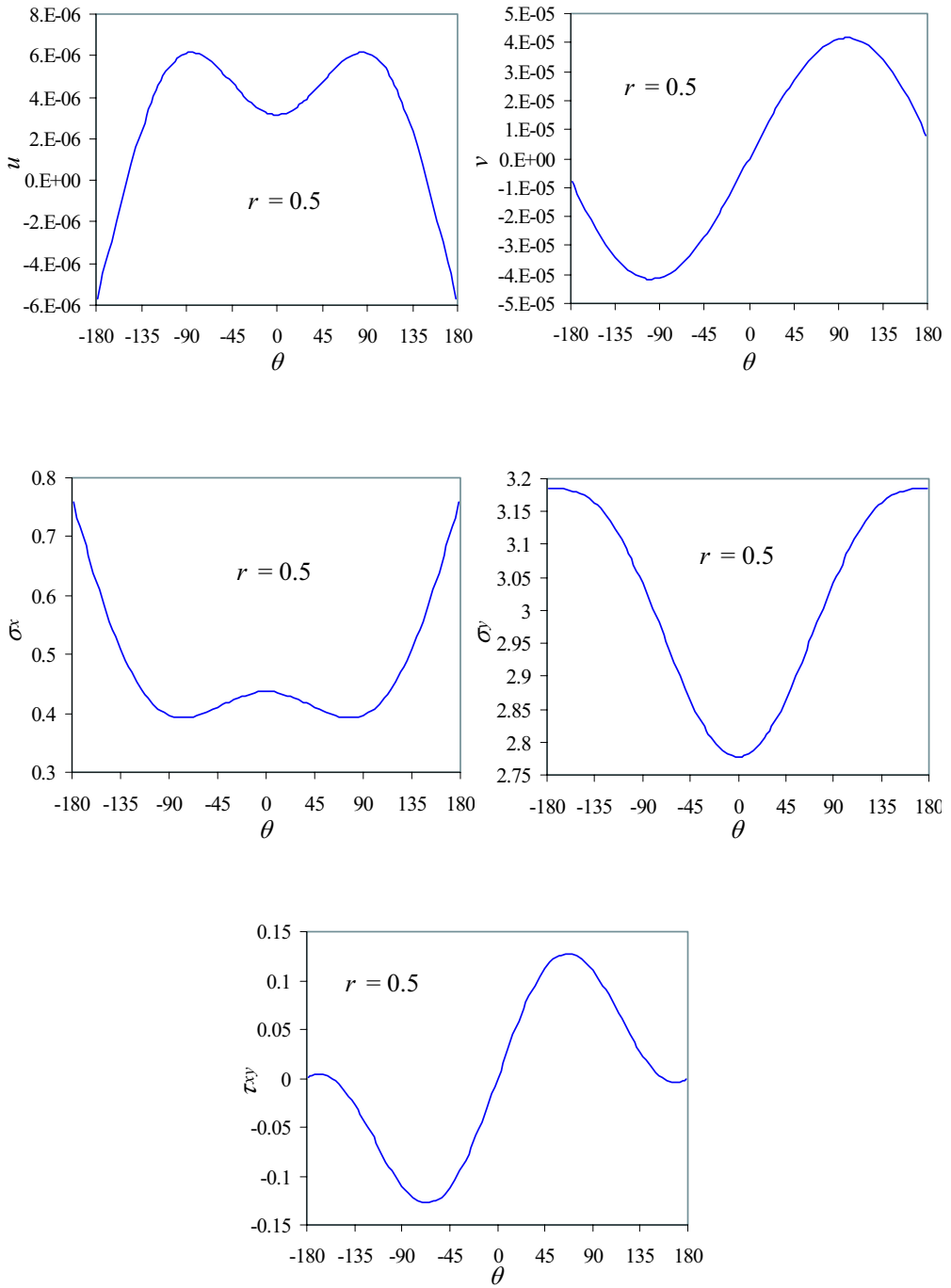
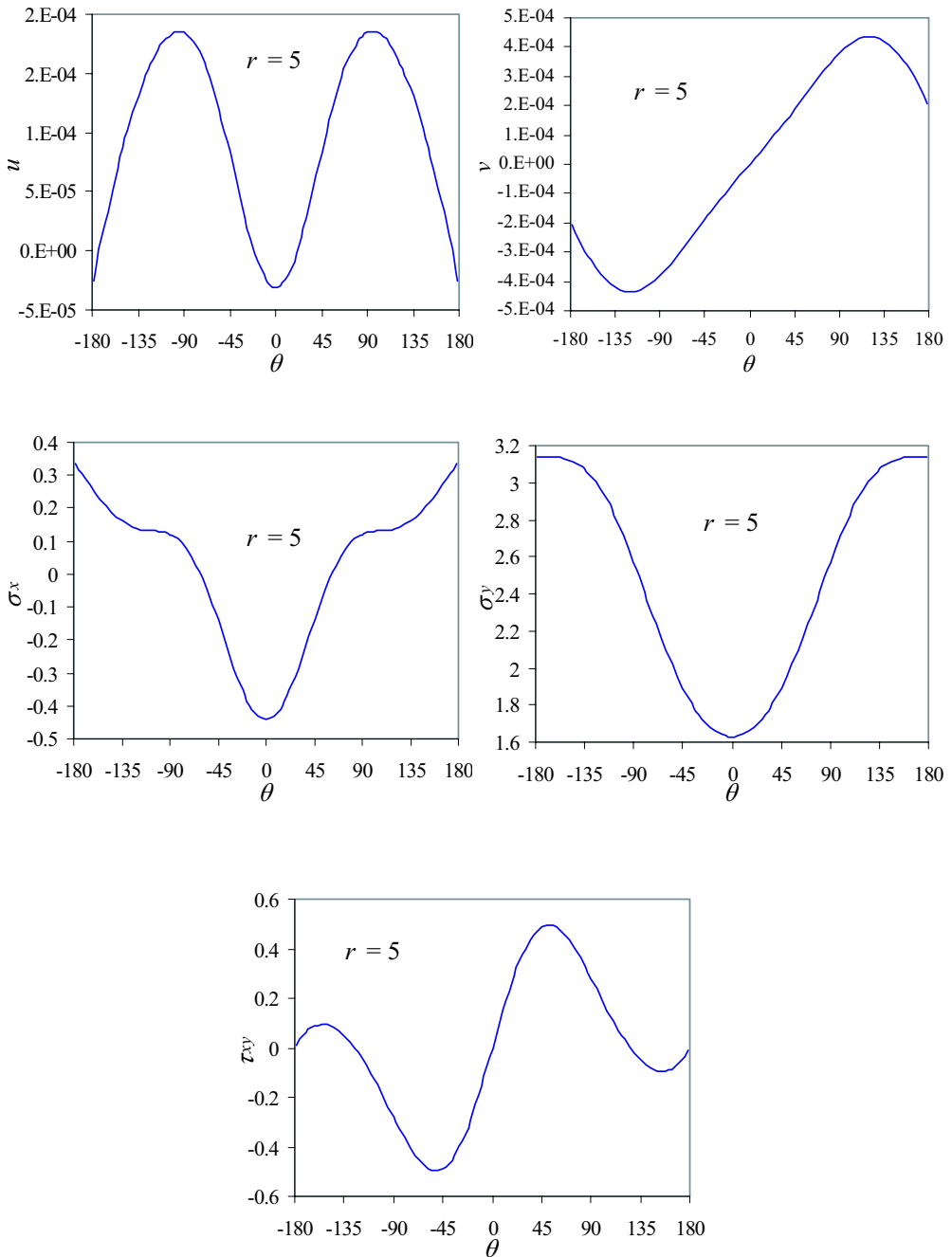**Figure 15.** Displacements and stresses along the circle $r = 0.5$.

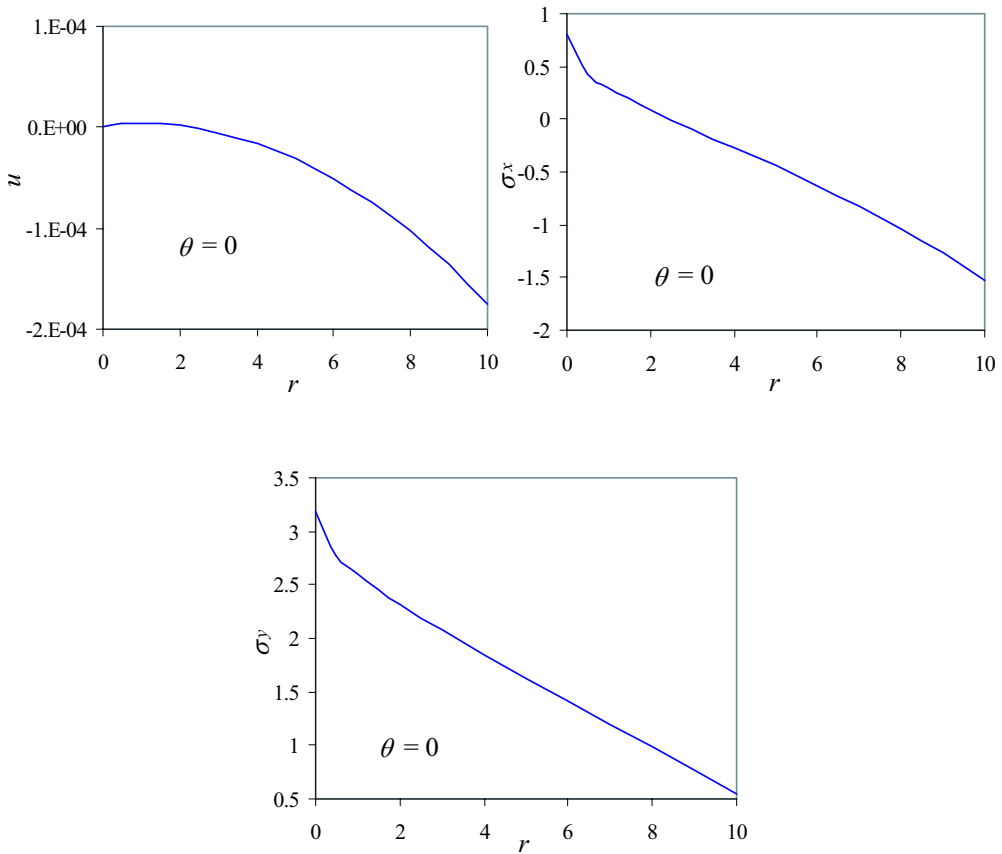**Figure 16.** Displacements and stresses along the circle $r = 5$.

Figure 17. Nonvanishing displacements and stresses along the line of extension of the cohesive crack ($\theta = 0$).

$a_{10} = 1$, $a_{11} = 10^{-3}$, $a_{12} = -10^{-5}$, and $a_{13} = -10^{-5}$ are chosen as initial guesses. We obtain the coefficients for integer eigenvalues as

$$a_{10} = 0.9604, \quad a_{11} = 0.0124, \quad a_{12} = -2.1858 \cdot 10^{-4}, \quad a_{13} = -1.9741 \cdot 10^{-5},$$
$$b_{10} = 0.595, \quad b_{11} = -0.0107, \quad b_{12} = 1.9553 \cdot 10^{-4}, \quad b_{13} = 1.8647 \cdot 10^{-5},$$

with very small σ_x and τ_{xy} on the segment $x = -10.5$, $0 \le y \le 10.5$. Coefficients of the higher order terms are negligibly small.

The displacements and stresses corresponding to these coefficients along two circles surrounding the crack tip and the line of extension of the cohesive crack are plotted in Figures 15–17. Away from the crack tip, the distribution of the displacement and/or stress may be quite different. As expected, the stress σ_y at the cohesive crack tip is equal to f_t, and no stress at any other locations reaches f_t.

These solutions are exact when the displacements or tractions corresponding to them are applied on the boundary, as in the broken lines shown in Figure 12.

Near the cohesive crack tip, the maximum values of the main stresses σ_x and σ_y occur on the cohesive crack faces. This is because we have chosen the cohesive crack opening profile (Figure 13) corresponding to a TPB test. This feature of the stress distribution for this type of specimen has also been noticed by Planas et al. [2003] and validated by our own numerical analysis.

7. Conclusions and discussion

The asymptotic fields obtained in Section 4 are universal for the prescribed normal cohesion-separation relation. However, relationships Equation (41) between the coefficients are dependent on the softening relation of the material. The actual values of the coefficients are dependent on the geometry and boundary/loading conditions of the problem, as in the Williams expansion of a traction-free crack. The cohesive crack asymptotic fields obtained here can be used as enrichment functions in the extended/generalized finite element method at the tip of long cohesive cracks, as well as short branches/kinks.

In traction-free cracks, terms in Equations (19)–(23) corresponding to different eigenvalues are independent, that is, controlled by independent coefficients. In the cohesive crack tip fields some of the terms can be dependent and not controlled by independent coefficients. Such a dependence also exists in the higher order terms of the crack tip fields in elastoplastic power-law hardening materials (see [Xia et al. 1993; Yang et al. 1993; 1996; Chao and Yang 1996] and a review by Karihaloo and Xiao [2003a]).

In this paper, we have applied a complex-function formulation of homogeneous isotropic linear elasticity for simplicity; however, the derivation can be easily extended to bimaterial interfacial cracks (see [Rice 1988]) and/or anisotropic elasticity. Anisotropy can be treated with the Stroh formalism [Stroh 1958; Suo 1990; Deng 1994].

Appendix: Derivatives of $f_i(r)$

The first five derivatives of $f_i(r)$ in Equation (39) are

$$f_i^{(1)}(r) = \tfrac{2}{3}i\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-1} \sum_{n=1}^{N} d_n n r^{n-1};$$

$$f_i^{(2)}(r) = \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-2}\left(\sum_{n=1}^{N} d_n n r^{n-1}\right)^2 + \tfrac{2}{3}i\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-1} \sum_{n=2}^{N} d_n n(n-1)r^{n-2};$$

$$\begin{aligned}
f_i^{(3)}(r) &= \tfrac{2}{3}i\left(\tfrac{2}{3}i - 1\right)\left(\tfrac{2}{3}i - 2\right)\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-3}\left(\sum_{n=1}^{N} d_n n r^{n-1}\right)^3 \\
&\quad + \tfrac{6}{3}i\left(\tfrac{2}{3}i - 1\right)\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-2} \sum_{n=1}^{N} d_n n r^{n-1} \sum_{n=2}^{N} d_n n(n-1)r^{n-2} \\
&\quad + \tfrac{2}{3}i\left(1 + \sum_{n=1}^{N} d_n r^n\right)^{\frac{2}{3}i-1} \sum_{n=3}^{N} d_n n(n-1)(n-2)r^{n-3};
\end{aligned}$$

$$f_i^{(4)}(r) = \tfrac{2}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(\tfrac{2}{3}i-3\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-4}\left(\sum_{n=1}^{N}d_n n r^{n-1}\right)^4$$

$$+\tfrac{12}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-3}\left(\sum_{n=1}^{N}d_n n r^{n-1}\right)^2\sum_{n=2}^{N}d_n n(n-1)r^{n-2}$$

$$+\tfrac{6}{3}i\left(\tfrac{2}{3}i-1\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-2}\left(\sum_{n=2}^{N}d_n n(n-1)r^{n-2}\right)^2$$

$$+\tfrac{8}{3}i\left(\tfrac{2}{3}i-1\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-2}\sum_{n=1}^{N}d_n n r^{n-1}\sum_{n=3}^{N}d_n n(n-1)(n-2)r^{n-3}$$

$$+\tfrac{2}{3}i\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-1}\sum_{n=4}^{N}d_n n(n-1)(n-2)(n-3)r^{n-4};$$

$$f_i^{(5)}(r) = \tfrac{2}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(\tfrac{2}{3}i-3\right)\left(\tfrac{2}{3}i-4\right)\times\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-5}\left(\sum_{n=1}^{N}d_n n r^{n-1}\right)^5$$

$$+\tfrac{20}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(\tfrac{2}{3}i-3\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-4}\times\left(\sum_{n=1}^{N}d_n n r^{n-1}\right)^3\sum_{n=2}^{N}d_n n(n-1)r^{n-2}$$

$$+\tfrac{30}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-3}\sum_{n=1}^{N}d_n n r^{n-1}\left(\sum_{n=2}^{N}d_n n(n-1)r^{n-2}\right)^2$$

$$+\tfrac{20}{3}i\left(\tfrac{2}{3}i-1\right)\left(\tfrac{2}{3}i-2\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-3}\left(\sum_{n=1}^{N}d_n n r^{n-1}\right)^2\times\sum_{n=3}^{N}d_n n(n-1)(n-2)r^{n-3}$$

$$+\tfrac{20}{3}i\left(\tfrac{2}{3}i-1\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-2}\sum_{n=2}^{N}d_n n(n-1)r^{n-2}\sum_{n=3}^{N}d_n n(n-1)(n-2)r^{n-3}$$

$$+\tfrac{10}{3}i\left(\tfrac{2}{3}i-1\right)\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-2}\sum_{n=1}^{N}d_n n r^{n-1}\sum_{n=4}^{N}d_n n(n-1)(n-2)(n-3)r^{n-4}$$

$$+\tfrac{2}{3}i\left(1+\sum_{n=1}^{N}d_n r^n\right)^{\frac{2}{3}i-1}\sum_{n=5}^{N}d_n n(n-1)(n-2)(n-3)(n-4)r^{n-5}.$$

Acknowledgment

References

[Abdalla and Karihaloo 2004] H. M. Abdalla and B. L. Karihaloo, "A method for constructing the bilinear tension softening diagram of concrete corresponding to its true fracture energy", *Mag. Concr. Res.* **56**:10 (2004), 597–604.

[Babuška et al. 2003] I. Babuška, U. Banerjee, and J. E. Osborn, "Survey of meshless and generalized finite element methods: A unified approach", *Acta Numerica* **12** (2003), 1–125.

[Barenblatt 1962] G. I. Barenblatt, "The mathematical theory of equilibrium cracks in brittle fracture", *Adv. Appl. Mech.* **7** (1962), 55–129.

[Bialas and Mróz 2005] M. Bialas and Z. Mróz, "Modelling of progressive interface failure under combined normal compression and shear stress", *Int. J. Solids Struct.* **42**:15 (2005), 4436–4467.

[de Borst et al. 2004] R. de Borst, M. A. Gutierrez, G. N. Wells, J. J. C. Remmers, and H. Askes, "Cohesive-zone models, higher-order continuum theories and reliability methods for computational failure analysis", *Int. J. Numer. Methods Eng.* **60**:1 (2004), 289–315.

[Carpinteri and Colombo 1989] A. Carpinteri and G. Colombo, "Numerical analysis of catastrophic softening behaviour (snapback instability)", *Comput. Struct.* **31**:4 (1989), 607–636.

[Chandra et al. 2002] N. Chandra, H. Li, C. Shet, and H. Ghonem, "Some issues in the application of cohesive zone models for metal-ceramic interfaces", *Int. J. Solids Struct.* **39**:10 (2002), 2827–2855.

[Chao and Yang 1996] Y. J. Chao and S. Yang, "Higher order crack tip fields and its implication for fracture of solids under mode II conditions", *Eng. Fract. Mech.* **55**:5 (1996), 777–794.

[Cocchetti et al. 2002] G. Cocchetti, G. Maier, and X. P. Shen, "Piecewise linear models for interfaces and mixed mode cohesive cracks", *Comput. Model. Eng. Sci.* **3**:3 (2002), 279–298.

[Cornelissen et al. 1986] H. A. W. Cornelissen, D. A. Hordijk, and H. W. Reinhardt, "Experimental determination of crack softening characteristics of normalweight and lightweight concrete", *Heron* **31**:2 (1986), 45–56.

[Deng 1994] X. M. Deng, "An asymptotic analysis of stationary and moving cracks with frictional contact along bimaterial interfaces and in homogeneous solids", *Int. J. Solids Struct.* **31**:17 (1994), 2407–2429.

[Dugdale 1960] D. S. Dugdale, "Yielding of steel sheets containing slits", *J. Mech. Phys. Solids* **8**:2 (1960), 100–104.

[Elices et al. 2002] M. Elices, G. V. Guinea, J. Gómez, and J. Planas, "The cohesive zone model: advantages, limitations and challenges", *Eng. Fract. Mech.* **69**:2 (2002), 137–163.

[Hansbo and Hansbo 2004] A. Hansbo and P. Hansbo, "A finite element method for the simulation of strong and weak discontinuities in solid mechanics", *Comput. Methods Appl. Mech. Eng.* **193**:33–35 (2004), 3523–3540.

[Hillerborg et al. 1976] A. Hillerborg, M. Modéer, and P. E. Petersson, "Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements", *Cem. Concr. Res.* **6**:6 (1976), 773–782.

[Hong and Kim 2003] S. S. Hong and K. S. Kim, "Extraction of cohesive-zone laws from elastic far-fields of a cohesive crack tip: a field projection method", *J. Mech. Phys. Solids* **51**:7 (2003), 1267–1286.

[Hutchinson and Evans 2000] J. W. Hutchinson and A. G. Evans, "Mechanics of materials: top-down approaches to fracture", *Acta Mater.* **48**:1 (2000), 125–135.

[Karihaloo 1995] B. L. Karihaloo, *Fracture mechanics and structural concrete*, Addison Wesley Longman, UK, 1995.

[Karihaloo 1999] B. L. Karihaloo, "Size effect in shallow and deep notched quasi-brittle structures", *Int. J. Fract.* **95**:1-4 (1999), 379–390.

[Karihaloo and Xiao 2003a] B. L. Karihaloo and Q. Z. Xiao, "Linear and nonlinear fracture mechanics", pp. 81–212 in *Comprehensive structural integrity, II: fundamental theories and mechanisms of failure*, vol. 2, edited by B. L. Karihaloo and W. G. Knauss, Elsevier Pergamon, UK, 2003.

[Karihaloo and Xiao 2003b] B. L. Karihaloo and Q. Z. Xiao, "Modelling of stationary and growing cracks in FE framework without remeshing: a state-of-the-art review", *Comput. Struct.* **81**:3 (2003), 119–129.

[Karihaloo et al. 2003] B. L. Karihaloo, H. M. Abdalla, and T. Imjai, "A simple method for determining the true specific fracture energy of concrete", *Mag. Concr. Res.* **55**:5 (2003), 471–481.

[Leblond and Frelat 2004] J. B. Leblond and J. L. Frelat, "Crack kinking from an initially closed, ordinary or interface crack, in the presence of friction", *Eng. Fract. Mech.* **71**:3 (2004), 289–307.

[Liu et al. 2004] X. Y. Liu, Q. Z. Xiao, and B. L. Karihaloo, "XFEM for direct evaluation of mixed mode SIFs in homogeneous and bi-materials", *Int. J. Numer. Methods Eng.* **59**:8 (2004), 1103–1118.

[Moës and Belytschko 2002] N. Moës and T. Belytschko, "Extended finite element method for cohesive crack growth", *Eng. Fract. Mech.* **69**:7 (2002), 813–833.

[Moës et al. 1999] N. Moës, J. Dolbow, and T. Belytschko, "A finite element method for crack growth without remeshing", *Int. J. Numer. Methods Eng.* **46**:1 (1999), 131–150.

[Mróz and Bialas 2005] Z. Mróz and M. Bialas, "A simplified analysis of interface failure under compressive normal stress and monotonic or cyclic shear loading", *Int. J. Numer. Anal. Methods Geomech.* **29**:4 (2005), 337–368.

[Muskhelishvili 1953] N. I. Muskhelishvili, *Some basic problems of mathematical theory of elasticity*, Noordhoff, Holland, 1953.

[Owen and Fawkes 1983] D. R. J. Owen and A. J. Fawkes, *Engineering fracture mechanics: numerical methods and applications*, Pineridge Press Ltd., Swansea, UK, 1983.

[Planas et al. 2001] J. Planas, Z. P. Bazant, and M. Jirasek, "Reinterpretation of Karihaloo's size effect analysis for notched quasibrittle structures", *Int. J. Fract.* **111**:1 (2001), 17–28.

[Planas et al. 2003] J. Planas, M. Elices, G. V. Guinea, F. J. Gómez, D. A. Cendón, and I. Arbilla, "Generalizations and specializations of cohesive crack models", *Eng. Fract. Mech.* **70**:14 (2003), 1759–1776.

[Remmers et al. 2003] J. J. C. Remmers, R. de Borst, and A. Needleman, "A cohesive segments method for the simulation of crack growth", *Comput. Mech.* **31**:1-2 (2003), 69–77.

[Rice 1988] J. R. Rice, "Elastic fracture mechanics concepts for interfacial cracks", *J. Appl. Mech. (Trans. ASME)* **55** (1988), 98–103.

[Sih and Liebowitz 1968] G. C. Sih and H. Liebowitz, *Mathematical theories of brittle fracture*, vol. II, Chapter Fracture: an advanced treatise, pp. 67–190, Academic Press, New York, 1968.

[Stroh 1958] A. N. Stroh, "Dislocations and cracks in anisotropic elasticity", *Philos. Mag. (8)* **3** (1958), 625–646.

[Strouboulis et al. 2001] T. Strouboulis, K. Copps, and I. Babuška, "The generalized finite element method", *Comput. Methods Appl. Mech. Eng.* **190**:32–33 (2001), 4081–4193.

[Suo 1990] Z. Suo, "Singularities, interfaces and cracks in dissimilar anisotropic media", *Proc. Roy. Soc. London Ser. A* **427**:1873 (1990), 331–358.

[Valente 1991] S. Valente, "Influence of friction on cohesive crack-propagation", pp. 695–704 in *Fracture processes in concrete, rock and ceramics*, edited by J. G. M. van Mier et al., RILEM Proceedings **13**, London, 1991.

[Wecharatana 1990] M. Wecharatana, "Brittleness index of cementitious composite", in *Serviceability and durability of construction materials*, edited by B. A. Suprenant, ASCE Publications, New York, 1990.

[Wells and Sluys 2001] G. N. Wells and L. J. Sluys, "A new method for modeling cohesive cracks using finite elements", *Int. J. Numer. Methods Eng.* **50**:12 (2001), 2667–2682.

[Williams 1957] M. L. Williams, "On the stress distribution at the base of a stationary crack", *J. Appl. Mech. (Trans. ASME)* **24** (1957), 109–114.

[Xia et al. 1993] L. Xia, T. C. Wang, and C. F. Shih, "Higher-order analysis of crack tip fields in elastic power-law hardening materials", *J. Mech. Phys. Solids* **41**:4 (1993), 665–687.

[Xiao and Karihaloo 2005] Q. Z. Xiao and B. L. Karihaloo, "Recent developments of the extended/generalized FEM and a comparison with the FEM", pp. 303–324 in *Developments and applications of solid mechanics, Proceedings of the Symposium for Professor M. G. Huang's 90th Birthday*, edited by X. P. Wu, Press of University of Science and Technology of China, China, 2005.

[Xiao et al. 2006, in press] Q. Z. Xiao, B. L. Karihaloo, and X. Y. Liu, "Incremental-secant modulus iteration scheme and stress recovery for simulating cracking process in quasi-brittle materials using XFEM", *Int. J. Numer. Mech. Eng.* (2006, in press).

[Yang et al. 1993] S. Yang, Y. J. Chao, and M. A. Sutton, "Higher order asymptotic crack tip fields in a power-law hardening material", *Eng. Fract. Mech.* **45**:1 (1993), 1–20.

[Yang et al. 1996] S. Yang, F. G. Yuan, and X. Cai, "Higher order asymptotic elastic-plastic crack-tip fields under antiplane shear", *Eng. Fract. Mech.* **54**:3 (1996), 405–422.

QIZHI XIAO: xiaoq@cardiff.ac.uk
School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, United Kingdom

BHUSHAN LAL KARIHALOO: karihaloob@cardiff.ac.uk
School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, United Kingdom

ANALYSIS OF ELECTROMECHANICAL BUCKLING OF A PRESTRESSED MICROBEAM THAT IS BONDED TO AN ELASTIC FOUNDATION

DAVID ELATA AND SAMY ABU-SALIH

The electromechanical buckling of a prestressed microbeam bonded to a dielectric elastic foundation is analyzed. It is shown that electrostatic forces can precipitately instigate buckling even when the prestress in the microbeam is lower than the critical value that would cause mechanical buckling. We show that electrostatic potential can be used to achieve on/off switching of surface flexures. An analytic solution of the critical electromechanical state is derived. In addition, an analytic approximation of the initial postbuckling state is also presented, and is validated numerically.

1. Introduction

Mechanical buckling is a well known phenomenon that occurs in thin elastic structures subjected to compressive loads. Mechanical buckling develops only if the compressive loads are larger than a critical value. In most buckled structures, reduction of the compressive load to a subcritical level will eliminate the buckling deformation. In thin sheet-like elastic solids that are bonded to an elastic foundation, a compressive stress can cause a dense occurrence of buckling flexures [Hetenyi 1946]. The buckling phenomenon is a *bifurcation* transition because the postbuckling state may arbitrarily develop into one of several modes [Gilmore 1981; Godoy 1999; Nguyen 2000].

A different instability that is prevalent in the field of microelectromechanical systems (MEMS) is the pull-in phenomenon in electrostatic actuators [Elata et al. 2003; Pelesko and Bernstein 2003]. This inherent instability is due to the nonlinear nature of electrostatic forces. This instability is known as a *limit point* or a *fold* in the equilibrium state of the electrostatic actuator [Elata et al. 2003; Gilmore 1981; Godoy 1999; Nguyen 2000], and is not a bifurcation.

However, a bifurcation transition is also possible in electrostatic actuators. A well known example is the *side pull-in* in electrostatic comb-drives [Elata et al. 2003; Elata and Leus 2005; Legtenberg et al. 1996]. This instability response is an electromechanical bifurcation because the comb-drive may collapse in more than one direction.

Recently, the bifurcation response of a clamped-clamped beam that is subjected to both compressive stress and an electrostatic field was investigated. This bifurcation instability was termed *electromechanical buckling* (EMB) [Elata and Abu-Salih 2005], because it is a true coupling between mechanical buckling and electromechanical bifurcation.

In the present study, the EMB response of a prestressed infinite beam, bonded to an elastic foundation, is analyzed. Specifically it is shown that buckling in this system can be reversibly switched on and off by application and elimination of a driving voltage.

In the next section the governing equation of the problem is presented. In Section 3 the parameters of the critical electromechanical buckling state are analytically derived. The initial postbuckling analysis is presented in Section 4 and includes an approximate analytic solution which is numerically validated.

2. Formulation

In this study we analyze the EMB of an infinite beam. To facilitate the analysis we use the following strategy. First, we assume that the critical and postbuckling states are both periodic. This assumption will be later confirmed by numerical simulations. Second, we derive the wavelength of the periodic solution. For the critical state the wavelength is derived analytically. For the postbuckled state the wavelength is computed in the following way: we first consider the postbuckled state of a finite beam with periodic boundary conditions. The length of the finite beam which is associated with the minimal energy per unit length of the postbuckled state is the correct postbuckling wavelength in the infinite beam. Finally, we use numerical simulations to confirm that the critical and postbuckling solutions indeed converge to a periodic waveform.

Consider an infinite beam which is bonded to an elastic foundation as illustrated in Figure 1a. The beam is conductive, the elastic foundation is an isolating dielectric, and the bottom substrate is a fixed rigid conductor. The beam is prestressed and is subjected to a potential V, and the substrate is grounded.

The field equation that governs the electromechanical response of the beam is given by

$$Db\frac{d^4y}{dx^4} - \sigma bh\frac{d^2y}{dx^2} - Ebh\left(\frac{1}{L}\int_L \frac{1}{2}\left(\frac{dy}{dx}\right)^2 dx\right)\frac{d^2y}{dx^2} + k_f by = \frac{\varepsilon b V^2}{2(g-y)^2}. \tag{1}$$

The four terms on the left hand side of Equation (1) are the distributed load due to bending, residual stress, membrane stiffening, and elastic foundation. The term on the right hand side is the electrostatic distributed force.

Here y is the deflection as a function of location x along the beam, $D = E^*h^3/12$ is the bending rigidity of the beam, b is the beam width, h is the beam thickness, and $E^* = E/(1-\nu^2)$ is the effective bending modulus assuming that the beam is wide, that is, $b \gg h$, where E is Young's modulus and ν is the Poisson ratio. The variable σ is the prestress (positive in tension), L is the length of the beam, k_f is the elastic modulus of the foundation measured in N/m^3, ε is the permittivity of the dielectric elastic foundation and g is its nominal thickness.

The third term in (1), which includes the beam length L, is correct for $L \to \infty$. The equilibrium equation also holds for a beam of finite length L with periodic boundary conditions. In this case, if L is equal to an integer number of wavelengths of the periodic solution of the infinite beam, then the solution of the finite beam is identical to the solution of the infinite beam.

In the third term of (1), it is assumed that the elastic foundation does not constrain the transverse strain ε_{zz} when the beam is in tension. This may be expected when the thickness of the elastic foundation is greater than the beam width. If, however, the thickness of the elastic foundation is much smaller than the beam width, then the elastic foundation may constrain the transverse strain ε_{zz} when the beam is subjected to axial tension. In this case the effective elastic modulus in the third term should be replaced with E^*.
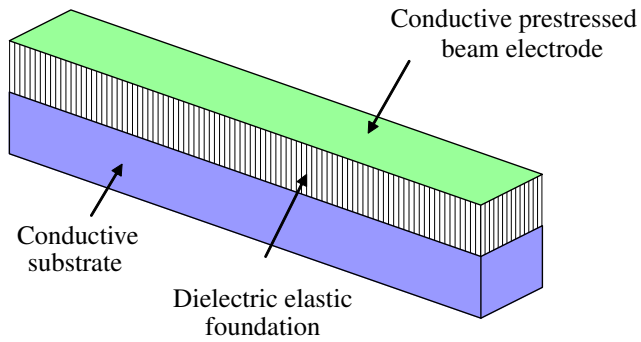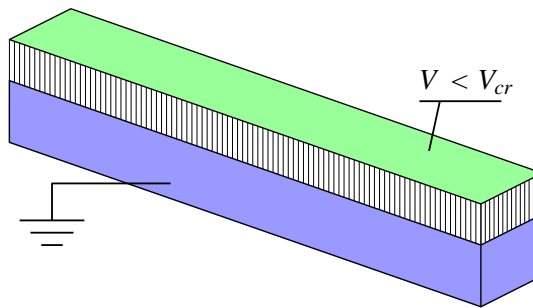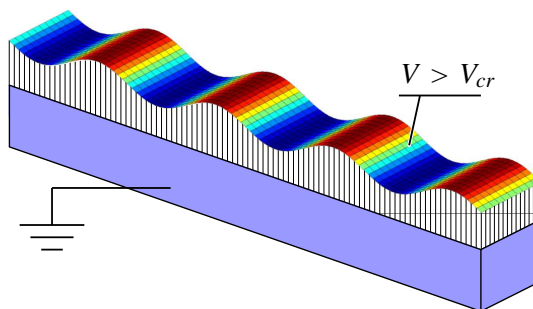
Figure 1. A prestressed beam bonded to an elastic foundation. (a) When the prestress in the conductive beam is below the critical value, and no voltage is applied, the beam remains flat; (b) when a subcritical voltage is applied, the beam deflection is uniform; (c) when a supercritical voltage is applied, electromechanical buckling occurs.

The effect of fringing fields is not considered in this work though simplified approximations of this effect may be added to the analysis [Leus and Elata 2004]. Also, the equilibrium equation (1) is valid for small rotations, that is, $|dy/dx| \ll 1$. If this condition does not hold, additional nonlinear terms must be included in the governing equation [Brush and Almaroth 1975].

The electrostatic distributed force is approximated by the local parallel-plates model which is valid for small rotations.

The governing equation may be written in the following normalized form

$$\frac{1}{(2\pi)^4}\frac{d^4\tilde{y}}{d\tilde{x}^4} + \frac{2\beta}{(2\pi)^2}\frac{d^2\tilde{y}}{d\tilde{x}^2} - \frac{\tilde{g}^2}{\alpha}\left[\int_0^{\alpha}\frac{1}{2}\left(\frac{d\tilde{y}}{d\tilde{x}}\right)^2 d\tilde{x}\right]\frac{1}{(2\pi)^4}\frac{d^2\tilde{y}}{d\tilde{x}^2} + \tilde{y} = \frac{\tilde{V}^2}{(1-\tilde{y})^2}, \tag{2}$$

where

$$\tilde{x} = \frac{x}{\Lambda_{cr}}, \qquad \tilde{y} = \frac{y}{g}, \qquad \beta = \frac{\sigma}{\sigma_{cr}}, \qquad \sigma_{cr} = -\frac{2\sqrt{k_f D}}{h},$$

$$\tilde{g} = \sqrt{12\frac{E}{E^*}}\frac{g}{h}, \qquad \alpha = \frac{L}{\Lambda_{cr}}, \qquad \tilde{V}^2 = \frac{\varepsilon V^2}{2k_f g^3}, \qquad \Lambda_{cr} = 2\pi\left(\frac{D}{k_f}\right)^{1/4}.$$

Here Λ_{cr} and σ_{cr} are the wavelength and the stress at the critical buckling state when no voltage is applied [Abu-Salih and Elata 2005; Hetenyi 1946], β is the normalized prestress parameter and α is the normalized wavelength [Abu-Salih and Elata 2005].

3. Critical state

When no electrostatic forces are applied, the prestressed beam will not buckle if the prestress is lower than the critical stress σ_{cr}. In this section we investigate the effect of electrostatic forces on the critical buckling state of the prestressed beam. Specifically, we show that electrostatic forces can induce buckling in a beam in which the prestress is subcritical (in such a beam buckling will not occur if electrostatic forces are not applied).

When a voltage is applied to the beam, the beam initially deflects uniformly, similar to a parallel-plates actuator [Pelesko and Bernstein 2003] (Figure 1b). In this case the dielectric elastic foundation constitutes both a dielectric substance between the electrodes and the elastic spring of the parallel-plates actuator.

When the deflection is nearly uniform, the nonlinear term (in square brackets) may be omitted from the governing equation (2), which reduces to

$$\frac{1}{(2\pi)^4}\frac{d^4\tilde{y}}{d\tilde{x}^4} + \frac{2\beta}{(2\pi)^2}\frac{d^2\tilde{y}}{d\tilde{x}^2} + \tilde{y} = \frac{\tilde{V}^2}{(1-\tilde{y})^2}. \tag{3}$$

The deflection of the infinite beam is assumed to be periodic and is postulated in the form

$$\tilde{y} = \tilde{y}_0 + B\sin\frac{2\pi\tilde{x}}{\alpha}, \tag{4}$$

where $\tilde{y}_0$ is an average value which is equal to the uniform displacement when no buckling occurs, and B is the amplitude of the structural waves that develop due to the electromechanical buckling response.

At the verge of buckling, B is small, and therefore the electrostatic force may be approximated by the following Taylor expansion

$$\frac{\tilde{V}^2}{(1-\tilde{y})^2} \approx \tilde{V}^2 \left[\frac{1}{(1-\tilde{y}_0)^2} + \frac{2}{(1-\tilde{y}_0)^3} B \sin \frac{2\pi \tilde{x}}{\alpha} + O(B^2) \right]. \tag{5}$$

Now, substituting the postulated deflection (4) and the approximated electrostatic force (5) into the linear governing equation (3), yields

$$B \frac{\sin(2\pi \tilde{x}/\alpha)}{\alpha^4} \left(\alpha^4(1-\delta) - 2\alpha^2\beta + 1 \right) = \frac{\tilde{V}^2}{(1-\tilde{y}_0)^2} - \tilde{y}_0. \tag{6}$$

Here δ is the normalized electrostatic stiffness that is given by

$$\delta = \frac{2\tilde{V}^2}{(1-\tilde{y}_0)^3}. \tag{7}$$

On the verge of buckling, where $B = 0$, the solution of (6) is

$$\tilde{V}^2 = \tilde{y}_0(1-\tilde{y}_0)^2. \tag{8}$$

The deflection of the beam in this case is uniform ($\tilde{y} = \tilde{y}_0$), and the force applied to the beam by the elastic foundation balances the electrostatic force. For incipient buckling, (8) holds and can be subtracted from (6) to yield

$$B \frac{\sin(2\pi \tilde{x}/\alpha)}{\alpha^4} \left(\alpha^4(1-\delta) - 2\alpha^2\beta + 1 \right) = 0. \tag{9}$$

The last equation can be solved for the critical buckling parameters (β, α), and the amplitude B remains arbitrary, as is usual in linear buckling analysis [Godoy 1999; Timoshenko 1936]. The stability equation of the beam is given by

$$\alpha^4(1-\delta) - 2\alpha^2\beta + 1 = 0.$$

The solution of the this equation is

$$\alpha = \sqrt{\frac{\beta \pm \sqrt{\beta^2 - (1-\delta)}}{1-\delta}}. \tag{10}$$

The normalized wavelength α must be a real positive value ($\alpha > 0$), therefore β is restricted by

$$\beta \geq \sqrt{1-\delta}. \tag{11}$$

Substituting the critical load $\beta_{cr} = \sqrt{1-\delta}$ into (10), the critical value of the wavelength is found

$$\alpha_{cr} = \frac{1}{\sqrt{\beta_{cr}}} = \frac{1}{(1-\delta)^{1/4}} = \left(1 - \frac{2\tilde{V}^2}{(1-\tilde{y}_0)^3} \right)^{-1/4}.$$

From this it is clear that when no electrostatic forces are applied (that is, $\tilde{V} = 0$) buckling cannot occur for $\beta < 1$ (that is, subcritical stress). The critical normalized wavelength in the case of $\tilde{V} = 0$ is $\alpha_{cr(\tilde{V}=0)} = 1$, which proves that Λ_{cr} is indeed the wavelength of the critical buckling state [Abu-Salih and Elata 2005].

From (11) it is clear that due to the destabilizing effect of the electrostatic forces, buckling can occur for $\beta < 1$ whenever $\delta > 0$. The buckling deflection (4) is therefore due to the combined effect of subcritical prestress and electrostatic forces. This justifies the term electromechanical buckling.

A schematic illustration of a prestressed beam that is bonded to an elastic foundation is presented in Figure 1a. When the prestress in the conductive beam is below the critical value, and no voltage is applied, the beam remains flat (Figure 1b). For the same level of prestress, a sufficient voltage will precipitately instigate buckling (Figure 1c).

For a given normalized prestress β, the critical switching voltage $\tilde{V}_{cr}$ that instigates buckling is extracted by solving (7), (8) and $\beta_{cr} = \sqrt{1 - \delta}$, to yield

$$\tilde{V}_{cr}^2 = 4 \frac{1 - \beta_{cr}^2}{\left(3 - \beta_{cr}^2\right)^3}.$$

Figure 2 presents the normalized voltage square $\tilde{V}_{cr}^2$, wavelength α_{cr}, and deflection $\tilde{y}_0$, at the critical states, for various values of normalized prestress β. At the limit of zero prestress, the critical wavelength becomes infinite and the voltage approaches an asymptotic value. This state is in essence the pull-in state of an infinite parallel-plates actuator. At this limit the deflection at the critical state is $\tilde{y}_0 = 1/3$. For such a large reduction of the elastic foundation thickness, the linear model of the foundation may be unrealistic.

4. Initial postbuckling state

In the postbuckled state, considerable membrane stresses develop and the mechanical response is governed by the *nonlinear* equilibrium Equation (2), including the term in square brackets. The periodic problem is solved for beams with various finite normalized lengths, α. In each of these lengths, it is assumed that the solution includes one period. The average total potential density (energy per unit length) associated with each length α is then computed. The solution of the infinite problem is identified with the periodic solution of the finite-length problem, for which the average total potential density is minimal.

The periodic boundary conditions of the finite beam are

$$\tilde{y}(\tilde{x} = 0) = \tilde{y}(\tilde{x} = \alpha), \qquad \left.\frac{d\tilde{y}}{d\tilde{x}}\right|_{\tilde{x}=0} = \left.\frac{d\tilde{y}}{d\tilde{x}}\right|_{\tilde{x}=\alpha}. \tag{12}$$

4.1. *Approximate analytic solution.* As in the previous linear analysis, it is postulated that the postbuckling deflection is of the form

$$\tilde{y} = \tilde{y}_0' + B' \sin \frac{2\pi \tilde{x}}{\alpha}, \tag{13}$$

where B' is the normalized amplitude of the postbuckling deflection. In initial postbuckled states we assume that B' is small. As in the preceding critical state analysis, for small values of B' the electrostatic force may be approximated by the following Taylor expansion

$$\frac{\tilde{V}^2}{(1 - \tilde{y})^2} \approx \tilde{V}^2 \left(\frac{1}{(1 - \tilde{y}_0')^2} + \frac{2}{(1 - \tilde{y}_0')^3} B' \sin \frac{2\pi \tilde{x}}{\alpha} + O(B'^2) \right). \tag{14}$$
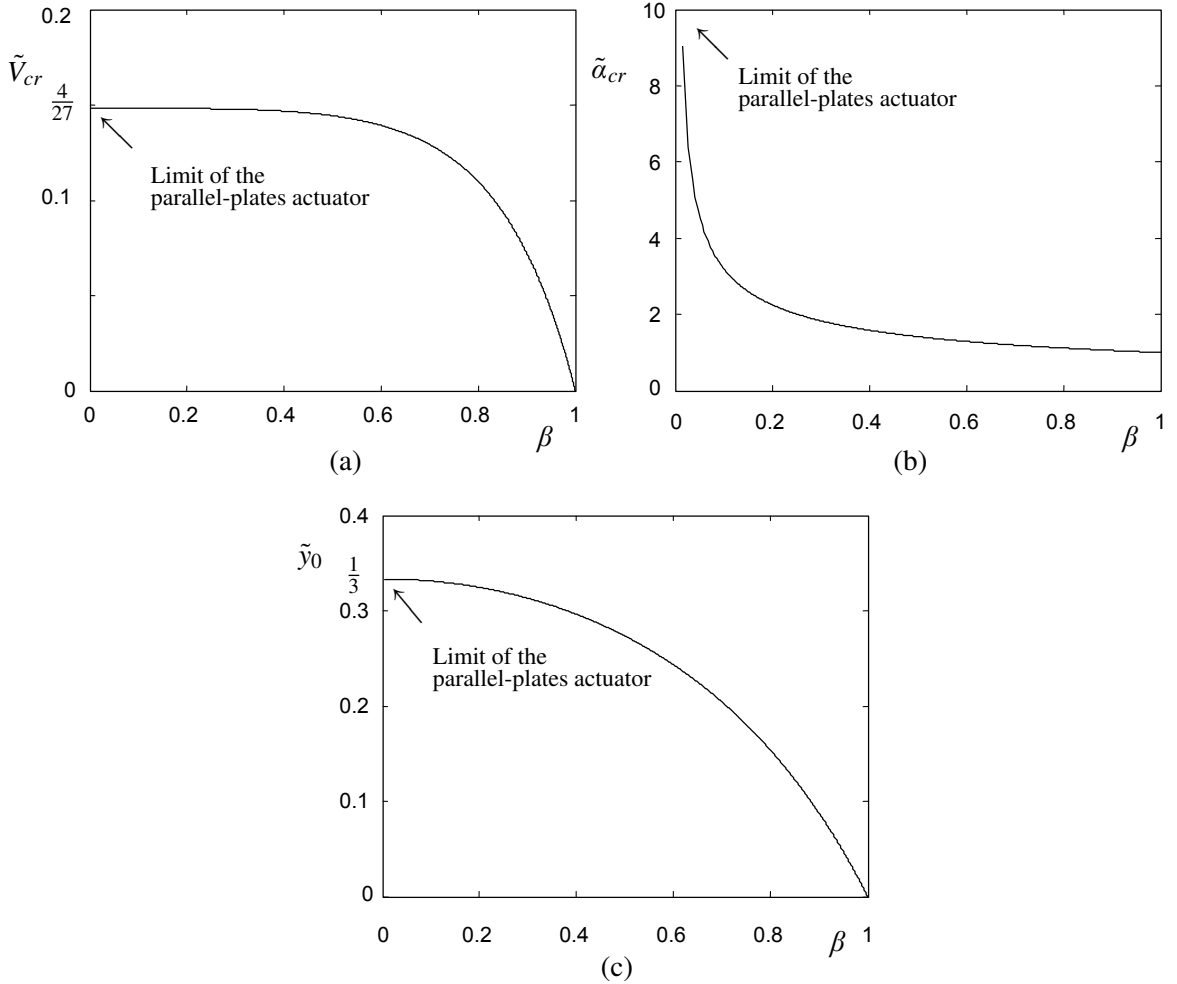
Figure 2. The critical electromechanical buckling state for different values of prestress β. (a) Normalized critical voltage squared $\tilde{V}_{cr}^2$. At zero prestress the critical voltage is the same as for a parallel-plates actuator. (b) Normalized critical wavelength α_{cr}. At zero prestress the normalized wavelength is infinite and the beam remains flat. (c) Normalized deflection $\tilde{y}_0$.

Substituting (13) and (14) into the nonlinear equation (2), and considering only small values of B', we further assume that the equilibrium equation holds separately for the average and the periodic parts of the postulated deflection.

With this assumption, the equilibrium equation (2) reduces to the following

$$\tilde{V}^2 = \tilde{y}_0'(1 - \tilde{y}_0')^2, \tag{15}$$

$$B'\frac{\sin(2\pi \tilde{x}/\alpha)}{\alpha^4}\left(\alpha^4(1-\delta) - 2\beta\alpha^2 + \tfrac{1}{4}\tilde{g}^2 B'^2 + 1\right) = 0. \tag{16}$$

The two assumptions, namely that the postbuckling deflection is of the form (13), and that when B' is small the solution of (2) may be substituted by the simultaneous solution of (15) and (16), will be validated numerically in the next subsection.

The nontrivial solution of (16) is given by

$$B' = \frac{2}{\tilde{g}}\sqrt{2\beta\alpha^2 - \alpha^4(1-\delta) - 1}\,.$$

In an infinite beam, the normalized wavelength associated with given values of β and δ is the one that minimizes the total potential per unit length, of the system. The total potential $\psi = U_B + U_A + U_{k_f} - U_E^*$ is the sum of four energy components associated with bending U_B, axial deformation U_A, deformation of the elastic foundation U_{k_f}, and complementary electrostatic energy U_E^* [Elata and Abu-Salih 2005]. Normalizing the strain energy components by the (axial) strain energy at the verge of buckling ($U_{cr} = \sigma_{cr}^2 A/2E$), yields

$$\tilde{U}_B = \frac{\tilde{g}^2}{4(2\pi)^4\alpha}\int_0^\alpha \left(\frac{d^2\tilde{y}}{d\tilde{x}^2}\right)^2 d\tilde{x},$$

$$\tilde{U}_A = \frac{\tilde{g}^4}{4\alpha^2}\left(-\frac{2(2\pi)^2\alpha\beta}{\tilde{g}^2} + \frac{1}{2}\int_0^\alpha \left(\frac{d\tilde{y}}{d\tilde{x}}\right)^2 d\tilde{x}\right)^2,$$

$$\tilde{U}_{k_f} = \frac{\tilde{g}^2}{4\alpha}\int_0^\alpha \tilde{y}^2 d\tilde{x},$$

$$\tilde{U}_E^* = \frac{\tilde{V}^2\tilde{g}^2}{2\alpha}\int_0^\alpha \frac{d\tilde{x}}{(1-\tilde{y})}.$$

The normalized total potential is accordingly given by

$$\tilde{\psi} = \tilde{U}_B + \tilde{U}_A + \tilde{U}_{k_f} - \tilde{U}_E^*. \tag{17}$$

Substituting the postulated deflection (13) into the normalized total potential, differentiating the total potential with respect to α, and setting this derivative to zero, yields a nonlinear equation for the postbuckling wavelength α. This equation was numerically solved and the results are presented by the *dashed lines* in Figure 3, showing α, B' and $\tilde{y}_0'$ as function of $\tilde{V}^2/\tilde{V}_{cr}$ for the parameters $\tilde{y}_0 = 0.1$ (that is, $\beta = 0.88$), $g/h = 10$ and $\upsilon = 0.25$.

To clarify the notion of minimal energy, the total potential of the system (17) is presented in Figure 4 as a function of voltage and wavelength (for $\tilde{y}_0 = 0.1$ (that is, $\beta = 0.88$), $g/h = 10$ and $\upsilon = 0.25$). In this figure the flat slopes are the regions in which no buckling occurs (the total energy is only a function of V). The first valley describes buckling with a single wavelength in a beam of finite length with periodic boundary conditions. The second and third valleys are repetitions of this first valley, and describe buckling into a double and a triple flexure waves for a beam with double and triple length, respectively (and periodic boundary conditions). The center lines running through the routes of the valleys describe the same postbuckling wavelength as a function of applied voltage (that is, repetitions of the same solution). The critical stability states are marked by the solid line on the rims of the valleys [Abu-Salih and Elata 2005].
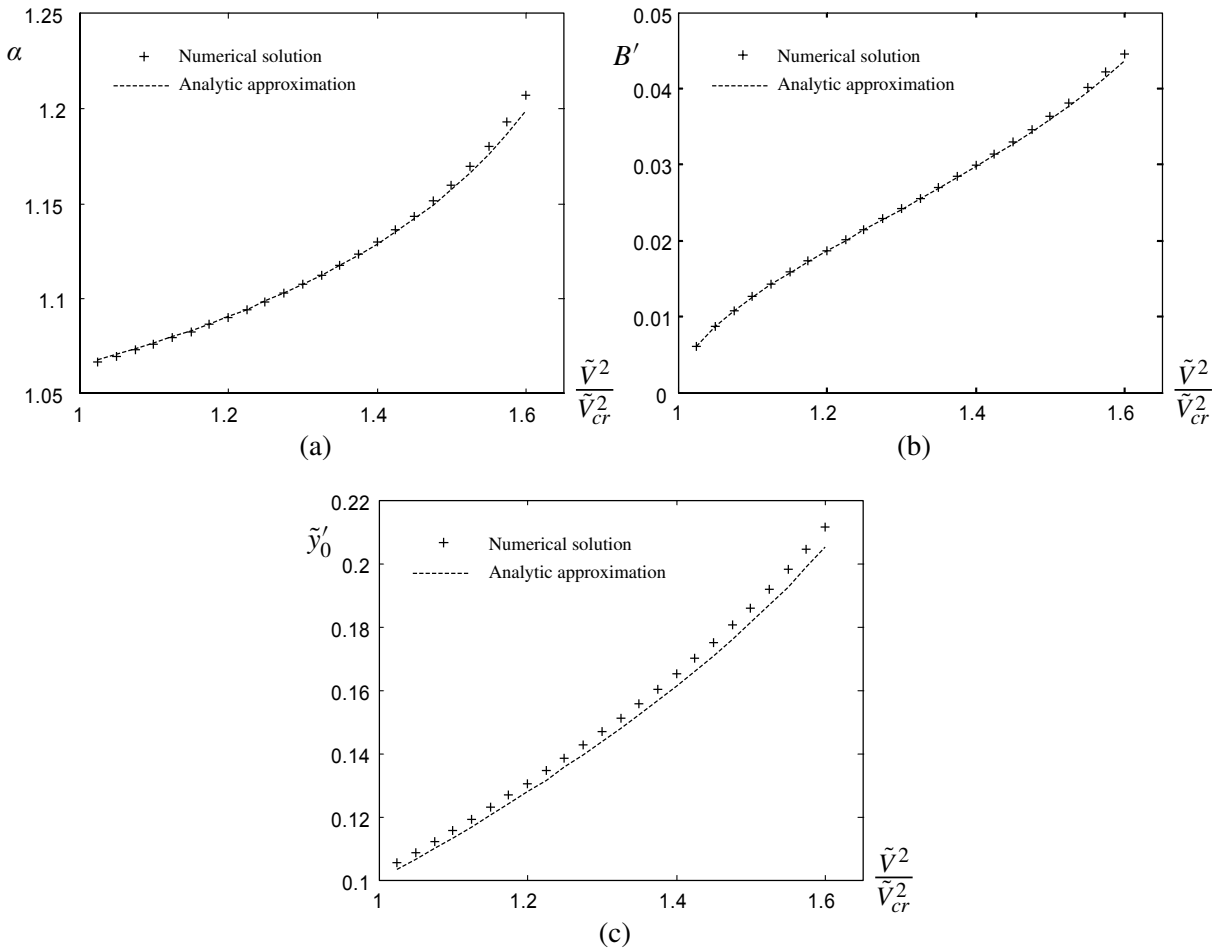
Figure 3. Postbuckling state as a function of applied voltage for $\tilde{y}_0 = 0.1$ (that is, $\beta = 0.88$), $g/h = 10$ and $\upsilon = 0.25$. (a) The normalized wavelength; (b) normalized amplitude of the sinusoidal component of the deflection; (c) the normalized deflection average.

4.2. *Numerical validation.* To validate the analytic solution presented in the previous subsection, a finite-difference numerical code for solving Equation (2) with periodic boundary conditions (12) was implemented in MATLAB®. For given supercritical voltages (that is, $\tilde{V}^2 > \tilde{V}_{cr}^2$), the code numerically computed the total potential of the system (17) and numerically found the normalized wavelength that minimizes this potential per unit length.

Figure 5 presents the convergence of the postbuckling deflection parameters as a function of the number of nodes n in the finite-difference mesh for the conditions $\beta = 0.9$, $g/h = 10$, $\upsilon = 0.25$, and $\tilde{V}^2/\tilde{V}_{cr}^2 = 1.35$. For these conditions the postbuckling deflection parameters are $\alpha = 1.092$, $\tilde{y}_0' = 0.1288$ and $B' = 0.02771$. To verify that the postbuckling deflection is of the functional form (13), the deflection average was subtracted from the deflection which was then normalized such that its amplitude was in the
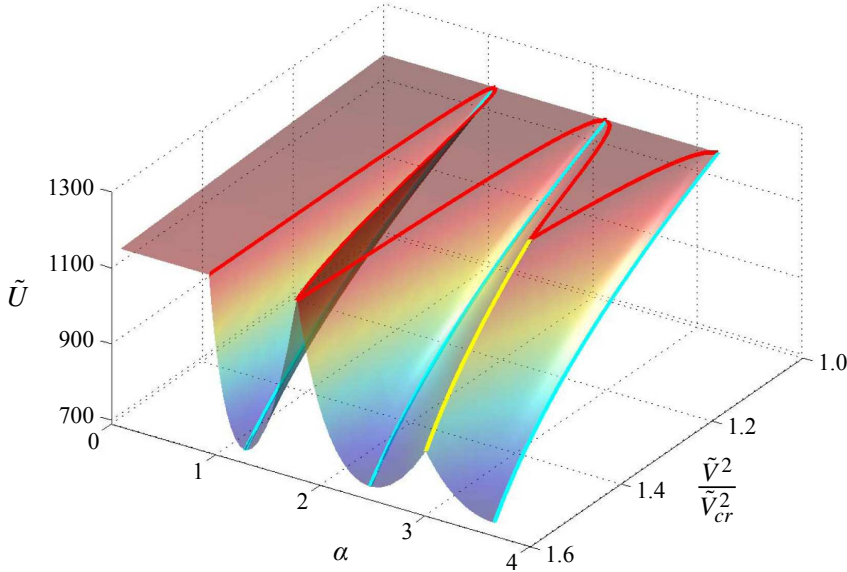
Figure 4. The total potential energy of a finite beam with periodic boundary conditions and $\tilde{y}_0 = 0.1$ (that is, $\beta = 0.88$), $g/h = 10$ and $\upsilon = 0.25$, as a function of the normalized applied voltage and the postulated wavelength. The flat regions describe states in which the applied voltage is insufficient to induce buckling for prescribed values of the postulated wavelength.

range $-1 \le \tilde{y} - \tilde{y}_0' \le 1$. The root-mean-square of the error from a perfect sine wave was then computed and found to be $rms = 6.5 \cdot 10^{-4}$.

Similar convergence properties and compliance with the postulated waveform (13) were found when the response of a beam with lengths $L/\Lambda_{cr} = 2\alpha$ and $L/\Lambda_{cr} = 3\alpha$ was simulated. In this case the solution was a double and triple repetition of the solution obtained for a beam of length α.

Last, it was assumed that the postbuckling deflection parameter B' is small and that the solution of (2) may be substituted by the simultaneous solution of (15) and (16). Figure 3 presents the numerically computed values of α, B' and $\tilde{y}_0'$ ('+' marks) as functions of $\tilde{V}^2/\tilde{V}_{cr}$ for the parameters $\tilde{y}_0 = 0.1$ (that is, $\beta = 0.88$), $g/h = 10$ and $\upsilon = 0.25$. These results are based on a finite-difference solution of (2) with boundary conditions (12). In these numerical simulations, the functional waveform of the deflection was not a priori constrained in any way (except for the periodic boundary conditions). The agreement between the numerically computed values and the analytic approximation verify that for small values of $\tilde{V}^2/\tilde{V}_{cr}^2$ these assumptions are valid.

Figure 6 presents the value of $\tilde{V}^2/\tilde{V}_{cr}^2$ for which B' is 1%, 3%, and 5%. It is clear that for small values of β the voltage range for which the analysis is applicable is rather limited. However, for $\beta \ge 0.85$, B' remains small for a considerable range of $\tilde{V}^2/\tilde{V}_{cr}^2$.

In a series of numerical simulations (not presented in this manuscript), for different values of prestress β, the applied voltage was increased up to the pull-in point [Elata and Abu-Salih 2005]. At these simulated pull-in states, not only was the average deflection $\tilde{y}_0'$ considerable (for example, $\tilde{y}_0' \approx 1/3$), but also
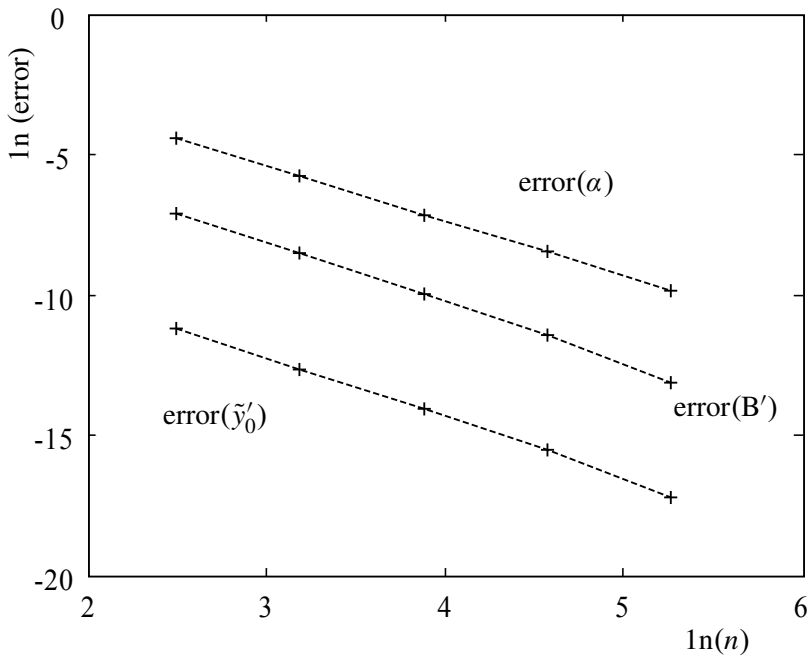
Figure 5. Convergence of the postbuckling deflection parameters as a function of the number of nodes n (for $\beta = 0.9$, $\alpha = 1.092$, $\tilde{V}^2 / \tilde{V}_{cr}^2 = 1.35$).

the amplitude of the sinusoidal deflection component, B', became rather large such that the deflection form (13) is no longer a valid approximation. Under these conditions the equilibrium equation becomes invalid due to the large rotations, and in any case, the high reduction of the elastic foundation thickness would suggest that a nonlinear foundation response should be considered.

5. Discussion

In this work the electromechanical buckling of a prestressed beam which is bonded to a dielectric elastic foundation is analyzed. An analytic solution of the critical electromechanical state is derived, and it is shown that electrostatic forces can precipitately instigate buckling even when the prestress in the beam is lower than the critical value that would cause mechanical buckling. An analytic approximation of the initial postbuckling state is also presented, and is validated numerically. The numerical simulations show that a stable, initial postbuckling state exists.

The analysis presented in this study considers perfect elastic response with no residual strains, for example, plastic deformation. The work done by the voltage source is invested in elastic strain energy (of the beam and foundation) and electrostatic potential between the beam and substrate. When the voltage source is turned off (effectively setting $V = 0$) the deformed system will return to the original state with a flat prestressed beam and an unloaded elastic foundation. This suggests that electrostatic potential can be used to achieve on/off switching of flexure waves in a prestressed beam. Such on/off
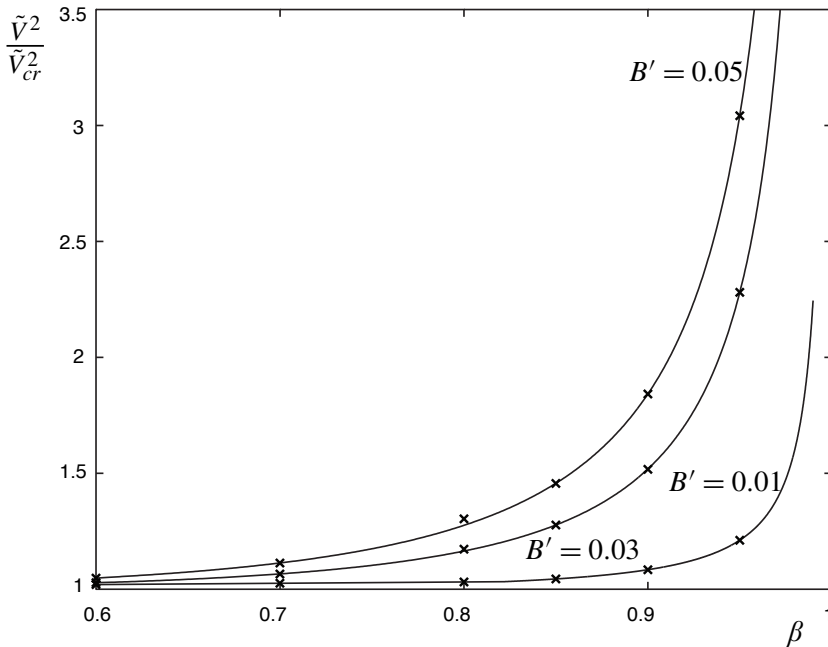
Figure 6. Validity range of the assumption of small B'.

switching of flexures may be useful in microsystems. For example, it may enable repeated reversible modification of the optical reflectivity of a microstructure.

Finally, one may always pose the question of how long must an actual beam be to justify its consideration as infinite. Electromechanical buckling may also occur in beams of finite length, and this response will be affected by the boundary conditions at the edges. In finite but long beams, the postbuckled flexures in regions that are sufficiently far from either edge (distance measured in number of wavelengths) will resemble the postbuckled flexures in an infinite beam.

References

[Abu-Salih and Elata 2005] S. Abu-Salih and D. Elata, "Analytic postbuckling solution of a pre-stressed infinite beam bonded to a linear elastic foundation", *Int. J. Solids Struct.* **42**:23 (2005), 6048–6058.

[Brush and Almaroth 1975] D. O. Brush and B. O. Almaroth, *Buckling of bars, plates and shells*, McGraw-Hill, New York, 1975.

[Elata and Abu-Salih 2005] D. Elata and S. Abu-Salih, "Analysis of a novel method for measuring residual stress in microsystems", *J. Micromech. Microeng.* **15** (2005), 921–927.

[Elata and Leus 2005] D. Elata and V. Leus, "How slender can comb-drive fingers be?", *J. Micromech. Microeng.* **15** (2005), 1055–1059.

[Elata et al. 2003] D. Elata, O. Bochobza-Degani, and Y. Nemirovsky, "Analytical approach and numerical alpha-Lines method for pull-in hyper-surface extraction of electrostatic actuators with multiple uncoupled voltage sources", *J. Microelectromech. S.* **12**:5 (2003), 681–691.

[Gilmore 1981] R. Gilmore, *Catastrophe theory for scientists and engineers*, Wiley, New York, 1981. Reprinted Dover, New York, 1993.

[Godoy 1999] L. A. Godoy, *Theory of elastic stability: analysis and sensitivity*, Taylor and Francis, Philadelphia, 1999.

[Hetenyi 1946] H. Hetenyi, *Beams on elastic foundation*, Ann Arbor, University of Michigan Press, 1946.

[Legtenberg et al. 1996] R. Legtenberg, A. W. Groeneveld, and M. Elwenspoek, "Comb-drive actuators for large displacements", *J. Micromech. Microeng.* **6** (1996), 320—329.

[Leus and Elata 2004] V. Leus and D. Elata, "Fringing field effect in electrostatic actuators", Technion, Haifa, 2004, Available at http://meeng.technion.ac.il/Research/TReports/2004/ETR-2004-02.html.

[Nguyen 2000] Q. S. Nguyen, *Stability and nonlinear solid mechanics*, Wiley, New York, 2000.

[Pelesko and Bernstein 2003] J. A. Pelesko and D. H. Bernstein, *Modeling MEMS and NEMS*, Chapman & Hall/CRC, Boca Raton, FL, 2003.

[Timoshenko 1936] S. P. Timoshenko, *Theory of elastic stability*, McGraw-Hill, New York, 1936.

DAVID ELATA: elata@tx.technion.ac.il
Faculty of Mechanical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel

SAMY ABU-SALIH: samyas@tx.technion.ac.il
Faculty of Mechanical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel

ON UNIQUENESS IN THE AFFINE BOUNDARY VALUE PROBLEM OF THE NONLINEAR ELASTIC DIELECTRIC

R. J. KNOPS AND C. TRIMARCO

An integral identity is constructed from properties of the energy momentum tensor and is used to demonstrate uniqueness of the displacement on star-shaped regions to the affine boundary value problem of the nonlinear homogeneous elastic dielectric. The method of proof, nontrivially adapted from that of the corresponding elastic problem, assumes the electric enthalpy function to be rank-one convex and strictly quasiconvex. Furthermore, for a given displacement gradient, the electric quantities are proved unique for specified nonaffine and nonuniform electric boundary conditions subject to the electric enthalpy and strain energy functions satisfying additional convexity conditions.

1. Introduction

This paper considers uniqueness of smooth solutions to certain simple boundary value problems for the nonlinear homogeneous elastic dielectric in equilibrium and occupying a bounded region of n-dimensional Euclidean space subject to zero body-force and electric source charges. Uniqueness in corresponding electromagnetic problems is guaranteed by strict convexity of the energy. A similar condition on the strain energy function also ensures uniqueness to boundary value problems of nonlinear elastostatics, but several well known counterexamples demonstrate that universal uniqueness is untenable. The counterexamples mean that also for the nonlinear elastic dielectric unqualified uniqueness is unacceptable and consequently the condition of strict convexity is too restrictive even for ferroelectrics and similar materials exhibiting phase transitions. But equally, we do not expect there to be universal nonuniqueness since it is intuitively evident that certain simple problems should possess a unique solution.

In elastostatics, this topic has been treated in [Knops and Stuart 1984] where the notions of strict quasiconvexity and rank-one convexity of the strain energy function are introduced to establish uniqueness of a smooth solution to affine displacement boundary value problems on star-shaped bounded regions. Pivotal to the proof is a Noetherian conservation law [Gelfand and Fomin 1963] which in the context of elasticity has been separately derived either directly or from properties of the energy-momentum (Eshelby) tensor. Contributions notably include those by Chadwick [1975], Eshelby [1975], Green [1973], Günther [1962], Gurtin [2000], Hill [1986], Knowles and Sternberg [1972]. Pohozaev [1965] and Pucci and Serrin [1986] are among those who have used a similar law in partial differential equations.

The uniqueness proof presented here for the nonlinear elastic dielectric is patterned on that described in [Knops and Stuart 1984], and therefore is likewise restricted to star-shaped regions. Another application of the basic proof is by Mareno [2004] who investigated uniqueness in the second order theory of nonlinear elasticity. As a consequence of these previous studies, the intrinsic mathematical interest of the present paper is seen as lying not so much in the details required to extend the proof, but rather

---

in the electromagnetic problems under consideration, for which the uniqueness results are new. In this respect, it is perhaps worth remarking that uniqueness results generally are restricted to the linear theory and comparatively little information appears available for nonlinear elastic dielectrics. Therefore the following conclusions are viewed as contributing to a basic understanding of the coupled theory of elasticity and electromagnetism, with particular relevance for numerical computation and those aspects concerned with buckling and hysteresis.

For ease of presentation, it is convenient to ignore magnetic effects, although these may also be included. It is also convenient to conduct the analysis with respect to the reference configuration which requires the electric fields and the governing Maxwell–Lorentz equations to be appropriately transformed from their usual formulation in the current configuration. For this purpose, we appeal in part to a variational procedure in which the strain energy function is replaced by an electric enthalpy regarded as a function of the deformation gradient, electric field and the polarization referred to the reference configuration. The consequent enlarged set of Euler–Lagrange equations contain the transformed Maxwell–Lorentz equations, and enable certain properties of the energy-momentum tensor to be derived. These in turn lead to an identity, analogous to the elastostatic conservation law, whose construction, while nontrivial, appears to be more direct than is usual; compare for example, [Ericksen 2006, to appear; Maugin 1993; Maugin and Trimarco 1991; 2001, Trimarco 2002; 2003; and Pack and Herrmann 1986]. The Maxwell–Lorentz equations also imply that the electric enthalpy is independent of the polarization, so that the conditions for the electric enthalpy to be quasiconvex and rank-one convex, both essential for the proof, need only be defined in terms of the deformation gradient and the electric field. As is well known, these generalized notions of convexity hold for deformation gradients and electric fields that may be discontinuous across an internal surface provided both satisfy geometrical compatibility conditions that characterize coherent phases in solids. This aspect is not developed in what follows. Furthermore, the generalized convexity notions both reduce to convexity in the usual sense for functions whose arguments are vectors or scalars. The reduction is pertinent to the concluding discussion regarding uniqueness under mixed boundary conditions of the electric constituents for a given deformation gradient, which includes application to a dielectric embedded in a capacitor.

Section 2 assembles essential preliminaries, introduces the electric enthalpy function, states, and, for completeness, proves the conservation law. The notions of rank-one convexity and quasiconvexity are introduced in Section 3 which also specifies the affine boundary conditions and constructs the uniqueness proof. The concluding remarks, given in Section 4, include alternative conditions for uniqueness in the affine boundary value problem, and a discussion of uniqueness of the electric displacement and polarization in the affine problem. For completeness, we supplement these results by demonstrating that for a given deformation gradient the electric field remains unique when the affine electric boundary conditions are replaced by those of standard nonuniform mixed type. Under these boundary conditions we also prove uniqueness of the polarization and the electric displacement subject additionally to the strain energy being rank-one convex with respect to the polarization, a condition possibly too severe for ferroelectrics but for which uniqueness of the deformation and electric field remains valid. The conclusions are, of course, not unexpected, although perhaps less so in the nonlinear theory, and merely reflect the reduction, already remarked, of generalized notions of convexity to that of standard convexity for functions of vector quantities.

The direct tensor notation is mainly employed except when greater clarity is achieved by the corresponding suffix notation. Existence of a smooth solution is assumed, but in this respect we note that under the generalized convexity conditions discussed here, existence of elastostatic weak solutions has been established by Ball [1976], while Serre [2004] has discussed the corresponding electromagnetic problem.

2. Notation and other preliminaries

A nonlinear homogeneous elastic dielectric in its reference configuration occupies the bounded region $\Omega \subset \mathbb{R}^n$ of n-dimensional Euclidean space. The piecewise continuously differentiable boundary $\partial \Omega$ of Ω has unit outward normal N and is assumed to be star-shaped with respect to an interior point. Boundary conditions, precisely stated in the next section, produce an equilibrium deformation of the dielectric in which a point $X \in \Omega$ becomes displaced to the point x, where X, x represent vectors in $\mathbb{R}^n$ whose components with respect to a Cartesian coordinate system are X_A and x_i. The deformation, supposed smooth, possesses a gradient expressed by

$$F = \frac{\partial x}{\partial X}, \qquad F_{iA} = \frac{\partial x_i}{\partial X_A}, \tag{2–1}$$

while the inverse is given by

$$F^{-1} = \frac{\partial X}{\partial x}, \qquad F_{Ai}^{-1} = \frac{\partial X_A}{\partial x_i}, \tag{2–2}$$

and the determinant associated with Equation (2–1) is $J = det\, F$. Let $M^{m \times n}$ denote the set of $m \times n$ matrices and suppose that $J \in M_+^{n \times n}$, the set of square matrices with positive determinant. The transpose of a tensor A is denoted by A^T; the identity tensor by I; and the tensorial trace operator by tr. Tensor and vector multiplication is indicated by juxtaposition, the precise form being clear from the particular context, while the inner product of tensors AB is given by

$$AB = tr\, AB^T.$$

In the deformed dielectric there is an electric field e, an electric displacement d, and a polarization p per unit deformed volume. These vector quantities satisfy the appropriate time independent Maxwell–Lorentz equations which in the assumed absence of electric source charges lead to the following the expressions $e = -\,\mathrm{grad}\,\varphi$ and $d = \epsilon_0 e + p$, where grad denotes the gradient operator with respect to the system of current coordinates x, $\varphi(x)$ is a scalar potential function of the variables x, and ϵ_0 is the *in vacuo* dielectric constant supposed positive.

It is convenient to develop the subsequent analysis with respect to the reference configuration Ω. Now, in terms of the notation

$$E = F^T e, \quad D = J F^{-1} d, \quad P = J F^{-1} p, \tag{2–3}$$

it has been shown by, for example, Walker et al. [1965] that the Maxwell–Lorentz equations in the reference configuration imply the relations

$$E = -\,\mathrm{Grad}\,\Phi, \tag{2–4}$$

$$D = \epsilon_0 J F^{-1} F^{-T} E + P, \tag{2–5}$$

where Grad represents the standard gradient operator with respect to the system of reference coordinates X, and $\Phi(X)$ is a scalar potential function of the variable X.

Next, we suppose that the dielectric possesses a strain energy function

$$W : M^{(n+1)\times n} \to \mathbb{R}$$

per unit volume of the reference configuration that depends upon both the deformation gradient and the electric polarization:

$$W = W(F, FP). \tag{2--6}$$

Furthermore, we introduce the electric enthalpy function $H(F, P, E)$ [Maugin and Trimarco 1991; Trimarco 2002] defined by

$$H(F, P, E) = W(F, FP) - \frac{1}{2}\epsilon_0 J E F^{-1} F^{-T} E - E P, \tag{2--7}$$

from which the identity immediately follows:

$$D = -\frac{\partial H}{\partial E}(F, P, E). \tag{2--8}$$

The complete set of equations governing the deformation of the dielectric may be obtained by considering the stationary points of the electric enthalpy (2–7) with respect to independent variations of Φ, F, and P. (See, for example, [Trimarco 2002; 2003; Maugin and Trimarco 1991; 2001; Maugin 1993; Pack and Herrmann 1986; Yu 1995] and the important discussion in [Ericksen 2006, to appear].) This yields the Euler–Lagrange equations:

$$\text{Div}\,\frac{\partial H}{\partial F}(F, P, E) = 0, \tag{2--9}$$

$$\text{Div}\,D = 0, \tag{2--10}$$

$$\frac{\partial H}{\partial P} \equiv \frac{\partial W}{\partial P} - E = 0, \tag{2--11}$$

where Div denotes the divergence operator with respect to the system of reference coordinates X.

We observe that Equations (2–10) and (2–4) are the usual electrostatics equations in the absence of electric free charge, and in fact (2–10) implies that d is solenoidal in the current configuration of the elastic dielectric. The classical electrostatics equations are consequentially recovered. Moreover, the transformation (2–3) and (2–5) together with (2–10) are consistent with the fundamental requirement that total electric charge be conserved.

From (2–11) we conclude that H is independent of P so that $H(F, E) : M^{(n+1)\times n} \to \mathbb{R}$

A crucial ingredient of the uniqueness proof described in the next section is a conservation law (or integral identity) that is stated and proved in the following lemma.

Lemma 2.1 [Ericksen 2006, to appear; Maugin 1993; Maugin and Epstein 1991; Trimarco 2002; 2003]. *Let $\Omega \subset \mathbb{R}^n$ have smooth boundary $\partial\Omega$ with unit outward normal N. Let (2–4), (2–5) and (2–7) hold, let $(x, \text{Grad}\,\Phi)$ be a smooth solution to the equilibrium equations (2–9)–(2–11), and let the electric enthalpy*

satisfy $H \in C^2(M^{(n+1)\times n}, \mathbb{R})$. Then:

$$
n \int_\Omega H(F, E)\, dX = \int_{\partial\Omega} \Bigg((NX)H(F, E) + tr \frac{\partial H(F, E)}{\partial F}\left(x - (X\,\mathrm{Grad})x\right)
$$

$$
+ ND(F, E)\left(\Phi - (X\,\mathrm{Grad})\Phi\right)\Bigg)\, dS, \tag{2–12}
$$

where dX and dS represent respectively the volume and surface elements of integration in the reference configuration.

Proof. The identity Equation (2–12) may be established by application of the divergence theorem either to the surface integral on the right and noting that

$$
\frac{\partial H}{\partial X} = \frac{\partial H}{\partial F}\,\mathrm{Grad}\ F - (D\,\mathrm{Grad})E, \tag{2–13}
$$

or, after rearrangement of the integrand and appeal to (2–9), to the integral identity

$$
\int_\Omega \left(XF^T \,\mathrm{Div}\, \frac{\partial H}{\partial F}\right) dX = 0. \tag{2–14}
$$

Instead, we prefer to employ the energy-momentum, or Eshelby, tensor B defined by

$$
B = (W - EP)I - F^T\frac{\partial W}{\partial F} + E \otimes P, \tag{2–15}
$$

where $E \otimes P$ denotes the tensor product of the vectors E and P. The relation

$$
\frac{\partial H}{\partial F} = \frac{\partial W}{\partial F} + JTF^{-T}, \tag{2–16}
$$

where T is the Maxwell stress tensor given explicitly by

$$
T = \epsilon_0(e \otimes e - \frac{1}{2}ee I), \tag{2–17}
$$

enables (2–15) to be alternatively expressed as

$$
B = HI - F^T\frac{\partial H}{\partial F} + E \otimes D, \tag{2–18}
$$

from which by appeal to (2–9) and (2–10) we may directly prove that

$$
\mathrm{Div}\ B = 0. \tag{2–19}
$$

On following Chadwick's [Chadwick 1975] or Hill's [Hill 1986] approach to the corresponding elastic problem, we have

$$
\mathrm{Div}(XB) = tr\ B
$$

$$
= nH - \mathrm{Div}(x\frac{\partial H}{\partial F} - \mathrm{Div}(D\Phi), \tag{2–20}
$$

which by integration over Ω leads to (2–12). $\qquad\square$

3. Uniqueness in the affine boundary value problem

In this section we prove that the affine boundary value problem has a unique smooth solution provided the electric enthalpy satisfies generalized convexity conditions. The region Ω is supposed to be star-shaped with respect to an interior point which without loss may be taken as the origin of coordinates so that

$$NX > 0, \quad X \in \partial\Omega. \tag{3–1}$$

We commence by considering two distinct smooth equilibrium solutions (x, Φ) and (y, Ψ) to the dielectric Equations (2–9)–(2–11) that satisfy the same boundary conditions in the following sense

$$x - y = 0, \qquad X \in \partial\Omega, \tag{3–2}$$

$$\Phi - \Psi = 0, \qquad X \in \partial\Omega. \tag{3–3}$$

Then by Hadamard's lemma [Hadamard 1903] it follows that

$$\mathrm{Grad}(y - x) = \lambda \otimes N, \quad \lambda = N\,\mathrm{Grad}(y - x), \qquad X \in \partial\Omega, \tag{3–4}$$

$$\mathrm{Grad}(\Psi - \Phi) = \mu N, \qquad \mu = N\,\mathrm{Grad}(\Psi - \Phi), \quad X \in \partial\Omega. \tag{3–5}$$

The first convexity assumption imposed on the electric enthalpy is that of rank-one convexity. The precise notion used is defined in the second of the following two related definitions.

Definition 3.1 (Rank-one convexity at a point). The function

$$H \in C(M^{(n+1)\times n}, \mathbb{R})$$

is (strictly) rank-one convex at F and E if and only if

$$H(F + ta \otimes b, E + tQ) \le tH(F + a \otimes b, E + Q) + (1 - t)H(F, E), \tag{3–6}$$

for all

$$t \in [0, 1], \quad F \in M_+^{n\times n}, \quad E \in \mathbb{R}^n, \quad Q \in \mathbb{R}^n, \quad a \in \mathbb{R}^n, \quad b \in \mathbb{R}^n,$$

such that $F + ta \otimes b \in M_+^{n\times n}$. Strict rank-one convexity at a point holds when the inequality in (3–6) is strict.

When $H \in C^1(M^{(n+1)\times n}, \mathbb{R})$, an immediate deduction from (3–6), obtained on taking the limit $t \to 0$, is the further inequality

$$H(F + a \otimes b, E + Q) \ge H(F, E) + \frac{\partial H(F, E)}{\partial F}a \otimes b + \frac{\partial H(F, E)}{\partial E}Q, \tag{3–7}$$

for all

$$F \in M_+^{n\times n}, \quad a \in \mathbb{R}^n, l \quad b \in \mathbb{R}^n, \quad E \in \mathbb{R}^n, \text{ and } Q \in \mathbb{R}^{n+},$$

such that $F + a \otimes b \in M_+^{n\times n}$.

Definition 3.2 (Rank-one convexity). The function $H(F, E)$ is (strictly) rank-one convex if and only if H is (strictly) rank-one convex at F and E for all $F \in M_+^{n\times n}$, and all $E \in \mathbb{R}^n$.

We can now state and prove the first lemma needed in the proof of uniqueness.

Lemma 3.1. *Let Ω be star-shaped with respect to the origin, and let*

$$H : M^{(n+1)\times n} \to \mathbb{R}$$

be rank-one convex. Let (x, Φ) and (y, Ψ) be distinct pairs of equilibrium smooth solutions to (2–9)–(2–11) *that satisfy the same Dirichlet boundary conditions in the sense of* (3–2) *and* (3–3). *When*

$$H \in C^1(M^{(n+1)\times n}, \mathbb{R}),$$

we have

$$n \int_{\Omega} \big(H(\text{Grad } x, -\text{Grad } \Phi) - H(\text{Grad } y, -\text{Grad } \Psi) \big) \, dX$$

$$\leq \int_{\partial\Omega} \left\{ \frac{\partial H(\text{Grad } x, -\text{Grad } \Phi)}{\partial F} - \frac{\partial H(\text{Grad } y, -\text{Grad } \Psi)}{\partial F} \right\} \times \{ N \otimes (y - (X \, \text{Grad}) \, y) \} \, dS$$

$$+ \int_{\partial\Omega} \left\{ N \frac{\partial H(\text{Grad } x, -\text{Grad } \Phi)}{\partial E} - N \frac{\partial H(\text{Grad } y, -\text{Grad } \Psi)}{\partial E} \right\} \times \{ -\Psi + X \, \text{Grad } \Psi \} \, dS. \quad (3–8)$$

Proof. The conservation law (2–12) by hypothesis is satisfied by both solutions (x, Φ) and (y, Ψ). Consequently, subtraction of the respective identities and appeal to (3–2)–(3–5) leads to the relation

$$n \int_{\Omega} \big(H(\text{Grad } x, -\text{Grad } \Phi) - H(\text{Grad } y, -\text{Grad } \Psi) \big) \, dX$$

$$= \int_{\partial\Omega} NX \big(H(\text{Grad } x, -\text{Grad } \Phi) - H(\text{Grad } x + \lambda \otimes N, -\text{Grad } \Phi - \mu N) \big) \, dS$$

$$+ \int_{\partial\Omega} NX \Big(\frac{\partial H}{\partial F}(\text{Grad } x, \text{Grad } \Phi) N \otimes \lambda - \mu N \frac{\partial H}{\partial E}(\text{Grad } x, \text{Grad } \Phi) \Big) \, dS$$

$$+ \int_{\partial\Omega} \Big(\frac{\partial H}{\partial F}(\text{Grad } x, \text{Grad } \Phi) - \frac{\partial H}{\partial F}(\text{Grad } y, \text{Grad } \Psi) \Big) \times \big(N \otimes (y - X, \text{Grad } y) \big) \, dS$$

$$+ \int_{\partial\Omega} \big(ND(\text{Grad } x, \text{Grad } \Phi) - ND \, \text{Grad } (y, \text{Grad } \Psi) \big) \times (\Psi - X \, \text{Grad } \Psi) \, dS. \quad (3–9)$$

The first two terms on the right are nonpositive by virtue of the star-shaped assumption (3–1) and inequality (3–7) for the rank-one convex function H. Consequently, the lemma is proved. $\square$

Remark 3.1. It is apparent from the proof of Lemma 3.1 that rank-one convexity of H is required only on the set of surface values of $\text{Grad}(y - x)$ and $\text{Grad}(\Psi - \Phi)$.

We next restrict our attention to affine boundary conditions. We have as a corollary to Lemma 3.1 the following lemma:

Lemma 3.2. *Let Ω be star-shaped and let* (3–1) *be satisfied. Let (x, Φ) be a smooth equilibrium solution to* (2–9)–(2–11), *let the electric enthalpy H be rank-one convex, and let x and Φ satisfy the respective affine boundary conditions*

$$x = c + AX, \quad X \in \partial\Omega, \quad (3–10)$$

$$\Phi = d + bX, \quad X \in \partial\Omega, \quad (3–11)$$

where $A \in M_+^{n \times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$ are constant.

    Then

$$\int_\Omega H(\text{Grad } x, -\text{Grad } \Phi) \, dX \leq \int_\Omega H(A, -b) \, dX. \tag{3-12}$$

Proof. Consider the affine equilibrium solution pair (y, Ψ) given by

$$y = c + AX, \qquad X \in \bar{\Omega}, \tag{3-13}$$

$$\Psi = d + bX, \qquad X \in \bar{\Omega}, \tag{3-14}$$

where as usual the overbar denotes closure; that is, $\bar{\Omega} = \Omega \cup \partial\Omega$. It easily follows that (y, Ψ) satisfies the boundary conditions (3–10) and (3–11), and also the relationships

$$\text{Grad } y = A, \quad \text{Grad } \Psi = b, \quad X \in \bar{\Omega}, \tag{3-15}$$

and

$$y - X \text{ Grad } y = c, \qquad X \in \bar{\Omega}, \tag{3-16}$$

$$\Psi - X \text{ Grad } \Psi = d, \qquad X \in \bar{\Omega}. \tag{3-17}$$

    The proof of the Lemma is completed upon noticing that the right side of Equation (3–8) vanishes by virtue of (3–15)–(3–17), the divergence theorem, and the equilibrium equations (2–9)–(2–10). $\qquad\square$

    Uniqueness of the affine solution (3–13) and (3–14) requires the introduction of our second general convexity assumption defined as follows:

Definition 3.3 (Quasiconvexity). The function $H \in C(M^{(n+1)\times n}, \mathbb{R})$ is quasiconvex at (A, b) if and only if

$$\int_\Sigma H(A + \text{Grad } \chi, -b + \text{Grad } \theta) \, dX \geq \int_\Sigma H(A, -b) \, dX, \tag{3-18}$$

for every bounded open set Σ, and $\chi \in W_0^{1,\infty}(\Sigma, \mathbb{R})$, $\theta \in W_0^{1,\infty}(\Sigma, \mathbb{R})$.

Definition 3.4 (Strict Quasiconvexity.). The function H is strictly quasiconvex at (A, b) if and only if H is quasiconvex at (A, b) and equality holds only when $\chi = \theta = 0$.

    The relation between rank-one convexity, quasiconvexity, and other notions of convexity is further discussed in, for example, [Ball 1976] and [Knops and Stuart 1984]. For present purposes, it is sufficient to note that all generalized notions of convexity reduce to the standard condition of convexity when the functions concerned are defined only on scalar and vector quantities.

    We now proceed to establish uniqueness of the solution to the affine boundary problem. We have:

Proposition 3.1 (Uniqueness). *Let Ω be star-shaped with respect to the origin, and let the affine boundary conditions be (3–10) and (3–11). Let $H \in C^2(M^{(n+1)\times n}, \mathbb{R})$ be rank-one convex and strictly quasiconvex at $(A, -b)$. Then the unique smooth equilibrium solution is*

$$x = c + AX, \qquad X \in \bar{\Omega}, \tag{3-19}$$

$$\Phi = d + bX, \qquad X \in \bar{\Omega}. \tag{3-20}$$

Proof. Suppose that (x, Φ) and (y, Ψ) are equilibrium solutions satisfying the affine boundary conditions (3–10) and (3–11) such that $x \not\equiv y \equiv c + AX$ and $\Phi \not\equiv \Psi \equiv d + bX$ for $X \in \Omega$. Consider the volume integral on the left of Equation (2–12), which may be rewritten

$$
\int_\Omega H(\text{Grad } x, -\text{Grad } \Phi) \, dX
$$
$$
= \int_\Omega H\big(\text{Grad } y + \text{Grad}(x - y), -\text{Grad } \Psi + \text{Grad}(\Psi - \Phi)\big) \, dX
$$
$$
= \int_\Omega H\big(A + \text{Grad}(x - y), -b + \text{Grad}(\Psi - \Phi)\big) \, dX. \tag{3–21}
$$

By hypothesis, $x - y = \Psi - \Phi = 0$ for $X \in \partial\Omega$, and consequently strict quasiconvexity of H at $(A, -b)$ implies that

$$
\int_\Omega H(\text{Grad } x, -\text{Grad } \Phi) \, dX > \int_\Omega H(A, -b) \, dX, \tag{3–22}
$$

which contradicts inequality Equation (3–12) and the Proposition is proved. $\square$

4. Concluding Remarks

This final section provides several remarks that supplement the previous results. In particular, we explore the implication of Proposition 3.1 for the uniqueness of the electric displacement and polarization, and consequently the electric free charge density on the surface $\partial\Omega$. For completeness, we also demonstrate for a given deformation gradient that the electric constituents are uniquely determined subject to mixed boundary conditions and a rank-one convex electric enthalpy. The conclusion represents a slight extension of the familiar property in electrostatics.

We commence with an observation whose validity is evident from an examination of the proof of Proposition 3.1.

Remark 4.1 (Alternate conditions). The conditions stipulated in Proposition 3.1 for H may be replaced by the alternative conditions of strict rank-one convexity and quasiconvexity at $(A, -b)$.

Remark 4.2 (Electric displacement and polarization). Suppose for simplicity that the strain energy function W is convex with respect to P so that (2–11) is invertible to give P uniquely in terms of E and F. We conclude that the conditions of Proposition 3.1 uniquely determine P to be constant. Consequently, (2–8) yields a unique constant value for D under the same conditions. Furthermore, the electric free charge surface density $\sigma(X)$ for $X \in \partial\Omega$ is given by

$$
DN = \sigma. \tag{4–1}
$$

where from (2–10) it necessarily follows that $\int_{\partial\Omega} \sigma \, dS = 0$. An appeal to (2–5) and (4–1) shows that σ is uniquely determined by P, E, and F, and therefore under the stipulated conditions is likewise unique.

On the other hand, when the electric potential is constant on the boundary, it follows as a special case of Proposition 3.1 that the electric field E vanishes everywhere in Ω, and by the assumed unique invertibility of (2–11), that the polarization is also identically zero. (See also Remark 4.4).

Remark 4.3 (Mixed boundary conditions: electric field). Suppose that

$$H(F, E) : M^{(n+1)\times n} \to \mathbb{R}$$

is strictly rank-one convex with respect to $E \in \mathbb{R}^n$ at each $F \in M_+^{n\times n}$ so that

$$H(F, tE + (1 - t)Q) < tH(F, E) + (1 - t)H(F, Q), \tag{4-2}$$

for all $E \in \mathbb{R}^n$, $Q \in \mathbb{R}^n$ and $t \in [0, 1]$. Notice, as already observed, that (4–2) is the usual condition for strict convexity as generalized definitions of convexity reduce to the corresponding standard definitions for scalar and vector quantities. Now consider the function defined by

$$I(t) = \int_\Omega \big(H(F, tE + (1 - t)Q) - tH(F, E) - (1 - t)H(F, Q)\big)\, dX, \tag{4-3}$$

which by inspection and (4–2) possesses the properties

$$I(0) = I(1) = 0, \tag{4-4}$$

$$I(t) < 0, \qquad t \in (0, 1). \tag{4-5}$$

Next assume that for each F there exist two distinct electric fields E, Q with potentials Φ, Ψ and corresponding electric displacements $D(F, E)$ and $D(F, Q)$. Instead of the affine boundary condition (3–11) we suppose nonaffine and nonuniform mixed boundary conditions such that for all $F \in M_+^{n\times n}$ and $E \not\equiv Q$ we have:

$$\Phi = \Psi, \qquad X \in \partial\Omega_1, \tag{4-6}$$

$$D(F, E)N = D(F, Q)N, \qquad X \in \partial\Omega_2, \tag{4-7}$$

where $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$, and

$$E = -\operatorname{Grad}\Phi, \qquad Q = -\operatorname{Grad}\Psi, \qquad X \in \Omega. \tag{4-8}$$

Let a superposed prime denote differentiation with respect to t. Examination of the graph of $I(t)$ immediately shows that

$$\begin{aligned}
0 > I'(0) &= \int_\Omega \Big(\frac{\partial H(F, Q)}{\partial E}(E - Q) - H(F, E) + H(F, Q)\Big)\, dX \\
&= \int_\Omega \Big(\frac{\partial H(F, Q)}{\partial E}(\operatorname{Grad}\Psi - \operatorname{Grad}\Phi) - H(F, E) + H(F, Q)\Big)\, dX \\
&= \int_\Omega \Big(-D(F, Q)(\operatorname{Grad}\Psi - \operatorname{Grad}\Phi) - H(F, E) + H(F, Q)\Big)\, dX,
\end{aligned}$$

which after an integration by parts and appeal to Equation (2–10) and (4–6) gives

$$\int_\Omega H(F, E)\, dX > \int_\Omega H(F, Q)\, dX + \int_{\partial\Omega_2} D(F, Q)N(\Phi - \Psi)\, dS. \tag{4-9}$$

On noting Equation (4–7) and either by reversing the roles of E and Q, or by evaluating $I'(1) > 0$, we are led to a contradiction and consequently we conclude that

$$\Phi \equiv \Psi, \qquad X \in \bar{\Omega}, \tag{4–10}$$

and uniqueness of the electric field is established. This is not necessarily constant, unlike the case of the affine boundary value problem.

Remark 4.4 (Mixed boundary conditions: electric displacement and polarization). The conditions introduced into the previous remark are insufficient to provide uniqueness of the corresponding electric displacement and polarization vectors, which is not surprising, especially for ferroelectrics and similar materials. We emphasize, however, that for such materials the argument can easily be modified as follows to additionally obtain uniqueness of the electric displacement and polarization. Assume the conclusion is false and that P and R are the distinct respective polarizations. Let the nonaffine and nonuniform mixed boundary conditions be such that Equations (4–6) and (4–7) hold, and in addition to (rank-one) convexity of the electric enthalpy (4–2) with respect to E, suppose that the strain energy is strictly (rank-one) convex with respect to P in the sense that for each given F

$$W(F, tFP + (1-t)FR) < tW(F, FP) + (1-t)W(F, FR), \tag{4–11}$$

for $P, R \in \mathbb{R}^n$. The function $G(t)$, defined by

$$G(t) =$$
$$\int_\Omega \big(W(F, tFP + (1-t)FR) - tW(F, FP) - (1-t)W(F, FR)\big)\, dX, \tag{4–12}$$

satisfies $G(0) = G(1) = 0$, and $G(t) < 0$, $0 < t < 1$, so that $G'(0) < 0$ and therefore by Equation (2–11) we have

$$\int_\Omega \big(E(P - R) - W(F, FP) + W(F, FR)\big)\, dX < 0. \tag{4–13}$$

But for each F we have shown already that E is uniquely determined, and so by interchange of P and R we are led to a contradiction and the polarization is unique. It is worth remarking that uniqueness of P is established here subject to conditions more general than those assumed in Remark 4.2. Uniqueness of the electric displacement now follows from relation (2–5).

References

[Ball 1976] J. M. Ball, "Convexity conditions and existence theorems in nonlinear elasticity", *Arch. Ration. Mech. An.* **63**:4 (1976), 337–403.

[Chadwick 1975] P. Chadwick, "Applications of an energy–momentum tensor in non-linear elastostatics", *J. Elasticity* **5**:3-4 (1975), 249–258.

[Ericksen 2006, to appear] J. L. Ericksen, "Theory of elastic dielectrics revisited", 2006, to appear. Arch. Ration. Mech. An.

[Eshelby 1975] J. D. Eshelby, "The elastic energy-momentum tensor", *J. Elasticity* **5**:3-4 (1975), 321–335.

[Gelfand and Fomin 1963] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Prentice-Hall, Englewood Cliffs, N. J., 1963.

[Green 1973] A. E. Green, "On some general formulae in finite elastostatics", *Arch. Ration. Mech. An.* **50**:1 (1973), 73–80.

[Günther 1962] W. Günther, "Über einige randintegrale der elastomechanik", *Abh. der Braunschweigishen wiss. Ges.* **14** (1962), 54.

[Gurtin 2000] M. E. Gurtin, *Configurational forces as basic concepts of continuum physics*, Springer, New York, 2000.

[Hadamard 1903] J. Hadamard, *Leçons sur la propagation des ondes*, Hermann, Paris, 1903. Reprinted by Dover 1952.

[Hill 1986] R. Hill, "Energy momentum tensors in elastostatics: some reflections on the general theory", *J. Mech. Phys. Solids* **34**:3 (1986), 305–317.

[Knops and Stuart 1984] R. J. Knops and C. A. Stuart, "Quasiconvexity and uniqueness of equilibrium solutions in nonlinear elasticity", *Arch. Ration. Mech. An.* **86**:3 (1984), 233–249.

[Knowles and Sternberg 1972] J. K. Knowles and E. Sternberg, "On a class of conservation laws in linearized and finite elastostatics", *Arch. Ration. Mech. An.* **44**:3 (1972), 187–211.

[Mareno 2004] A. Mareno, "Uniqueness of equilibrium solutions in second-order gradient nonlinear elasticity", *J. Elasticity* **74**:2 (2004), 99–107.

[Maugin 1993] G. A. Maugin, *Material inhomogeneities in elasticity*, Chapman and Hall, London, 1993.

[Maugin and Epstein 1991] G. A. Maugin and M. Epstein, "The electroelastic energy-momentum tensor", *P. Roy. Soc. Lond. A Mat.* **433**:1888 (1991), 299–312.

[Maugin and Trimarco 1991] G. A. Maugin and C. Trimarco, "Pseudomomentum and material forces in electromagnetic solids", *Int. J. Appl. Electrom. Mat.* **2** (1991), 207–216.

[Maugin and Trimarco 2001] G. A. Maugin and C. Trimarco, "Material mechanics of electromagnetic solids", pp. 129–172 in *Configurational mechanics of materials*, edited by R. Kienzler and G. A. Maugin, CISM Courses and Lectures **427**, Springer, Vienna, 2001.

[Pack and Herrmann 1986] Y. E. Pack and G. Herrmann, "Conservation laws and the material momentum tensor for the elastic dielectric", *Int. J. Eng. Sci.* **24**:8 (1986), 1365–1374.

[Pohozaev 1965] S. I. Pohozaev, "Eigenfunctions of the equation $\Delta u + \lambda f(u) = 0$", *Soviet Mathematics Dokl.* **6** (1965), 1408–1411. English translation ≡ Dokl. Akad. Nauk. SSSR. 165, 33–39.

[Pucci and Serrin 1986] P. Pucci and J. Serrin, "A general variational identity", *Indiana U. Math. J.* **35**:3 (1986), 681–703.

[Serre 2004] D. Serre, "Hyperbolicity of the nonlinear models of Maxwell's equations", *Arch. Ration. Mech. An.* **172**:3 (2004), 309–331.

[Trimarco 2002] C. Trimarco, "A Lagrangian approach to electromagnetic bodies", *Technische Mech.* **22** (2002), 175–180.

[Trimarco 2003] C. Trimarco, "The Green's and Eshelby's identities in generalised continua and in dielectrics", *Theor. Appl. Mech.* **30** (2003), 41–53.

[Walker et al. 1965] J. B. Walker, A. C. Pipkin, and R. C. Rivlin, "Maxwell's equations in a deformed body", *Accad. Naz. Lincei.* **38**:8 (1965), 674–676.

[Yu 1995] Y.-Y. Yu, "On the ordinary, generalized, and pseudo-variational equations of motion in nonlinear elasticity, piezo-electricity, and classsical plate theories", *J. Appl. Mech. (Trans. ASME))* **62** (1995), 471–478.

R. J. KNOPS: r.j.knops@ma.hw.ac.uk
School of Mathematics and Computer Sciences and the Maxwell Institute of Mathematical Sciences, Heriot–Watt University, Edinburgh EH14, Scotland, United Kingdom

C. TRIMARCO: trimarco@dma.unipi.it
Dipartimento di Matematica Applicata, Universitá di Pisa, Via Buonarroti 1, I-56127 Pisa, Italy

TWO-WAY THERMOMECHANICALLY COUPLED MICROMECHANICAL ANALYSIS OF SHAPE MEMORY ALLOY COMPOSITES

JACOB ABOUDI AND YUVAL FREED

A previously established micromechanical model whose capability to analyze and predict the behavior of thermoinelastic fibrous composites with one-way thermomechanical coupling, in which the temperature is prescribed in advance, was verified. This model is extended herein to incorporate two-way thermomechanical coupling effects in thermoinelastic composites. As a result of this generalization, the temperature which is coupled to the mechanical effects, is governed by the energy equation and is induced into the composite's constituents as a result of the application of mechanical loadings. The model is applied to predict the behavior of composites that consist of shape memory alloy fibers embedded in metallic and polymeric matrices. Results exhibit the response of the composites to various types of loading, and the effect of the two-way thermomechanical coupling that induces temperature deviations from reference temperatures at which shape memory and pseudoelasticity effects take place at the fibers.

1. Introduction

Shape memory alloy (SMA) materials undergo phase transformation which is caused by the application of stress and/or change in temperature. At high temperatures, the material behavior is nonlinear and hysteretic, but, at the end of a mechanical loading-unloading cycle, still yields the original stress-strain-free state (pseudoelastic behavior). At lower temperatures, a mechanical loading-unloading results in a residual deformation which can be recovered by a temperature increase (shape memory effect). The latter effect can be utilized to control the behavior of structures in which SMA materials have been embedded.

There are numerous micromechanical models that can predict the overall (macroscopic) behavior of composite materials with embedded shape memory alloy fibers. Examples for such micromechanical investigations are those of Boyd and Lagoudas [1994], Kawai et al. [1999], Carvelli and Taliercio [1999], Song et al. [1999], Kawai [2000], Gilat and Aboudi [2004] and Marfia [2005]. When thermal effects in these investigations are involved, they are treated as in thermal stress problems where a prescribed constant temperature is imposed throughout the composite. The thermomechanical coupling (TMC) in such thermoelastic problems is referred to as one-way coupling because only the mechanical field is affected by the temperature. A micro-macro-structural analysis with one-way TMC was recently employed by Gilat and Aboudi [2006] to investigate the thermal buckling of shape memory reinforced laminated plates.

In thermomechanical problems with two-way TMC, the temperature and mechanical effects are coupled to each other and the energy equation that governs the temperature field variation in the material involves the effect of total strain rate, the effect of inelastic strain rate in metallic materials and the effect

of transformation strain rate in SMA materials. In order to investigate the behavior of the monolithic SMA material with two-way TMC, appropriate constitutive relations and the energy equation need to be established. To this end, Auricchio and Petrini [2004b] presented a free-energy function from which the required two-way TMC constitutive and energy equations were derived. The implementation of these equations requires the development of a computational algorithm based on an implicit time procedure, in conjunction with the radial return method [Simo and Hughes 1998].

The purpose of the present paper is two-fold. First, a micromechanical model, referred to as a *high-fidelity generalized method of cells* (HFGMC) is generalized to incorporate two-way TMC capability. The HFGMC with one-way coupling can predict the behavior of multiphase inelastic composites with periodic microstructure by employing the homogenization technique. Its accuracy and reliability were demonstrated [Aboudi et al. 2002; 2003] by comparisons with analytical solutions that can be established in certain cases and with finite element solutions. The method has been employed also for the prediction of the behavior of viscoelastic-viscoplastic composites [Aboudi 2005], electro-magneto-thermoelastic composites, and composites that are subjected to large deformations, see the recent review by Aboudi [2004] (that includes also references to its predecessor GMC micromechanical model). It should be noted that the HFGMC with one-way TMC has been implemented in the recently developed micromechanics analysis code MAC/GMC by NASA Glenn Research Center, which has many user friendly features and significant flexibility; see [Bednarcyk and Arnold 2002] for the most recent version of its user guides. The predecessor GMC micromechanical model was employed by Williams and Aboudi [1999] to investigate two-way TMC of metal matrix composites.

As a result of the generalization offered by the HFGMC to incorporate the two-way TMC, macroscopic constitutive equations that govern the thermomechanical behavior of the composite are established. These relations are based on the micromechanical derivation of the effective stiffness tensor of the composite as well as the mechanical, thermal and inelastic concentration tensors and scalars. These concentration tensors and scalars are established by the homogenization of the periodic composite, in conjunction with the imposition of the coupled equilibrium and energy equations, and by imposing the continuity of tractions, displacements, heat fluxes and temperatures at the interfaces between the various materials, and by the application of the periodic boundary conditions. The latter conditions ensure that the tractions, displacements, heat fluxes and temperatures are identical at the opposite boundaries of a repeating unit cell that characterizes the periodic composite. These three types of concentration tensors are interrelated due to the TMC effects. In particular, the effective stiffness tensor of the composite involves the mechanical and thermal concentration tensors.

The second purpose of this paper is the investigation of the overall (macroscopic) behavior of composites consisting of continuous SMA fibers embedded in metallic and polymeric matrices with two-way TMC effects. In particular, the induced average temperatures that result from the two-way TMC are computed and presented under various circumstances. To this end, the constitutive two-way modeling of the SMA fibers of Auricchio and Petrini [2004b] is employed.

Results are given for SMA continuous fiber composites with a metallic (aluminum) matrix and polymeric (epoxy) matrix. It is shown that the TMC has little effect on the average stress-strain of the composite, but has a significant effect on the induced temperature that is generated due to the application of mechanical loadings. This effect is due mainly to the term that appears in the coupled energy equation that involves the inelastic strain rate in metallic materials, and to a lesser extent, to the transformation

strain rate in the SMA fibers. These results are given at two reference temperatures at which shape memory and pseudoelasticity effects in the SMA fibers take place.

2. Two-way TMC constitutive equations of the monolithic SMA

The fully coupled thermomechanical equations of the monolithic SMA material were presented by Auricchio and Petrini [2002; 2004b; 2004a], who extended and improved the thermodynamical model of Souza et al. [1998]. These equations are briefly presented below. The strain ϵ is decoupled into a dilatation θ and deviatoric $\mathbf{e}$ parts as follows:

$$\epsilon = \frac{\theta}{3}\mathbf{I} + \mathbf{e}, \tag{1}$$

where $\mathbf{I}$ is the unit tensor. The free energy function ψ is a function of the dilatation strain θ, the deviatoric strain $\mathbf{e}$, the transformation strain $\mathbf{e}^{tr}$ and absolute temperature T, given by

$$\rho\psi(\theta, \mathbf{e}, \mathbf{e}^{tr}, T) = \frac{1}{2}K\theta^2 + G||\mathbf{e} - \mathbf{e}^{tr}||^2 - 3\alpha K\theta(T - T_0) + \beta < (T - M_f) > ||\mathbf{e}^{tr}||$$
$$+ \frac{h}{2}||\mathbf{e}^{tr}||^2 + (u_0 - T\eta_0) + \rho c_v\left[T - T_0 - T log\frac{T}{T_0}\right] + \Upsilon(\mathbf{e}^{tr}), \tag{2}$$

where ρ, K, G, α, T_0, β, M_f, h, and ρc_v are the mass density, bulk modulus, shear modulus, coefficient of thermal expansion, reference temperature, slope of the stress-temperature relation, martensite final temperature, slope of the stress-strain relation during the stress-induced phase transformation and the heat capacity, respectively. In addition, u_0 and η_0 are the internal energy and entropy at the reference state. In Equation (2), $||.||$ and $<.>$ denote the Euclidean norm and the positive part of the argument respectively, and $\Upsilon(\mathbf{e}^{tr})$ is an indicator function defined as

$$\Upsilon(\mathbf{e}^{tr}) = \begin{cases} 0 & ||\mathbf{e}^{tr}|| \le \epsilon_L, \\ \infty & ||\mathbf{e}^{tr}|| > \epsilon_L, \end{cases}$$

where ϵ_L is the norm of the maximum transformation strain reached at the end of the transformation during uniaxial test.

The resulting constitutive relations obtained from Equation (2) are given by

$$p = \frac{\partial\psi}{\partial\theta} = K[\theta - 3\alpha(T - T_0)],$$

$$\mathbf{s} = \frac{\partial\psi}{\partial\mathbf{e}} = 2G(\mathbf{e} - \mathbf{e}^{tr}),$$

$$\eta = -\frac{\partial\psi}{\partial T} = \eta_0 + 3K\alpha\theta - \beta||\mathbf{e}^{tr}||\frac{<T - M_f>}{|T - M_f|} + \rho c_v log\frac{T}{T_0},$$

$$\mathbf{X} = -\frac{\partial\psi}{\partial\mathbf{e}^{tr}} = \mathbf{s} - \left[\beta(T - M_f) + h||\mathbf{e}^{tr}|| + \frac{\partial\Upsilon(\mathbf{e}^{tr})}{\partial||\mathbf{e}^{tr}||}\right]\frac{\partial||\mathbf{e}^{tr}||}{\partial\mathbf{e}^{tr}},$$

where p and $\mathbf{s}$ are the volumetric and deviatoric parts of the stress $\boldsymbol{\sigma}$, η is the entropy and $\mathbf{X}$ is the transformation stress associated with $\mathbf{e}^{tr}$.

The evolution law of $\mathbf{e}^{tr}$ is

$$\dot{\mathbf{e}}^{tr} = \dot{\xi} \frac{\partial F(\mathbf{X})}{\partial \boldsymbol{\sigma}} \tag{3}$$

where the dot denotes a derivative with respect to time t, and F and $\dot{\xi}$ play the roles of limit function and plastic consistent parameter, subject to the Kuhn–Tucker conditions. For symmetric SMA behavior, the function F is given by

$$F(\mathbf{X}) = ||\mathbf{X}|| - R \le 0, \tag{4}$$

where R is the radius of the elastic domain.

The resulting coupled energy equation is given by

$$\rho c_v \dot{T} + \bigtriangledown \cdot \mathbf{q} = \left[\mathbf{X} + T\beta \frac{\mathbf{e}^{tr}}{||\mathbf{e}^{tr}||} \right] : \dot{\mathbf{e}}^{tr} - 3\,\mathrm{T}\,\mathrm{K}\alpha\dot{\theta}, \tag{5}$$

in conjunction with the Fourier law that relates the heat flux $\mathbf{q}$ to the temperature gradient:

$$\mathbf{q} = -k \, \bigtriangledown T, \tag{6}$$

with k being the heat conductivity.

The above nonlinear equations have been treated by Auricchio and Petrini [2002; 2004b; 2004a] who employed for their solution a computational algorithm based on an implicit time procedure. This algorithm consists of integrating these equations over a time interval $[t_n, t_{n+1}]$ using an implicit backward Euler scheme. Thus, assuming the knowledge of the solution at time t_n, as well as the strain $\boldsymbol{\epsilon}$ at time t_{n+1}, the stresses are computed using the radial return mapping algorithm [Simo and Hughes 1998].

It should be noted that in the case of standard isotropic materials (for example, metallic materials) the energy equation is of a form similar to Equation (5) [Allen 1991], but with the term

$$\left[\mathbf{X} + T\beta \frac{\mathbf{e}^{tr}}{||\mathbf{e}^{tr}||} \right] : \dot{\mathbf{e}}^{tr}$$

replaced by the rate of inelastic work: $\dot{W}_I = \boldsymbol{\sigma} : \dot{\boldsymbol{\epsilon}}^I$, where $\boldsymbol{\epsilon}^I$ is the inelastic strain which replaces the transformation strain $\boldsymbol{\epsilon}^{tr}$ of the SMA material. In addition, $\dot{W}_I$ is usually multiplied by a partition factor ζ to indicate that only a portion of the inelastic work (about 90%) is transformed into heat [Hunter 1983]. Thus, the final form of the energy equation for conventional isotropic materials is given by

$$\rho c_v \dot{T} + \bigtriangledown \cdot \mathbf{q} = \zeta \boldsymbol{\sigma} : \dot{\boldsymbol{\epsilon}}^I - 3\,\mathrm{T}\,\mathrm{K}\alpha\dot{\theta}. \tag{7}$$

As it is shown in the following section, the spatial derivatives can be eliminated. As a result, Equations (5) and (7) are reduced to an ordinary differential equation in time. Consequently, let us represent (5) and (7) in the following compact form

$$[M]\dot{T} = [S]T + [\dot{Q}], \tag{8}$$

where $[\dot{Q}]$ denotes the right side of Equations (5) and (7). The implicit difference in time of Equation (8) yields [Mitchell and Griffiths 1980]

$$\left\{ [M] - \omega \Delta t [S] \right\} T^{n+1} = \left\{ [M] + (1-\omega) \Delta t [S] \right\} T^n + [Q]^n - [Q]^{n-1}, \tag{9}$$
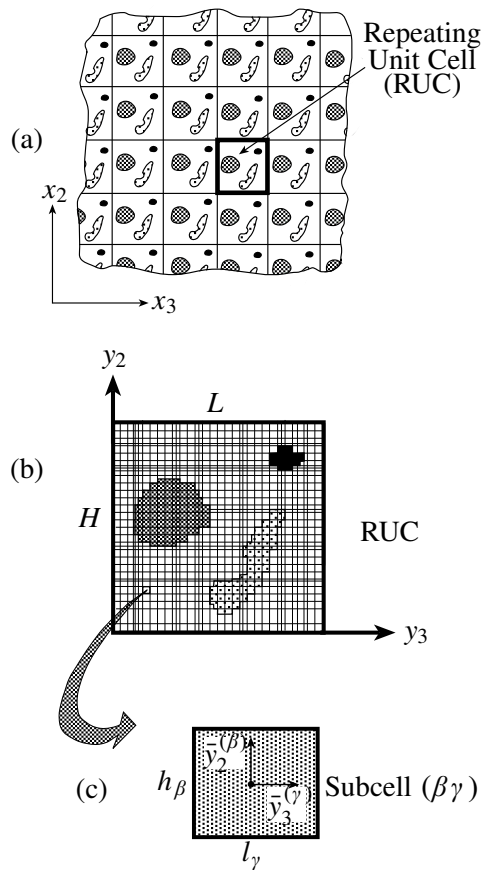
Figure 1. (a) A multiphase composite with doubly-periodic microstructures defined with respect to global coordinates (x_2, x_3). (b) The repeating unit cell is represented with respect to local coordinates (y_2, y_3). It is divided into N_β and N_γ subcells, in the y_2 and y_3 directions, respectively. (c) A characteristic subcell $(\beta\gamma)$ with local coordinates $\bar{y}_2^{(\beta)}$ and $\bar{y}_3^{(\gamma)}$ whose origin is located at its center.

where $\Delta t = t_{n+1} - t_n$ and ω is a parameter (for the Crank–Nicolson scheme: $\omega = 1/2$).

3. Two-way TMC micromechanical analysis

The HFGMC micromechanical model, which is extended herein to a two-way TMC, is used to predict the fully coupled thermoinelastic behavior of doubly periodic composites with SMA continuous fibers. For doubly periodic elastoplastic composites with one-way TMC, this theory has been fully described by [Aboudi et al. 2002; 2003] and by [Bednarcyk et al. 2004] for elastoplastic composites with imperfect bonding between the constituents. In these publications, the reliability and accuracy of the microme-chanical predictions were examined by comparisons with analytical solutions that can be established in some special cases and with a finite element procedure. In this paper, this micromechanical model with two-way TMC is briefly outlined.

This model is based on a homogenization technique for composites with periodic microstructure as shown in Figure 1(a) in terms of the global coordinates (x_2, x_3). The repeating unit cell, Figure 1(b), defined with respect to local coordinates (y_2, y_3), of such a composite is divided into N_β and N_γ subcells in the y_2 and y_3 directions. Each subcell is labeled by the indices $(\beta\gamma)$ with $\beta = 1, ..., N_\beta$ and $\gamma = 1, ..., N_\gamma$, and may contain a distinct homogeneous material. The dimensions of subcell $(\beta\gamma)$ in the y_2 and y_3 directions are denoted by h_β and l_γ. A local coordinate system $(\bar{y}_2^{(\beta)}, \bar{y}_3^{(\gamma)})$ is introduced in each subcell whose origin is located at its center, see Figure 1(c).

The local (subcell) constitutive equation of the material which, in general, is assumed to be thermoinelastic is given by

$$\boldsymbol{\sigma}^{(\beta\gamma)} = \mathbf{C}^{(\beta\gamma)}(\boldsymbol{\epsilon}^{(\beta\gamma)} - \boldsymbol{\epsilon}^{I(\beta\gamma)}) - \Gamma^{(\beta\gamma)}\Delta T^{(\beta\gamma)} \tag{10}$$

where $\boldsymbol{\sigma}^{(\beta\gamma)}$, $\boldsymbol{\epsilon}^{(\beta\gamma)}$, $\boldsymbol{\epsilon}^{I(\beta\gamma)}$ and $\Gamma^{(\beta\gamma)}$ are the stress, total strain, inelastic strain and thermal stress coefficients tensors in subcell $(\beta\gamma)$. In Equation (10), $\mathbf{C}^{(\beta\gamma)}$ is the stiffness tensor of the material in the subcell $(\beta\gamma)$, and $\Delta T^{(\beta\gamma)}$ denotes the temperature deviation from a reference temperature. The inelastic strain $\boldsymbol{\epsilon}^{I(\beta\gamma)}$ is governed by the flow rule (3) for SMA materials, by the Prandtl–Reuss equations of the classical plasticity or by an appropriate viscoplastic flow rule.

The basic assumption in HFGMC with one-way TMC is that the displacement vector $\mathbf{u}^{(\beta\gamma)}$ in each subcell is expanded into quadratic forms in terms of its local coordinates $(\bar{y}_2^{(\beta)}, \bar{y}_3^{(\gamma)})$, as follows

$$\mathbf{u}^{(\beta\gamma)} = \bar{\boldsymbol{\epsilon}} \cdot \mathbf{x} + \mathbf{W}_{(00)}^{(\beta\gamma)} + \bar{y}_2^{(\beta)}\mathbf{W}_{(10)}^{(\beta\gamma)} + \bar{y}_3^{(\gamma)}\mathbf{W}_{(01)}^{(\beta\gamma)}$$
$$+ \frac{1}{2}\left(3\bar{y}_2^{(\beta)2} - \frac{h_\beta^2}{4}\right)\mathbf{W}_{(20)}^{(\beta\gamma)} + \frac{1}{2}\left(3\bar{y}_3^{(\gamma)2} - \frac{l_\gamma^2}{4}\right)\mathbf{W}_{(02)}^{(\beta\gamma)} \tag{11}$$

where $\bar{\boldsymbol{\epsilon}}$ is the externally applied average strain, $\mathbf{W}_{(00)}^{(\beta\gamma)}$ is the volume-averaged displacement, and the higher-order terms

$$\mathbf{W}_{(mn)}^{(\beta\gamma)}$$

must be determined as discussed below.

In the two-way thermomechanically coupled HFGMC, the unknown temperature deviation $\Delta T^{(\beta\gamma)}$ in the subcell is also expanded as follows

$$\Delta T^{(\beta\gamma)} = \Delta T_{(00)}^{(\beta\gamma)} + \bar{y}_2^{(\beta)}\Delta T_{(10)}^{(\beta\gamma)} + \bar{y}_3^{(\gamma)}\Delta T_{(01)}^{(\beta\gamma)}$$
$$+ \frac{1}{2}\left(3\bar{y}_2^{(\beta)2} - \frac{h_\beta^2}{4}\right)\Delta T_{(20)}^{(\beta\gamma)} + \frac{1}{2}\left(3\bar{y}_3^{(\gamma)2} - \frac{l_\gamma^2}{4}\right)\Delta T_{(02)}^{(\beta\gamma)}, \tag{12}$$

where $\Delta T_{(00)}^{(\beta\gamma)}$ is the volume-averaged temperature and the higher-order terms $\Delta T_{(mn)}^{(\beta\gamma)}$ are additional unknowns.

The unknown terms

$$\mathbf{W}_{(mn)}^{(\beta\gamma)} \quad \text{and} \quad \Delta T_{(mn)}^{(\beta\gamma)}$$

are determined from the fulfillment of the coupled equilibrium and energy equations, the periodic boundary conditions, and the interfacial continuity conditions of displacements, tractions, temperatures and heat fluxes between subcells. The periodic boundary conditions ensure that the displacements, tractions, temperatures and heat fluxes at opposite surfaces of the repeating unit cell (that is, at $y_2 = 0$ and H as

well as $y_3 = 0$ and L) are identical, see Aboudi [2004] for more details pertaining to the micromechanical analysis with one-way TMC. A requisite in the present micromechanical analysis is that all these conditions are imposed in the average (integral) sense.

As a result of the imposition of these conditions, a linear system of algebraic equations at the current time step is obtained which can be represented in the following form

$$\mathbf{KU} = \mathbf{f} + \mathbf{g}, \tag{13}$$

where the matrix $\mathbf{K}$ contains information on the geometry and thermomechanical properties of the materials within the individual subcells $(\beta\gamma)$, and the displacement-temperature vector $\mathbf{U}$ contains the unknown displacement and temperature coefficients:

$$\mathbf{U} = [\mathbf{U}^{(11)}, \ldots, \mathbf{U}^{(N_\beta N_\gamma)}], \tag{14}$$

where in subcell $(\beta\gamma)$ these coefficients, which appear on the right side of Equations (11)–(12), are

$$\mathbf{U}^{(\beta\gamma)} = (\mathbf{W}_{(00)}, \Delta T_{(00)}, \mathbf{W}_{(10)}, \Delta T_{(10)}, \mathbf{W}_{(01)}, \Delta T_{(01)}, \mathbf{W}_{(20)}, \Delta T_{(20)}, \mathbf{W}_{(02)}, \Delta T_{(02)})^{(\beta\gamma)}. \tag{15}$$

The mechanical vector $\mathbf{f}$ contains information on the applied average (far-field) strains $\bar{\epsilon}$. The inelastic force vector $\mathbf{g}$ appearing on the right side of Equation (13) contains the inelastic effects given in terms of the integrals of the inelastic strain distributions. These integrals depend implicitly on the elements of the displacement-temperature coefficient vector $\mathbf{U}$, requiring an incremental procedure of Equation (13) at each point along the loading path, see [Aboudi et al. 2003] for more details.

Due to the dependence of the elements of $\mathbf{K}$ on the temperature, it is necessary to invert this matrix at every time step. The solution of Equation (13) at a given time step yields the following localization expression which relates the average strain $\bar{\epsilon}^{(\beta\gamma)}$ and temperature $\Delta\bar{T}^{(\beta\gamma)}$ in the subcell $(\beta\gamma)$ to the externally applied average strain $\bar{\epsilon}$ in the form:

$$\left\{ \begin{array}{c} \bar{\epsilon}^{(\beta\gamma)} \\ \Delta\bar{T}^{(\beta\gamma)} \end{array} \right\} = \left\{ \begin{array}{c} \mathbf{A}^{M(\beta\gamma)} \\ \mathbf{A}^{T(\beta\gamma)} \end{array} \right\} \bar{\epsilon} + \left\{ \begin{array}{c} \mathbf{V}^{T(\beta\gamma)} \\ v^{T(\beta\gamma)} \end{array} \right\} + \left\{ \begin{array}{c} \mathbf{V}^{I(\beta\gamma)} \\ v^{I(\beta\gamma)} \end{array} \right\}, \tag{16}$$

where $\mathbf{A}^{M(\beta\gamma)}$ and $\mathbf{A}^{T(\beta\gamma)}$ are the mechanical and thermal concentration tensors of the subcell $(\beta\gamma)$, $\mathbf{V}^{T(\beta\gamma)}$ and $\mathbf{V}^{I(\beta\gamma)}$ are second-order tensors that involve thermal and inelastic effects in the subcell, and $v^{T(\beta\gamma)}$ and $v^{I(\beta\gamma)}$ are the corresponding scalars. These second-order thermal and inelastic tensors and the corresponding scalars arise due to the existence of $[Q]^{n-1}$ in Equation (9) at the previous time step. It should be noted that in the present case of two-way TMC, the application of the far-field strain $\bar{\epsilon}$ induces a temperature deviation from the reference temperature $\Delta T^{(\beta\gamma)}$ in the subcell.

In order to establish the global (macroscopic) constitutive equation of the composite, we use the definition of the average stress in the composite in terms of average stress in the subcells:

$$\bar{\sigma} = \frac{1}{HL} \sum_{\beta=1}^{N_\beta} \sum_{\gamma=1}^{N_\gamma} h_\beta l_\gamma \bar{\sigma}^{(\beta\gamma)}, \tag{17}$$

where $\bar{\sigma}^{(\beta\gamma)}$ is the average stress in the subcell. By substituting Equation (10) and (16) into (17), one obtains the final form of the effective constitutive law of the multiphase fully coupled thermo-inelastic

Property	Value
E	70 GPa
ν	0.33
k	18 W/(m K)
α	1×10^{-6}/K
ρc_v	5.44 MJ/(m^3 K)
h	500 MPa
R	45 MPa
β	7.5 MPa/K
ϵ_L	0.03
M_f	253.15 K

Table 1. Material properties of the SMA fibers [Auricchio and Petrini 2002]. E, ν denote the Young's modulus and Poisson's ratio. The other parameters were already defined in Section 2.

composite, which relates the average stress $\bar{\boldsymbol{\sigma}}$, strain $\bar{\boldsymbol{\epsilon}}$, thermal stress $\bar{\boldsymbol{\sigma}}^T$ and inelastic stress $\bar{\boldsymbol{\sigma}}^I$ as follows

$$\bar{\boldsymbol{\sigma}} = \mathbf{C}^* \, \bar{\boldsymbol{\epsilon}} - (\bar{\boldsymbol{\sigma}}^T + \bar{\boldsymbol{\sigma}}^I). \tag{18}$$

In this equation $\mathbf{C}^*$ is the effective stiffness tensor which is given by

$$\mathbf{C}^* = \frac{1}{HL} \sum_{\beta=1}^{N_\beta} \sum_{\gamma=1}^{N_\gamma} h_\beta l_\gamma \left(\mathbf{C}^{(\beta\gamma)} \mathbf{A}^{M(\beta\gamma)} - \Gamma^{(\beta\gamma)} \mathbf{A}^{T(\beta\gamma)} \right). \tag{19}$$

The global thermal stress $\bar{\boldsymbol{\sigma}}^T$ is determined from

$$\bar{\boldsymbol{\sigma}}^T = -\frac{1}{HL} \sum_{\beta=1}^{N_\beta} \sum_{\gamma=1}^{N_\gamma} h_\beta l_\gamma \left(\mathbf{C}^{(\beta\gamma)} \mathbf{V}^{T(\beta\gamma)} - \Gamma^{(\beta\gamma)} v^{T(\beta\gamma)} \right). \tag{20}$$

The global inelastic stress $\bar{\boldsymbol{\sigma}}^I$ is of the form

$$\bar{\boldsymbol{\sigma}}^I = -\frac{1}{HL} \sum_{\beta=1}^{N_\beta} \sum_{\gamma=1}^{N_\gamma} h_\beta l_\gamma \left(\mathbf{C}^{(\beta\gamma)} \mathbf{V}^{I(\beta\gamma)} - \Gamma^{(\beta\gamma)} v^{I(\beta\gamma)} - \bar{\boldsymbol{\sigma}}^{I(\beta\gamma)} \right), \tag{21}$$

where the inelastic stress in the subcell is: $\bar{\boldsymbol{\sigma}}^{I(\beta\gamma)} = \mathbf{C}^{(\beta\gamma)} \bar{\boldsymbol{\epsilon}}^{I(\beta\gamma)}$. Finally, the global average temperature $\Delta \bar{T}$ in the composite can be determined from a simple averaging of all temperature deviations $\Delta \bar{T}^{(\beta\gamma)}$ in the subcells.

4. Applications

The established two-way TMC micromechanics is applied herein to predict the behavior of SMA composites in various circumstances. Two types of matrices are chosen to illustrate the responses of the

E (GPa)	ν	$\alpha(10^{-6}/\text{K})$	σ_y (MPa)	E_s (GPa)	$k(W/(\text{mK}))$	$\rho c_v(MJ/(\text{m}^3\text{K}))$
72.4	0.33	22.5	371.5	23	116.7	2.25

Table 2. Elastic, plastic and thermal parameters of the isotropic elastoplastic aluminum matrix. E, ν, α, σ_y, E_s, k and ρc_v denote the Young's modulus, Poisson's ratio, coefficient of thermal expansion, yield stress, secondary modulus, thermal conduction and heat capacity, respectively.

E(GPa)	ν	$\alpha(10^{-6}/\text{K})$	$k(W/(\text{mK}))$	$\rho c_v(MJ/(\text{m}^3\text{K}))$
3.45	0.35	54	0.18	1.28

Table 3. Elastic and thermal parameters of the isotropic epoxy polymeric matrix. E, ν, α, k and ρc_v denote the Young's modulus, Poisson's ratio, coefficient of thermal expansion, thermal conduction and heat capacity, respectively.

composites. In the first case an aluminum alloy is chosen as a matrix. In this case the inelasticity of the metallic matrix and its two-way TMC are involved (see Equation (7)) in addition to the two-way TMC of the SMA fibers (see Equation (5)). In the second type, a polymeric matrix (epoxy) is chosen. Here, the TMC arises due to the existence of the volumetric strain rate $\dot{\theta}$ in Equation (7) only (since $\dot{W}_I = 0$). The material properties of the SMA fibers and the aluminum and epoxy matrices are given in Tables 1–3. The present results were obtained by dividing the repeating unit cell, Figure 1(b), into $N_\beta = 3$ and $N_\gamma = 3$ subcells with the SMA fiber occupying the central subcell. In all cases the volume fraction of the SMA fibers is 0.3.

4.1. *Monolithic SMA behavior.* The uniaxial stress-strain behavior of the monolithic SMA fibers is shown in Figures 2 and 3 at reference temperatures $T_0 = 253.15$ K and $T_0 = 285.15$ K. At the first temperature the material is stable in the martensitic phase and the shape memory effect takes place: the mechanical loading-unloading results in a residual deformation which can be recovered by a temperature increase. At the second temperature the material is stable in the austenitic phase and the pseudoelasticity effect takes place: the material behavior is nonlinear and hysteretic but yields the original stress-strain-free state at the end of the mechanical loading-unloading. The figures show also the induced temperature deviations due to the TMC. These temperatures are caused mainly by the existence of the term

$$\left(\mathbf{X} + T\beta \frac{\mathbf{e}^{tr}}{\|\mathbf{e}^{tr}\|} \right) : \dot{\mathbf{e}}^{tr}$$

in the energy Equation (5) which is the counterpart of the rate of the inelastic work $\dot{W}_I$ in Equation (7). Indeed, the standard thermomechanical coupling coefficient

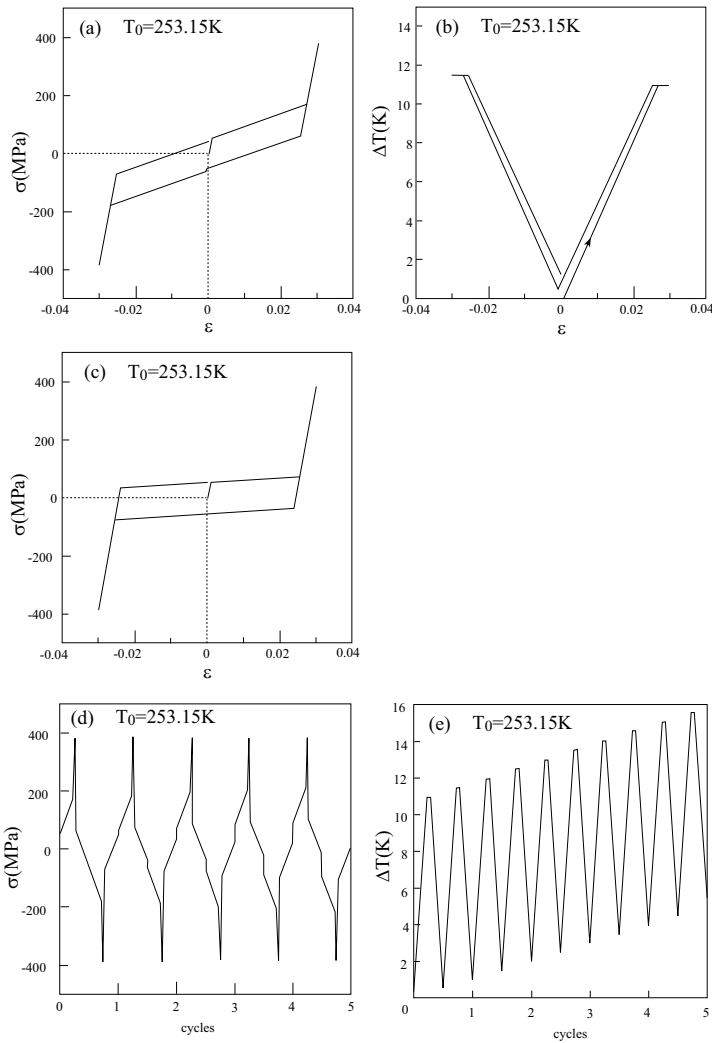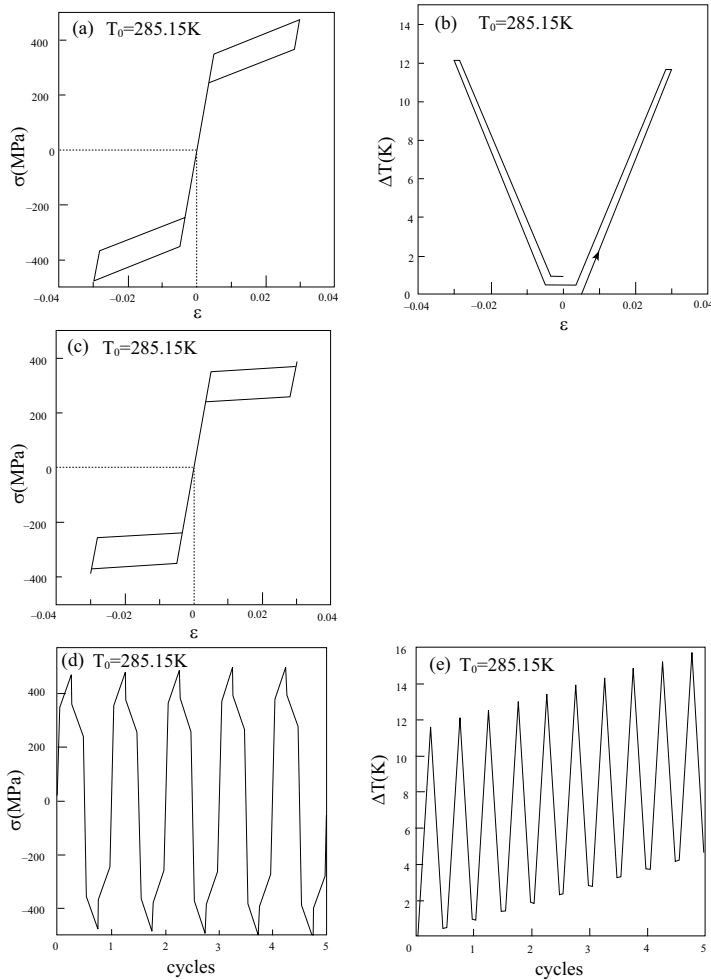$$\delta = \frac{E(1+\nu)\alpha^2 T_0}{(1-2\nu)(1-\nu)\rho c_v}, \tag{22}$$

Figure 2. (a) The uniaxial stress-strain response in one cycle of SMA and, (b) the corresponding induced temperature deviation. (c) The corresponding uniaxial stress-strain response of SMA when the TMC is neglected. (d) The uniaxial stress-strain response in five cycles of SMA and (e) the corresponding induced temperature deviation. In all cases the reference temperature is $T_0 = 253.15$ K.

that characterizes the amount of coupling in thermoelastic material (that is, the coupling that is caused by the term $3 \, \mathrm{T} \, \mathrm{K} \alpha \dot{\theta}$ in Equation (5)) is equal to 0.2×10^{-4} for the considered SMA material, which is very small. The rate effects in the present SMA material are negligibly small. Figures 2(c) and 3(c) (which coincide with those given by Auricchio and Petrini [2002]) show the SMA response at these reference temperatures when the TMC is neglected. It is well observed that the effect of TMC on the SMA behavior is significant. This is reflected in the change of the stress-strain slope when the phase transformation takes place due to the induced temperature changes. Figures 2(b) and 3(b) show that there

Figure 3. (a) The uniaxial stress-strain response in one cycle of SMA and, (b) the corresponding induced temperature deviation. (c) The corresponding uniaxial stress-strain response of SMA when the TMC is neglected. (d) The uniaxial stress-strain response in five cycles of SMA and (e) the corresponding induced temperature deviation. In all cases the reference temperature is $T_0 = 285.15$ K.

is a residual increase in the temperature (about 1 K) upon a complete uniaxial stress loading-unloading cycle of the homogeneous SMA material (where neither convection nor adiabatic conditions exist). It should be noted that the effect of the TMC on the response of the SMA that is depicted in Figure 2(a) and 3(a), as compared to Figure 2(c) and 3(c), results from the induced temperatures that affect the phase transformation and not by the generated thermal strains. For the reference temperature $T_0 = 285.15$ K, for example, the transformation strain at $\epsilon = 3\%$ is equal to 0.023 whereas the corresponding thermal strain is 0.12×10^{-4}.
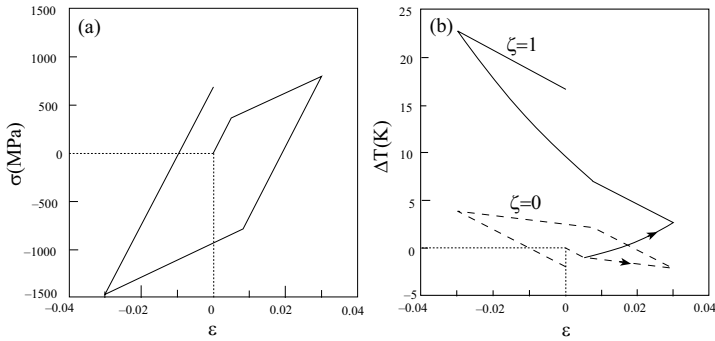
Figure 4. (a) The uniaxial stress-strain response in one cycle of the unreinforced aluminum and, (b) the resulting induced temperature deviation when $\zeta = 1$ and 0 in Equation (7).

Let us investigate the behavior of the SMA material that is subjected to a uniaxial stress cyclic loading. To this end, the SMA fiber at reference temperatures $T_0 = 253.15$ K and 285.15 K is subjected to five cycles of strain loading-unloading, and the resulting stress response and induced temperature are shown in Figure 2(d, e) and Figure 3(d, e). The stress shows a repetitive behavior but the continuously rising heat generation can be clearly observed.

4.2. *SMA/aluminum composite.* In order to investigate the behavior of the elastoplastic work-hardening aluminum matrix that has been characterized in Table 2, the uniaxial stress response of the (unreinforced) material to a cyclic loading is shown in Figure 4(a). The thermomechanical coupling coefficient δ that is given by Equation (22) is: $\delta = 0.027$. Therefore it is instructive to exhibit the induced temperature deviation during this cyclic loading in two different cases. To this end, Figure 4(b) shows the induced temperature when in Equation (7) $\zeta = 1$ and $\zeta = 0$. In the first case full TMC is taken into account, while in the second case the heat generated by the rate of plastic work $\dot{W}_I$ is neglected while retaining the coupling caused by the term $3 \, T \, K\alpha\dot{\theta}$ in this equation (which is accounted for by the coupling coefficient δ). Figure 4(b) clearly shows that the heat generated by the rate of plastic work is predominant. Except in Figure 4(b), ζ is taken in all cases to be equal to 1. The effect of TMC on the stress-strain response is negligibly small. Indeed, under a uniaxial stress loading at a strain of $\epsilon = 3\%$, the resulting inelastic and thermal strains are given by 0.019 and 0.56×10^{-4}, respectively.

Let us consider presently a composite in which the aluminum matrix is reinforced by SMA fibers that are oriented in the 1-direction. Figures 5 and 6 show the average uniaxial stress response $\bar{\sigma}_{11}$ to a cyclic loading of the composite in the axial direction and the resulting induced global temperature deviations $\Delta\bar{T}$ at two reference temperatures:

$$T_0 = 253.15 \, \text{K} \quad \text{and} \quad 285.15 \, \text{K}.$$

Although the curves in these figures appear to be quite similar, some differences can be detected which shows the effects of the SMA behavior at these two different reference temperatures. A comparison of these figures with the response of the unreinforced aluminum (that has been exhibited in Figure 4) provides further indications about the influence of the SMA fibers. The unreinforced aluminum is
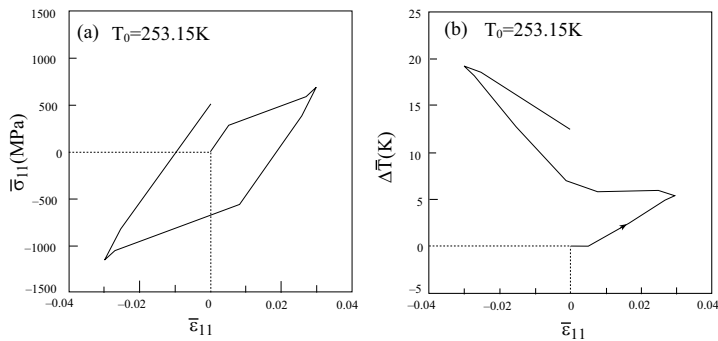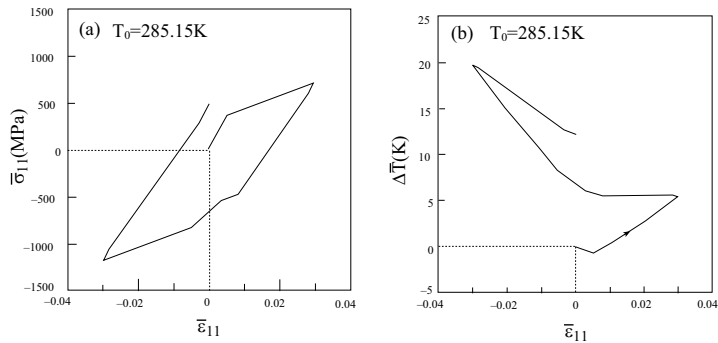
Figure 5. (a) The uniaxial stress-strain response in one cycle of the SMA/aluminum composite loaded in the fibers direction and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 253.15$ K.

Figure 6. (a) The uniaxial stress-strain response in one cycle of the SMA/aluminum composite loaded in the fibers direction and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 285.15$ K.

modeled as an elastic-plastic material with linear hardening, see Figure 4(a). A comparison of Figure 4(a) with Figure 5(a) and 6(a) shows that due to the existence of the SMA fibers, the linear hardening behavior is lost. Here too, just like the unreinforced aluminum, the effect of the TMC on the global stress-strain response of the SMA/aluminum composite is negligible.

It is interesting that the maximum induced temperature by the TMC in the unreinforced aluminum is greater than that obtained in the SMA/aluminum composite. This observation can be explained by the fact that the maximum temperature generated in the monolithic SMA is about 12 K (see Figure 2(b) and 3(b)), while in the monolithic aluminum it is about 23 K (see Figure 4(b)). The reinforcement by SMA fibers results in a decrease of the dominant effect of the aluminum which decreases the global temperature that is generated in the composite by the TMC.

The Young's moduli of the SMA fibers and the aluminum matrix are quite close to each other ($70 MPa$ and $72.4 MPa$, respectively). Consequently the response of the SMA/aluminum in the transverse 2-direction is quite similar to its response in the axial direction (shown in Figures 5 and 6).
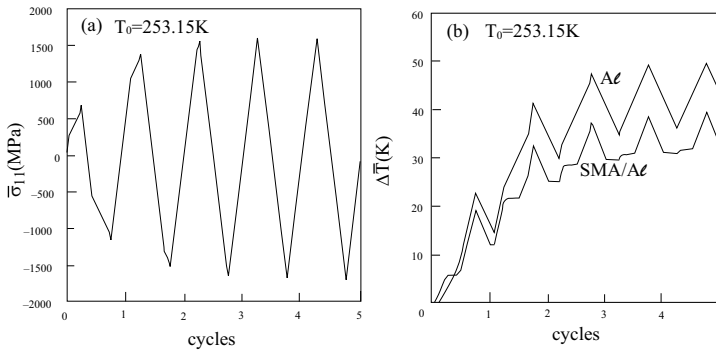
Figure 7. (a) The uniaxial response of the SMA/aluminum composite in 5 cycles loaded in the fibers direction and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 253.15$ K. Also shown is the corresponding induced temperature deviation in the unreinforced aluminum.
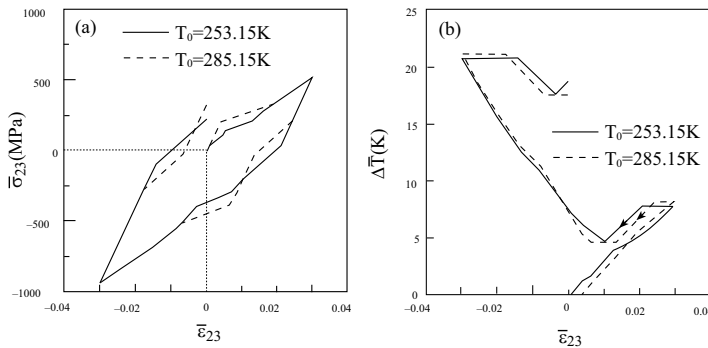

Figure 8. (a) The transverse shear stress-strain response in one cycle of the SMA/aluminum composite and, (b) the resulting induced average temperature deviation at reference temperatures: $T_0 = 253.15$ K and 285.15 K.

A further investigation of the induced temperature due to the two-way TMC can be performed by subjecting the SMA/aluminum composite to 5 cycles of uniaxial stress loading-unloading in the fibers direction at a reference temperature $T_0 = 253.15$ K. The resulting average axial stress response of the composite and the induced average temperature deviation are shown in Figure 7. The corresponding response to 5 cyclic loadings of the unreinforced aluminum matrix is quite similar (but with higher amplitude) to the composite response that is exhibited in Figure 7(a) and is not shown. The induced temperature of the unreinforced aluminum, on the other hand, is shown in Figure 7(b). Figure 7(b) reveals, as expected, the effect of TMC caused by the inelastic effects that appear in Equation (5) for the SMA fibers and Equation (7) for the aluminum. The latter effect appears to be predominant.

As a final illustration of the TMC effects in SMA/aluminum, we consider a cyclic transverse shear loading of the SMA/aluminum composite. Here average transverse shear stress-strain responses $\bar{\sigma}_{23} - \bar{\epsilon}_{23}$
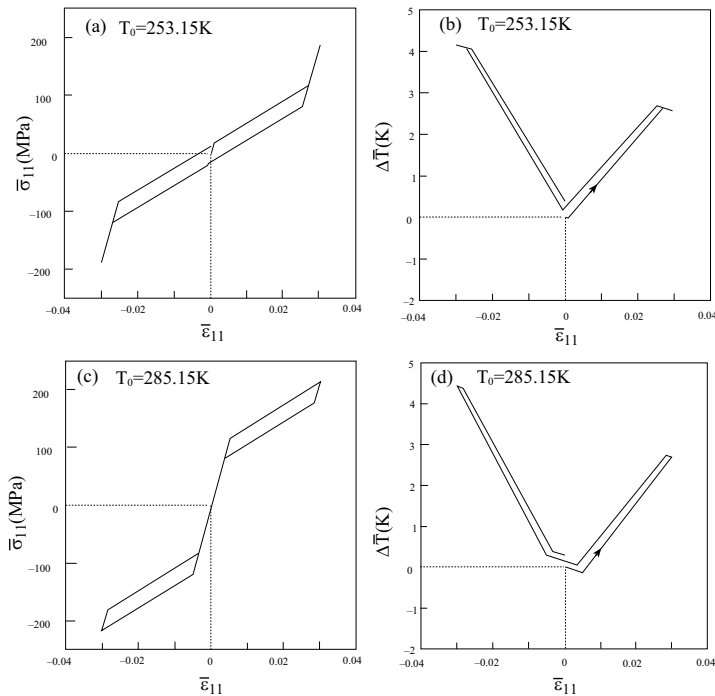
Figure 9. (a) The uniaxial stress-strain response in one cycle of the SMA/epoxy composite loaded in the fibers direction and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 253.15$ K. (c) The uniaxial stress-strain response in one cycle of the SMA/epoxy composite loaded in the fibers direction and, (d) the resulting induced average temperature deviation at a reference temperature $T_0 = 285.15$ K.

of the composite at two reference temperatures:

$$T_0 = 253.15 \, \text{K} \quad \text{and} \quad 285.15 \, \text{K}$$

are shown in Figure 8. In addition, the resulting induced global temperature deviations $\Delta \bar{T}$ due to the TMC effects are also compared. Here too, the induced temperature is generated mainly by the rate of plastic work of the aluminum matrix. It should be noted that under a pure shear loading, the TMCs which cause the generation of temperatures in the monolithic SMA and the unreinforced aluminum are due to the first terms that represent the inelastic effects in the right side of Equation (5) and (7). The second term $3 \, T \, K\alpha\dot{\theta}$ in both equations vanishes in this case.

4.3. *SMA/epoxy composite.* Consider the epoxy matrix whose material parameters are given in Table 3. Its thermomechanical coupling coefficient is quite small: $\delta = 0.016$. Consequently, due to the absence of inelastic effects in this polymer, it is expected that the effect of TMC will be weak and the induced temperature deviation will not be significant. In Figure 9, the uniaxial response to a cyclic loading in the
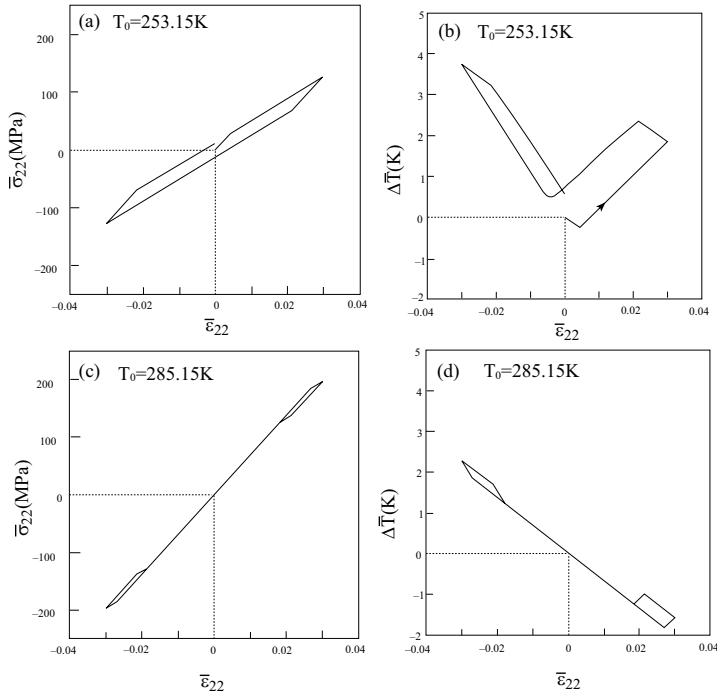
Figure 10. (a) The uniaxial stress-strain response in one cycle of the SMA/epoxy composite loaded in the perpendicular direction of to the fibers and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 253.15$ K. (c) The uniaxial stress-strain response in one cycle of the SMA/epoxy composite loaded in the perpendicular direction of to the fibers and, (d) the resulting induced average temperature deviation at a reference temperature $T_0 = 285.15$ K.

fibers' direction of SMA/epoxy composite is shown at two reference temperatures:

$$T_0 = 253.15 \text{ K} \quad \text{and} \quad T_0 = 285.15\text{K}.$$

Also shown are the resulting average temperatures that are generated due to the TMC effect. The figure shows that quite different stress-strain behaviors are obtained by loading-unloading of the composite at the two reference temperatures, but the induced temperatures are very similar. The Young's moduli of the SMA fiber and epoxy matrix are quite different ($70GPa$ and $3.45GPa$, respectively). Thus the uniaxial responses of the composite to loading in the axial direction (that is, parallel to the fibers) and transverse direction (that is, perpendicular to the fibers) should be quite different. This difference can be observed by comparing Figure 9 with Figure 10 which exhibits the transverse response of the SMA/epoxy composite at the same reference temperatures.

A careful check of the temperatures that are generated by a uniaxial stress loading-unloading of the pure SMA, Figures 2(b) and 3(b), reveal that there is asymmetry that can be clearly observed at the final strain of $\epsilon = \pm 3\%$ at which the temperature deviations ΔT are not equal. The combined effects of the (asymmetric) temperature generated by loading-unloading cycle of the SMA fibers and the (symmetric)
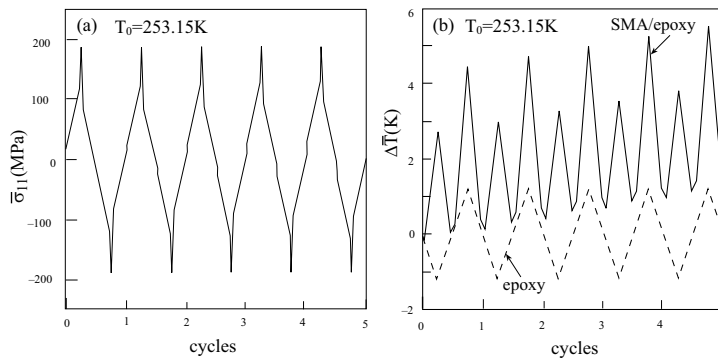
Figure 11. (a) The uniaxial response of the SMA/epoxy composite in 5 cycles loaded in the fibers direction and, (b) the resulting induced average temperature deviation at a reference temperature $T_0 = 253.15$ K. Also shown is the corresponding induced temperature deviation in the unreinforced epoxy.

temperature generated in the epoxy matrix that are shown in Figures 9(b), 9(d), 10(b) and 10(d) at a finer scale, exhibit the resulting asymmetries in the global induced temperatures of the SMA/epoxy composite.

As a final illustration of the SMA/epoxy composite behavior, we present in Figure 11 the response of the composite to 5 cycles of uniaxial stress loading in the fibers direction at a reference temperature

$$T_0 = 253.15 \, \text{K}.$$

Of particular interest is the uniform periodic temperature that is induced in the unreinforced epoxy caused by the TMC (in the absence of any irreversible effects). This is in contrast to the nonuniform and continuously increasing temperature in the monolithic SMA fiber that was shown in Figure 2(e) which is caused by the irreversible effects in this material. As a result, the induced temperature $\Delta \bar{T}$ caused by the TMC effects in the SMA/epoxy composite exhibits an irregular pattern as shown in Figure 11(b).

We conclude the paper by mentioning that by subjecting the SMA/epoxy composite to a transverse shear loading, the amount of the induced temperature is very small implying that the effect of TMC is negligibly small in this case.

5. Conclusions

A micromechanical model which can predict the behavior of multiphase inelastic composites with one-way TMC has been generalized to incorporate two-way TMC effects in the presence of SMA phases. The generalization results in coupled mechanical and energy equations in which the temperature is induced in the phases as a result of the application of far field mechanical loadings. Overall macroscopic thermoinelastic constitutive equations of composites with two-way TMC have been established. They represent the global behavior of the SMA fiber composite at every increment of loading. The results show that two-way TMC must be taken into account when dealing with SMA/metallic composites, and that it has a minor effect in SMA/polymeric composites. This is due to the inelastic effects that generate heat in the metallic phase in addition to the portion of heat obtained from the volumetric strain rate.

Damage effects were not taken into account in the present modeling of the monolithic SMA material or in the SMA composite. Considerations of such effects is a subject for future research.

The present micromechanical model can be employed to investigate the behavior of SMA composite structures with two-way TMC. In particular, the induced temperatures that are caused by bending, buckling, postbuckling and vibrations of a SMA composite structure can be determined by employing a micro-macro-structural analysis according to which the global constitutive equations that are obtained from the proposed model are applied at every point of the structure.

Acknowledgments

References

[Aboudi 2004] J. Aboudi, "The generalized method of cells and high-fidelity generalized method of cells micromechanical models — a review", *Mech. Adv. Mater. Struct.* **11** (2004), 329–366.

[Aboudi 2005] J. Aboudi, "Micromechanically established constitutive equations for multiphase materials with viscoelastic-viscoplastic phases", *Mech. Time-Depend. Mat.* **9**:2-3 (2005), 121–145.

[Aboudi et al. 2002] J. Aboudi, M.-J. Pindera, and S. M. Arnold, "High-fidelity generalization method of cells for inelastic periodic multiphase materials", NASA, Houston, 2002, Available at http://gltrs.grc.nasa.gov/citations/all/tm-2002-211469.html.

[Aboudi et al. 2003] J. Aboudi, M.-J. Pindera, and S. M. Arnold, "Higher-order theory for periodic multiphase materials with inelastic phases", *Int. J. Plasticity* **19**:6 (2003), 805–847.

[Allen 1991] D. H. Allen, "Thermo-mechanical coupling in inelastic solids", *Appl. Mech. Rev.* **44** (1991), 361–373.

[Auricchio and Petrini 2002] F. Auricchio and L. Petrini, "Improvements and algorithmical considerations on a recent three-dimensional model describing stress-induced solid phase transformations", *Int. J. Numer. Meth. Eng.* **55**:11 (2002), 1255–1284.

[Auricchio and Petrini 2004a] F. Auricchio and L. Petrini, "A three-dimensional model describing stress-temperature induced solid phase transformations: solution algorithm and boundary value problems", *Int. J. Numer. Meth. Eng.* **61**:6 (2004), 807–836.

[Auricchio and Petrini 2004b] F. Auricchio and L. Petrini, "A three-dimensional model describing stress-temperature induced solid phase transformations: thermomechanical coupling and hybrid composite applications", *Int. J. Numer. Meth. Eng.* **61**:5 (2004), 716–737.

[Bednarcyk and Arnold 2002] B. A. Bednarcyk and S. M. Arnold, "MAC/GMC 4.0 user's manual", NASA, Houston, 2002, Available at http://gltrs.grc.nasa.gov/cgi-bin/GLTRS/browse.pl?2002/TM-2002-212077-VOL3.% html.

[Bednarcyk et al. 2004] B. A. Bednarcyk, S. M. Arnold, J. Aboudi, and M.-J. Pindera, "Local field effects in titanium matrix composites subjected to fiber-matrix debonding", *Int. J. Plasticity* **20**:8-9 (2004), 1707–1737.

[Boyd and Lagoudas 1994] J. G. Boyd and D. C. Lagoudas, "Thermomechanical response of shape memory composites", *J. Intel. Mat. Syst. Struct.* **5**:3 (1994), 333–346.

[Carvelli and Taliercio 1999] V. Carvelli and A. Taliercio, "A micromechanical model for the analysis of unidirectional elasto-plastic composites subjected to 3D stresses", *Mech. Res. Commun.* **26**:5 (1999), 547–553.

[Gilat and Aboudi 2004] R. Gilat and J. Aboudi, "Dynamic response of active composite plates: shape memory alloy fibers in polymeric/metallic matrices", *Int. J. Solids Struct.* **41**:20 (2004), 5717–5731.

[Gilat and Aboudi 2006] R. Gilat and J. Aboudi, "Thermal buckling of activated shape memory reinforced laminated plates", *Smart Mater. Struct.* **15** (2006), 829–838.

[Hunter 1983] S. C. Hunter, *Mechanics of continuous media*, 2nd ed. ed., Ellis Horwood Ltd, Chichester, 1983.

[Kawai 2000] M. Kawai, "Effects of matrix inelasticity on the overall hysteretic behavior of TiNi-SMA fiber composites", *Int. J. Plasticity* **16**:3-4 (2000), 263–282.

[Kawai et al. 1999] M. Kawai, H. Ogawa, V. Baburaj, and T. Koga, "Micromechanical analysis for hysteretic behavior of unidirectional TiNi SMA fiber composites", *J. Intel. Mat. Syst. Struct.* **10**:1 (1999), 14–28.

[Marfia 2005] S. Marfia, "Micro-macro analysis of shape memory alloy composites", *Int. J. Solids Struct.* **42**:13 (2005), 3677–3699.

[Mitchell and Griffiths 1980] A. R. Mitchell and D. F. Griffiths, *The finite difference method in partial differential equations*, John Wiley, Chichester, 1980.

[Simo and Hughes 1998] J. C. Simo and T. J. R. Hughes, *Computational inelasticity*, Springer, New York, 1998.

[Song et al. 1999] G. Q. Song, Q. P. Sun, and M. Cherkaoui, "Role of microstructures in the thermomechanical behavior of SMA composites", *J. Eng. Mater. Tech.* **121** (1999), 86–92.

[Souza et al. 1998] A. C. Souza, E. N. Mamiya, and N. Zouain, "Three-dimensional model for solids undergoing stress-induced phase transformations", *Eur. J. Mech. A/Solids* **17**:5 (1998), 789–806.

[Williams and Aboudi 1999] T. O. Williams and J. Aboudi, "A fully coupled thermomechanical micromechanics model", *J. Therm. Stresses* **22**:9 (1999), 841–873.

JACOB ABOUDI: aboudi@eng.tau.ac.il
Department of Solid Mechanics, Materials and Systems, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Israel
http://www.eng.tau.ac.il/~aboudi/

YUVAL FREED: yuval@eng.tau.ac.il
Department of Solid Mechanics, Materials and Systems, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

SUBMISSION GUIDELINES

ORIGINALITY

Authors may submit manuscripts in PDF format on-line. Submission of a manuscript acknowledges that the manuscript is *original and has neither previously, nor simultaneously, in whole or in part, been submitted elsewhere.* Information regarding the preparation of manuscripts is provided below. Correspondence by email is requested for convenience and speed. For further information, consult the web site at http://www.jomms.org or write to

Marie-Louise Steele
Division of Mechanics and Computation
Durand Building, Room 262
Stanford University
Stanford CA 94305

LANGUAGE

Manuscripts must be in English. A brief abstract of about 150 words or less must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate), and email address if available. A home-page URL is optional.

FORMAT

Authors are encouraged to use LaTeX and the standard article class, but submissions in other varieties of TeX, and, exceptionally in other formats, are acceptable. Electronic submissions are strongly encouraged in PDF format only; after the refereeing process we will ask you to submit all source material.

REFERENCES

Bibliographical references should be listed alphabetically at the end of the paper and include the title of the article. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format (see a current issue for examples), however, in the manuscript, the citation should be by first author's last name and year of publication, e.g. "as shown by Kramer, et al. (1994)". Links will be provided to all literature with known web locations and authors are encouraged to provide their own links on top of the ones provided by the editorial process.

FIGURES

Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple, or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to

production@mathscipub.org

Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by comments such as "Place Figure 1 here". The same considerations apply to tables.

WHITE SPACE

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

PROOFS

Page proofs will be made available to authors (or to the designated corresponding author) at a web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Journal of Mechanics of Materials and Structures