Pergamon

Int. J. Solids Structures Vol. 33, No. 7, pp. 991–1003, 1996

0020–7683/96 $15.00 + .00

0020–7683(95)00082–8

GENERALIZED PLASTICITY AND SHAPE-MEMORY ALLOYS

J. LUBLINER and F. AURICCHIO†
Department of Civil Engineering, University of California at Berkeley, Berkeley, CA 94720, U.S.A

(*Received* 16 *October* 1994; *in revised form* 20 *March* 1995)

Abstract—The theory of the generalized plasticity model is reviewed. A special form for multiaxial behavior, based on the Drucker–Prager flow potential, is proposed. The model is applied to a simplified representation of the behavior of shape-memory alloys, with numerical examples.

1. INTRODUCTION

Generalized plasticity is an internal-variable model of rate-independent inelasticity that includes conventional or classical plasticity as a special case (Lubliner, 1984, 1987). It was developed in order to account for the behavior of elastic-plastic solids in which, following initial plastic loading and elastic unloading, the reloading is not necessarily elastic up to the state at which unloading began; such solids include graphite, some stainless steels, some rocks, and others. Simple versions of the model that are easy to implement numerically have recently been proposed and tested (Lubliner, 1991; Lubliner *et al.*, 1993; Auricchio and Taylor, 1994).

In spite of the name chosen for it, however, the model is not a mere generalization of plasticity theory. Rather, it is based on some fundamental axioms and on results from elementary set theory and topology. Thus, for example, there is no requirement that the elastic domain (the set of states at which only elastic processes are possible) be connected.

This generality makes the model potentially capable of describing solid behavior that may be vastly different from elastic-plastic behavior. Solids known as *shape-memory alloys* exhibit such behavior, in the form of *pseudoelasticity* (the recovery of large deformations in a loading-unloading cycle, occurring at sufficiently high temperatures) and the *shape-memory effect* (recovery of large deformations by a combination of mechanical and thermal processes). Shape-memory alloys have been studied experimentally for the last three decades, and a plethora of constitutive models has been proposed over the past 10 years. Virtually all these models, however, are *ad hoc* descriptions of observed behavior, and are limited to uniaxial loading.

This paper represents a first attempt to apply generalized plasticity to the behavior of shape-memory alloys. Section 2 gives a review of the underlying theory, with the formal mathematics relegated to an appendix. In Section 3, a simple three-dimensional form is proposed for solids that may have an inelastic volume deformation, as shape-memory alloys appear to, on the basis of the Drucker–Prager flow potential. Section 4 contains the application of the model to a simplified representation of the behavior of shape-memory alloys, which nevertheless preserves its salient features. The application is confirmed by numerical examples. A concluding section indicates directions for future research.

---

† On leave from Dipartimento di Ingegneria Civile, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 00133, Roma, Italy

2. GENERALIZED PLASTICITY: THEORY

2.1. *States and processes*

A local thermomechanical state is represented by the couple $(\mathbf{G}, \mathbf{q})$, where $\mathbf{G}$ (belonging to a space $\mathscr{G}$) stands for the "controllable" state variables and $\mathbf{q}$ for the internal variables. Under strain-temperature control, $\mathbf{G}$ may be $(\mathbf{E}, T)$, where $\mathbf{E}$ is the Green–Saint-Venant strain tensor (which may be replaced by the infinitesimal strain $\boldsymbol{\varepsilon}$ under infinitesimal displacement gradients) and T is the temperature. Under stress-temperature control, $\mathbf{G}$ may be $(\mathbf{S}, T)$, where $\mathbf{S}$ is the second Piola–Kirchhoff stress tensor, under large displacement gradients and $(\boldsymbol{\sigma}, T)$, where $\boldsymbol{\sigma}$ is the conventional stress, under infinitesimal displacement gradients.

The set of realizable states is denoted $\mathscr{S}$, and we define $\mathscr{S}_{\mathbf{q}} \overset{\text{def}}{=} \{\mathbf{G} \mid (\mathbf{G}, \mathbf{q}) \in \mathscr{S}\}$, that is, a projection of $\mathscr{S}$ onto a plane $\mathbf{q} =$ constant. The constitutive equations consist of the thermomechanical equations of state (that is, the stress-strain-temperature relations, in general dependent on the internal variables as well) and the rate equations for the internal variables. The latter are assumed to take the form

$$\dot{\mathbf{q}} = \mathbf{f}(\mathbf{G}, \mathbf{q}, \dot{\mathbf{G}}). \tag{1}$$

Rate-independence means that eqn (1) is invariant under a replacement of t by $\phi(t)$, where $\phi(\cdot)$ is any monotonically increasing, continuously differentiable function. It can easily be shown that a necessary and sufficient condition for rate-independence is that $\mathbf{f}(\mathbf{G}, \mathbf{q}, \cdot)$ be homogenous of the first degree, that is, for any positive number c,

$$\mathbf{f}(\mathbf{G}, \mathbf{q}, c\dot{\mathbf{G}}) = c\mathbf{f}(\mathbf{G}, \mathbf{q}, \dot{\mathbf{G}}). \tag{2}$$

If a material neighborhood of a rate-independent inelastic body is treated as a small thermodynamic system, with a local process defined as a mapping $t \mapsto (\mathbf{G}(t), \mathbf{q}(t))$ of a time interval—say $[t_0, t_1]$, in which case the process is said to go *from* $(\mathbf{G}(t_0), \mathbf{q}(t_0))$ *to* $(\mathbf{G}(t_1), \mathbf{q}(t_1))$—then every state is an equilibrium state, and consequently *every process is quasi-static but not, in general, reversible*.

A function $t \mapsto \phi(t)$ such as was mentioned above in defining rate-independence produces a process $t \mapsto (\mathbf{G}(\phi(t)), \mathbf{q}(\phi(t)))$. Since any such function is invertible, the relation between this process and the original process $t \to (\mathbf{G}(t), \mathbf{q}(t))$ is an equivalence on the set of processes. The corresponding equivalence classes are called *paths*.

A process is called *elastic* if $\mathbf{q}(\cdot)$ is a constant function. If the Kelvin inequality (Lubliner, 1990; p. 62) is assumed to hold as representing the second law of thermodynamics, then in an elastic process (with heat conduction neglected) the internal entropy production vanishes: the process is not necessarily reversible, but may be called *quasi-reversible* (Fosdick and Serrin, 1975). It is obvious that a process is elastic if and only if all processes having the same path are elastic, and hence one can speak of elastic and inelastic paths.

As a result of rate-independence—embodied in the constraint (2) on the rate equation (1)—any process with $\mathbf{G}(\cdot)$ equal to a constant function is necessarily elastic. In a rate-dependent (e.g. viscoplastic) body, such a process would be relaxation process and hence inelastic, unless $\mathbf{G}$ is in the elastic region. In the present context, such a process (whose path consists of one point) may be called *trivial*.

2.2. *Elastic range, state and domain*

Rate-independent plasticity is closely tied to the concept of *elastic range*, first formalized by Pipkin and Rivlin (1965) and later expanded by Owen (1968, 1970). These studies used the framework of the theory of materials with memory without reference to internal variables. Here the concept is defined with the use of internal variables.

The *elastic range* of a state $(\mathbf{G}, \mathbf{q}) \in \mathscr{S}$ is defined as

$$\mathscr{E}(\mathbf{G},\mathbf{q}) = \{\mathbf{G}^* | \quad \text{there exists an elastic process from} \quad (\mathbf{G},\mathbf{q}) \text{ to } (\mathbf{G}^*,\mathbf{q})\}.$$

Since a trivial process is elastic, it is obvious that $\mathbf{G} \in \mathscr{E}(\mathbf{G},\mathbf{q})$ and, therefore, every state has a non-empty elastic range. The set $\mathscr{E}(\mathbf{G},\mathbf{q}) \times \{\mathbf{q}\}$, which may be naturally identified with $\mathscr{E}(\mathbf{G},\mathbf{q})$, is obviously the union of the ranges of all elastic processes (paths) from $(\mathbf{G},\mathbf{q})$ and is, therefore, path-connected and hence connected. The elastic range of every state will furthermore be assumed to be closed in $\mathscr{S}_{\mathbf{q}}$. A similar assumption was made by Pipkin and Rivlin (1965). In the work of Owen (1968, 1970), however, the elastic range, differently defined, is an open set.

A state $(\mathbf{G},\mathbf{q}) \in \mathscr{S}$ will be called *elastic* if $\mathbf{G}$ is an interior point of $\mathscr{E}(\mathbf{G},\mathbf{q})$ and *plastic* or *inelastic* if $\mathbf{G}$ is a boundary point of $\mathscr{E}(\mathbf{G},\mathbf{q})$. The set of all elastic states in $\mathscr{S}$ will be called the *elastic* domain and denoted $\mathscr{S}^{\mathrm{E}}$; the projection of $\mathscr{S}^{\mathrm{E}}$ into $\mathscr{S}_{\mathbf{q}}$, that is, the set $\{\mathbf{G} | (\mathbf{G},\mathbf{q}) \in \mathscr{S}^{\mathrm{E}}\}$ for a given $\mathbf{q}$, will be denoted $\mathscr{S}^{\mathrm{E}}_{\mathbf{q}}$ and also called the elastic domain (at $\mathbf{q}$). Unlike $\mathscr{E}(\mathbf{G},\mathbf{q})$, $\mathscr{S}^{E}_{\mathbf{q}}$ need not, in general, be connected. If it is, then it is easy to show that $\mathscr{S}^{\mathrm{E}}_{\mathbf{q}} \subset \mathscr{E}(\mathbf{G},\mathbf{q})$ for every $\mathbf{G} \in \mathscr{S}_{\mathbf{q}}$.† The set of all inelastic states may analogously be denoted $\mathscr{S}^{\mathrm{I}} = \mathscr{S} - \mathscr{S}^{\mathrm{E}}$, and its projections into $\mathscr{S}_{\mathbf{q}}$ is $\mathscr{S}^{\mathrm{I}}_{\mathbf{q}} = \mathscr{S}_{\mathbf{q}} - \mathscr{S}^{\mathrm{E}}_{\mathbf{q}}$. It is, furthermore, convenient to define

$$\mathscr{G}^{\mathrm{I}} = \bigcup_{\mathbf{q}} \mathscr{S}^{\mathrm{I}}_{\mathbf{q}},$$

so that a state $(\mathbf{G},\mathbf{q})$ is elastic if (but not only if) $\mathbf{G} \notin \mathscr{G}^{\mathrm{I}}$.

2.3. *Rate equations*

Since it is the rate equations that determine which processes are elastic and inelastic, it is the nature of these equations that ultimately decides the structure of the elastic range and, therefore, of the elastic domain. In general it is quite difficult to deduce this structure from the rate equations, and, rather than attempt to do this, we assume the sets in question to be sufficiently regular in some sense, and then deduce some necessary properties of the function $\mathbf{f}$ in eqn (1).

For example, if $\mathbf{G} \in \mathscr{S}^{\mathrm{E}}_{\mathbf{q}}$ then $\mathbf{f}(\mathbf{G},\mathbf{q},\dot{\mathbf{G}}) = 0$, since every $(\mathbf{G}^*,\mathbf{q})$ with $\mathbf{G}^*$ in a small enough neighborhood of $\mathbf{G}$ is attainable elastically and, therefore, $\dot{\mathbf{q}} = 0$ in any possible process through $(\mathbf{G},\mathbf{q})$.

If the elastic range of a state $(\mathbf{G},\mathbf{q})$ has a non-empty interior, then its boundary $\partial\mathscr{E}(\mathbf{G},\mathbf{q})$ may be assumed to be a piecewise smooth surface in $\mathscr{S}_{\mathbf{q}}$. If, in particular, $(\mathbf{G},\mathbf{q})$ is a plastic state and $\partial\mathscr{E}(\mathbf{G},\mathbf{q})$ is locally smooth at $\mathbf{G}$ (in which case the state may be called a *regular* plastic state), with a normal $\mathbf{N}$ pointing away from $\mathring{\mathscr{E}}(\mathbf{G},\mathbf{q})$, then $\mathbf{f}(\mathbf{G},\mathbf{q},\cdot)$, if continuous, must have the property that $\mathbf{f}(\mathbf{G},\mathbf{q},\dot{\mathbf{G}}) = 0$ if and only if $\mathbf{N}\cdot\dot{\mathbf{G}} \leqslant 0$.‡ (*Proof*: for a sufficiently small positive number h, $\mathbf{G}+h\dot{\mathbf{G}}$ is in the interior or exterior of $\partial\mathscr{E}(\mathbf{G},\mathbf{q})$, respectively, as $\mathbf{N}\cdot\dot{\mathbf{G}} < 0$ or $\mathbf{N}\cdot\dot{\mathbf{G}} > 0$. In the former case $\mathbf{G}+h\dot{\mathbf{G}}$ is attainable elastically, hence $\dot{\mathbf{q}} = 0$: in the latter case $\mathbf{G}+h\dot{\mathbf{G}}$ is not attainable elastically, and in particular not by a straight-line path, hence $\dot{\mathbf{q}} \neq 0$. The limiting case $\mathbf{N}\cdot\dot{\mathbf{G}} = 0$ follows by continuity.) The simplest form of $\mathbf{f}$ having this property, as well as obeying the homogeneity condition (2), is

$$\mathbf{f}(\mathbf{G},\mathbf{q},\dot{\mathbf{G}}) = \mathbf{g}(\mathbf{G},\mathbf{q})\langle\mathbf{N}\cdot\dot{\mathbf{G}}\rangle, \tag{3}$$

where $\langle\cdot\rangle$ is the Macaulay bracket, that is $\langle x\rangle = 1/2(x+|x|)$. More complex forms were considered by Darve and collaborators (Darve *et al.*, 1978; Darve and Labanieh, 1982).

---

† Note that a process whose range is entirely in the elastic domain is reversible and not merely quasi-reversible.

‡ The centred dot · designates the scalar product in an arbitrary n-dimensional vector space. In spaces of symmetric second-rank tensors, however, the conventional colon : will be used.

2.4. *Thermomechanical state equations and Coleman method*

In the contemporary literature of thermomechanics, it is common to use a method developed by Coleman and various collaborators, and in particular by Coleman and Gurtin (1967) in the context of internal variables, for the derivation of the thermomechanical equations of state on the basis of the Clausius–Duhem inequality. It is instructive to test the applicability of the method to the rate-independent inelastic continuum discussed here, with the rate equations taking in particular the form (3). In the absence of heat conduction, the Clausius–Duhem inequality may be written in the form of the Clausius–Planck inequality,

$$\boldsymbol{\Sigma}\cdot\dot{\mathbf{G}}-\dot{\Psi}=(\boldsymbol{\Sigma}-\partial\Psi/\partial\mathbf{G})\cdot\dot{\mathbf{G}}+D(\mathbf{G},\mathbf{q},\dot{\mathbf{q}})\geqslant 0, \tag{4}$$

where $\boldsymbol{\Sigma}=(\mathbf{S},-\eta)$ (η being the entropy density), ψ is the Helmholtz free-energy density, and

$$D(\mathbf{G},\mathbf{q},\dot{\mathbf{q}})\overset{\text{def}}{=}-\partial\Psi(\mathbf{G},\mathbf{q})/\partial\mathbf{q}\cdot\dot{\mathbf{q}}$$

is the internal dissipation density (all densities are per unit reference volume). The aforementioned Kelvin inequality is just

$$D(\mathbf{G},\mathbf{q},\dot{\mathbf{q}})\geqslant 0.$$

With the rate equation (3) inserted, inequality (4) becomes

$$(\boldsymbol{\Sigma}-\partial\Psi/\partial\mathbf{G})\cdot\dot{\mathbf{G}}+D(\mathbf{G},\mathbf{q},\mathbf{g}(\mathbf{G},\mathbf{q}))\langle\mathbf{N}\cdot\dot{\mathbf{G}}\rangle\geqslant 0;$$

but this is satisfied at all $(\mathbf{G},\mathbf{q},\dot{\mathbf{G}})$ if

$$\boldsymbol{\Sigma}=\partial\Psi/\partial\mathbf{G}-\lambda D(\mathbf{G},\mathbf{q},\mathbf{g}(\mathbf{G},\mathbf{q}))\mathbf{N}$$

for any $\lambda\in[0,1]$, since for any such λ and any x, $\langle x\rangle-\lambda x\leqslant 0$. Thus, as was noted by Lubliner (1972), the Coleman method does not lead to a unique dependence of $\boldsymbol{\Sigma}$ on the state without a further assumption, namely, that at a plastic state an elastic process, in which $\mathbf{N}\cdot\dot{\mathbf{G}}\leqslant 0$ by hypothesis, is quasi-reversible, that is, that in such a process the Clausius–Planck inequality holds as an equality. This assumption leads to $\lambda=0$ as the only possible value and hence yields the classical thermomechanical state equation $\boldsymbol{\Sigma}=\partial\Psi/\partial\mathbf{G}$, that is, $\mathbf{S}=\partial\Psi/\partial\mathbf{E}$ and $\eta=-\partial\Psi/\partial T$.

2.5. *Properties of the elastic range and domain*

If, as mentioned above, the boundary $\partial\mathscr{E}(\mathbf{G},\mathbf{q})$ of the elastic range of a state $(\mathbf{G},\mathbf{q})$ is a piecewise smooth surface, then it may, following Eisenberg and Phillips (1971), be called a *loading surface*. Similarly, the boundary $\partial\mathscr{S}_{\mathbf{q}}^{E}$ of the elastic domain at $\mathbf{q}$ may be assumed to form a piecewise smooth surface, which may be called the *yield surface* at $\mathbf{q}$.† The special case where the loading and yield surface coincide, that is, in which $\partial\mathscr{E}(\mathbf{G},\mathbf{q})=\partial\mathscr{S}_{\mathbf{q}}^{E}$ for every $\mathbf{G}\in\mathscr{S}_{\mathbf{q}}$ at a given $\mathbf{q}$, corresponds to *classical* or *conventional* plasticity, a fact that has given the more general case studied here the name *generalized plasticity*.

Some additional properties can be proved with the help of some propositions, stated and proved in Appendix 1 and based on elementary set theory and topology.

A corollary of Proposition 1 is the following:

If, at each $\mathbf{q}$, $\partial\mathscr{E}(\cdot,\mathbf{q})$ is continuous (in the sense defined in the Appendix) in $\mathscr{S}_{\mathbf{q}}$, then $\mathscr{S}_{\mathbf{q}}^{E}$ is open in $\mathscr{S}_{\mathbf{q}}$.

The following is a corollary of Proposition 2:

---

† It should be noted that in later work by Phillips and his collaborators the term *loading surface* took on a different meaning.

If $\mathscr{E}(\mathbf{G},\mathbf{q})$ equals the closure of its interior, then for every $\mathbf{G} \in \mathscr{S}_{\mathbf{q}}^{E}$ we have $\mathscr{E}(\mathbf{G},\mathbf{q}) = \overline{\mathscr{S}_{\mathbf{q}}^{E}}$ and $\partial\mathscr{E}(\mathbf{G},\mathbf{q}) = \partial\mathscr{S}_{\mathbf{q}}^{E}$. This result may also be phrased as follows (Eisenberg and Phillips, 1971): *the yield surface is the initial loading surface.*

2.6. *Classical plasticity as a special case*

Consider the case where $\mathscr{E}(\mathbf{G},\mathbf{q})$ is independent of $\mathbf{G}$. It then follows from Proposition 3 of Appendix 1 that $\overset{\circ}{\mathscr{E}}(\mathbf{G},\mathbf{q})$ equals $\mathscr{S}_{\mathbf{q}}^{E}$ at every $\mathbf{q}$. Moreover, since $\mathscr{E}(\mathbf{G},\mathbf{q})$ is connected, it contains no isolated points and is, therefore, equal to the closure of its interior; thus $\partial\mathscr{E}(\mathbf{G},\mathbf{q}) = \partial\mathscr{S}_{\mathbf{q}}^{E}$, and this is precisely the definition of classical plasticity given above. It further follows that $\partial\mathscr{S}_{\mathbf{q}}^{E}$ is the set of all plastic states:

$$\partial\mathscr{S}_{\mathbf{q}}^{E} = \{\mathbf{G} | \mathbf{G} \in \partial\mathscr{S}_{\mathbf{q}}^{E}\} = \{\mathbf{G} | \mathbf{G} \in \partial\mathscr{E}(\mathbf{G},\mathbf{q})\},$$

and, therefore,

$$\mathscr{S}_{\mathbf{q}}^{E} = \mathscr{S}_{\mathbf{q}}$$

Another definition of classical plasticity, given by Pipkin and Rivlin (1965) is that *the elastic range is unaffected by elastic deformations*, that is, $\mathscr{E}(\mathbf{G}^*,\mathbf{q}) = \mathscr{E}(\mathbf{G},\mathbf{q})$ if $\mathbf{G}^* \in \mathscr{E}(\mathbf{G},\mathbf{q})$. It is easy to show that this definition is implied by the one adopted here: If $\mathscr{E}(\mathbf{G},\mathbf{q}) = \mathscr{E}_{\mathbf{q}}$ for all $\mathbf{G} \in \mathscr{S}_{\mathbf{q}}$, then in particular $\mathscr{E}(\mathbf{G}^*,\mathbf{q}) = \mathscr{E}_{\mathbf{q}}$ for all $\mathbf{G}^* \in \mathscr{E}(\mathbf{G},\mathbf{q}) = \mathscr{E}_{\mathbf{q}}$.

3. GENERALIZED PLASTICITY OF DRUCKER–PRAGER TYPE

3.1. *Flow rule and inelastic potential*

The following discussion will be limited to stress-temperature control with infinitesimal displacement gradients. Moreover, the control state will be denoted explicitly as $(\boldsymbol{\sigma}, T)$ rather than $\mathbf{G}$. If the thermomechanical strain-stress-temperature relation is given by

$$\varepsilon = \hat{\varepsilon}(\boldsymbol{\sigma}, T, \mathbf{q}),$$

then the *inelastic strain rate* is defined as

$$\dot{\varepsilon}^{i} = \partial\hat{\varepsilon}/\partial\mathbf{q} \cdot \dot{\mathbf{q}}.$$

Note that this definition does not assume the existence of an inelastic strain. Under infinitesimal displacement gradients, however, it can usually be assumed (Lubliner, 1972) that

$$\hat{\varepsilon}(\boldsymbol{\sigma}, T, \mathbf{q}) = \varepsilon^{e}(\boldsymbol{\sigma}, T) + \varepsilon^{i}(\mathbf{q}),$$

in which case $\dot{\boldsymbol{\varepsilon}}^{i}$ is just the time derivative of the inelastic strain ε^{i}, while the thermoelastic strain $\boldsymbol{\varepsilon}^{e}$ is given by the usual linear stress-strain-temperature relations.

It can furthermore be assumed, with no loss of generality, that the internal-variable array $\mathbf{q}$ consists of $\boldsymbol{\varepsilon}^{i}$ and an additional array $\xi = (\xi_1, \xi_2, \ldots)$, in which case the dependence $\boldsymbol{\varepsilon}^{i}(\mathbf{q})$ is given by an identity relation.

An equation in which $\dot{\boldsymbol{\varepsilon}}^{i}$ is given to within a multiplicative scalar is known as a *flow rule*.

If there exists a scalar function $F(\boldsymbol{\sigma}, T, \mathbf{q})$ such that

$$\dot{\boldsymbol{\varepsilon}}^{i} = \dot{\lambda}\,\partial F/\partial\boldsymbol{\sigma},$$

where $\dot{\lambda}$ is a scalar, then F is called an *inelastic potential.*

3.2. *Drucker–Prager potential*

Let $\mathbf{s}$ denote the stress-deviator tensor, and let $J_2 = \frac{1}{2}\mathbf{s}:\mathbf{s}$, while $I_1 = tr\,\boldsymbol{\sigma}$, so that $\mathbf{s} = \boldsymbol{\sigma} - \frac{1}{3}I_1\mathbf{1}$, where $\mathbf{1}$ is the identity tensor. The inelastic potential is of *Drucker-Prager* type if it takes the form

$$F(\boldsymbol{\sigma}, T, \mathbf{q}) = \rho(\boldsymbol{\sigma}) - \phi(T), \tag{5}$$

where ϕ is some known function and

$$\rho(\boldsymbol{\sigma}) \overset{\text{def}}{=} \sqrt{3J_2} + \alpha I_1 .$$

The special case $\alpha = 0$ corresponds to equal response in tension and compression; the inelastic potential is then of *von Mises* type.

The flow rule for the Drucker-Prager potential takes the form

$$\dot{\boldsymbol{\varepsilon}}^{i} = \dot{\lambda}\left(\frac{\sqrt{3}}{2\sqrt{J_2}}\,\mathbf{s} + \alpha\mathbf{1}\right). \tag{6}$$

The rate of inelastic work (per unit volume) is

$$\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{i} = \dot{\lambda}\rho(\boldsymbol{\sigma}). \tag{7}$$

Note that in the von Mises case, $\dot{\lambda} = \sqrt{(2/3)\dot{\boldsymbol{\varepsilon}}^{i} : \dot{\boldsymbol{\varepsilon}}^{i}}$.

In radial loading, the stress can be expressed as $\boldsymbol{\sigma} = \mathbf{v}\rho(\boldsymbol{\sigma})$, where $\mathbf{v}$ is a constant tensor. Then

$$\dot{\lambda} = \mathbf{v} : \dot{\boldsymbol{\varepsilon}}^{i}$$

and, therefore,

$$\lambda = \mathbf{v} : \boldsymbol{\varepsilon}^{i} .$$

For example, in simple shear, with the shear stress given by τ and the conventional shear strain by γ, $\rho(\boldsymbol{\sigma}) = \sqrt{3}\,\tau$ and $\lambda = \gamma^{i}/\sqrt{3}$. In uniaxial tension and compression, if σ and ε denote, respectively, the magnitudes of stress and strain, then

$$\rho(\boldsymbol{\sigma}) = (1 \pm \alpha)\sigma,$$

$$\lambda = \frac{1}{1 \pm \alpha}\varepsilon^{i},$$

where the plus and minus signs hold in tension and compression, respectively.

3.3. *Simplifying assumptions*

It will first be assumed that in the general case the quantity λ, obtained by integrating $\dot{\lambda}$ as given by eqn (7), is uniquely determined by $\boldsymbol{\xi}$. It is a generalization of the "effective inelastic strain" commonly used in plasticity theory as a "hardening variable".

Suppose, next, that the loading surfaces in ($\boldsymbol{\sigma}$, T) space have the form $F(\boldsymbol{\sigma}, T, \mathbf{q}) =$ constant, with F given by eqn (5), that is,

$$\partial\mathscr{E}(\boldsymbol{\sigma}, T, \mathbf{q}) = \{(\boldsymbol{\sigma}^*, T^*) \mid \rho(\boldsymbol{\sigma}^*) - \phi(T^*) = \rho(\boldsymbol{\sigma}) - \phi(T)\}.$$

Since the normal $\mathbf{N}$ has been defined only as to direction, the quantity $\mathbf{N}\cdot\dot{\mathbf{G}}$ in eqn (3) may be given by

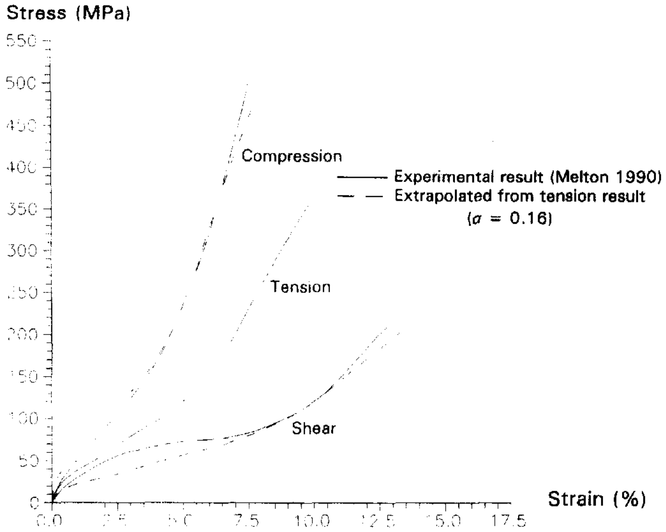

Fig. 1. Application of generalized plasticity with a Drucker–Prager potential to a martensitic Ni–Ti–10% Cu alloy.

$$\mathbf{N}\cdot\dot{\mathbf{G}} = \pm[(\partial\rho/\partial\boldsymbol{\sigma}):\dot{\boldsymbol{\sigma}} - \phi'(T)\dot{T}] = \pm\dot{F},$$

where the sign depends on whether the elastic range is on one side or the other of the loading surface. For the sake of definiteness, let the positive sign be taken. Then

$$\dot{\lambda} = \left[\frac{1}{\rho(\boldsymbol{\sigma})}\boldsymbol{\sigma} : \partial\boldsymbol{\varepsilon}^{i}/\partial\mathbf{q}\cdot\mathbf{g}(\boldsymbol{\sigma}, T, \mathbf{q})\right]\langle\dot{F}\rangle. \tag{8}$$

Finally, an assumption will be made concerning the quantity between brackets on the right-hand side of eqn (8): namely, that this function of $(\boldsymbol{\sigma}, T, \mathbf{q})$ depends on $(\boldsymbol{\sigma}, T)$ only through F and on $\mathbf{q}$ only through λ, that is,

$$\dot{\lambda} = h(F, \lambda)\langle\dot{F}\rangle. \tag{9}$$

Equation (9) is an ordinary differential equation in the variables λ and F. For example, a ρ–λ curve for isothermal initial loading can be obtained by setting $T = T_0$ (constant) and the initial condition $\lambda = 0$ when $\rho = 0$.

The hypothesis of a unique ρ–λ curve implies that if α is known, then a stress–strain curve in, say, tension may be used to generate the corresponding curves for shear (as given, for example, by the torsion of a thin-walled circular tube) and compression. Figure 1 shows the results of such an extrapolation, with an assumed value of $\alpha = 0.16$, compared with the experimental results obtained by Melton (1990) for a martensitic Ni–Ti–10% Cu alloy. As can be seen, the comparison is quite good for compression and fair for shear.

4. A SIMPLE MODEL FOR SHAPE-MEMORY ALLOYS

4.1. *Phase transformations*

Shape-memory alloys belong to a class of solids that can undergo reversible, diffusionless transformations between a highly ordered phase called *austenitic* (or the *parent phase*) and a less ordered phase called *martensitic*. Typically, the former is stable at higher temperatures and the latter at lower temperatures.

In a stress-free state, the austenite-to-martensite ($A \rightarrow M$) transformation begins at a temperature denoted M_s (for *martensite start*) and ends at the lower temperature M_f (for *martensite finish*); at intermediate temperatures the two phases coexist. The $A \rightarrow M$ transformation is accompanied by a deformation that is considerably greater than the thermal contraction occurring in the same temperature range; indeed, the coefficient of thermal expansion is nearly the same for austenite and martensite.

The reverse transformation, from martensite to austenite ($M \to A$), begins at the temperature A_s (for *austenite start*) and ends at the higher temperature A_f (for *austenite finish*); again, the two phases coexist at intermediate temperatures. The large deformation that occurred with the $A \to M$ transformation is removed in the $M \to A$ transformation.

At a constant temperature T, the same transformations may, depending on its value, be effected by subjecting the solid to stress, with an increase in stress being equivalent to a decrease in temperature. Thus, if the solid is initially austenitic, it may be transformed into martensite by stressing, and the resulting deformation is considerably greater than the accompanying elastic strain; the resulting stress–strain diagram resembles those found in elastic-plastic solids. Whether an $M \to A$ transformation, with the attendant removal of the deformation, occurs upon unloading depends on the temperature: it does not occur if $T \leqslant A_s$; it occurs partially if $A_s < T < A_f$, and completely if $T \geqslant A_f$. This last phenomenon, in which a large deformation is removed upon unloading, is known as *pseudoelasticity*. Otherwise, the removal of the deformation can be effected by heating, and subsequent cooling, following unloading. It is this recovery of the initial shape by a combination of mechanical and thermal processes that constitutes the *shape-memory effect*.

Numerous constitutive models have been proposed to describe the behavior of these alloys (Tanaka *et al.*, 1982, 1985, 1986, 1992; Cory and McNichols, 1985, 1987; Patoor *et al.*, 1988; Falk and Konopka, 1990; Liang and Rogers, 1990, 1992; Müller and Xu, 1991; Tobushi *et al.*, 1991; Brandon and Rogers, 1992; Abeyaratne and Knowles, 1993; Brinson, 1993; Ivshin and Pence, 1993, 1994; Sun and Hwang, 1993a, b; Wilmanski, 1993; Raniecki and Lexcellent, 1994). Most of these models, however, are *ad hoc* descriptions of uniaxial behavior, though some of them are based on formulations that are formally three-dimensional. The model presented here is fully three-dimensional and is based on generalized plasticity of Drucker–Prager type.

4.2. *Application of generalized plasticity model*

Experimental results (Funakubo, 1987; Chrysochoos, 1993) indicate that under a combination of a simple stress σ (whatever its nature) and temperature T, the $A \to M$ transformation takes place in that band in the σ-T plane that is bounded by the straight lines $\sigma = C_M(T - M_s)$ and $\sigma = C_M(T - M_f)$, and then only if $\sigma - C_M T$ increases. Similarly, the $M \to A$ transformation takes place in the band bounded by the lines $\sigma = C_A(T - A_s)$ and $\sigma = C_A(T - A_f)$, and only if $\sigma - C_A T$ decreases. Moreover, the coefficients C_A and C_M are very nearly equal. If the deformation accompanying the $A \to M$ transformation is identified with the inelastic deformation and denoted ε^i, then the union of the two bands constitutes the previously defined set $\mathscr{G}^1$ if $\mathscr{G}$ is identified with the σ-T plane. If we set $C_A = C_M = C'$ and define

$$
\begin{aligned}
F_1 &= \sigma - C'(T - M_f)\\
F_2 &= \sigma - C'(T - M_s)\\
F_3 &= \sigma - C'(T - A_s)\\
F_4 &= \sigma - C'(T - A_f)
\end{aligned}
$$

then

$$\mathscr{G}^1 = \mathscr{G}^1_{M \to A} \cup \mathscr{G}^1_{A \to M},$$

where

$$\mathscr{G}^1_{A \to M} = \{(\sigma, T) | F_1 F_2 \leqslant 0\}, \quad \mathscr{G}^1_{M \to A} = \{(\sigma, T) | F_3 F_4 \leqslant 0\},$$

while the loading surfaces are given by the lines

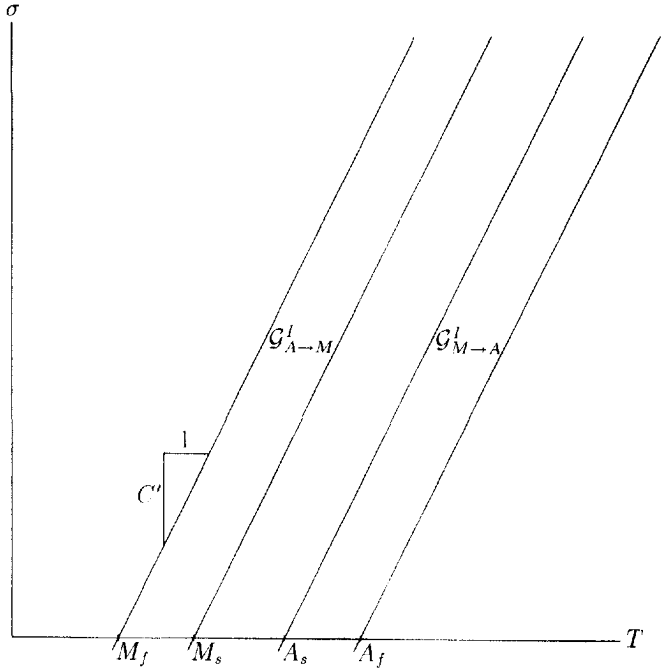

Fig. 2. Inelastic domains for $A \to M$ and $M \to A$ transformations.

$$F = \sigma - C'T = \text{constant}.$$

The geometry of the regions is shown in Fig. 2. It should be noted that the function defining the loading surfaces is independent of any internal variables, a simplification that nonetheless produces realistic results.

In a two-phase system, it can be assumed for the sake of simplicity that the only relevant internal variable (in addition to the inelastic deformation) is the fraction of mass occupied by one of the phases. For the sake of definiteness, let this variable be the mass fraction of martensite and let it be denoted ξ, so that $0 \leqslant \xi \leqslant 1$, with $\xi = 0$ denoting all austenite and $\xi = 1$ all martensite. Experiments indicate that the inelastic deformation is directly proportional to ξ, that is, $\varepsilon^I = \varepsilon_M \xi$, where ε_M is a constant equal to the maximum inelastic deformation, attained when the solid is all martensite.

It is important to acknowledge that this description represents a considerable oversimplification, in that it ignores the important fact that martensite, when first formed, may be present in multiple orientations (variants), and that it is not until the stress-induced reorientation of the variants takes place that relatively large deformation can take place. The extension of the model to account for reorientation is the subject of ongoing research.

The preceding results can be extended to general multiaxial loading by assuming that the inelastic potential and the loading surfaces are of Drucker–Prager form, and by replacing σ everywhere by ρ and ε^I by λ, as defined in the preceding section. Let the constants ε_M and C', moreover, be replaced by λ_M and C, respectively. Then $\lambda = \lambda_M \xi$, and the rate equation for ξ can be converted into one for λ, which can then be combined with the flow rule (6). Furthermore, $\phi(T) = CT$; thus $F = \rho - CT$, $F_1 = \rho - C(T - M_f)$, and so on.

4.3. *Rate equation for ξ*

The rate equation for ξ must reflect both the $A \to M$ and the $M \to A$ transformation. For the former, we may write

$$\dot{\xi}_{A\to M} = h_M \langle -F_1 F_2 \rangle \langle \dot{F} \rangle$$

and for the latter

$$\dot{\xi}_{M\to A} = -h_A \langle -F_3 F_4 \rangle \langle -\dot{F} \rangle.$$

since the production of austenite is equivalent to the loss of martensite. The quantities h_M and h_A may be assumed to be functions of F and ξ, and must reflect the fact the complete transformations lead ξ from 0 to 1 and from 1 to 0, respectively. Relatively simple forms that accomplish this goal, and that permit integration in closed form, are

$$h_M = \beta_M \frac{1-\xi}{|F_1 F_2| F^2}$$

and

$$h_A = \beta_A \frac{\xi}{|F_3 F_4| F_4^2},$$

where β_M and β_A are positive rate constants. The factors $1-\xi$ and ξ, respectively, represent a hypothesis of first-order reactions, while the indicated dependence on F is such as to permit closed-form solutions. For the complete $A \to M$ transformation beginning with $\xi = 0$ on the line $F_2 = 0$ (which corresponds to $F_1 = -C(M_s - M_f)$), the solution is

$$\xi = 1 - \exp\left[-\beta_M\left(\frac{1}{C(T-M_f)-\rho} - \frac{1}{C(M_s-M_f)}\right)\right],$$

while for the complete $M \to A$ transformation beginning with $\xi = 1$ on $F_3 = 0$ (corresponding to $F_4 = C(A_f - A_s)$), it is

$$\xi = \exp\left[-\beta_A\left(\frac{1}{\rho - C(T-A_f)} - \frac{1}{C(A_f-A_s)}\right)\right].$$

Since $\dot{\xi}_{A\to M}\dot{\xi}_{M\to A} = 0$, we may let $\dot{\xi} = \dot{\xi}_{A\to M} + \dot{\xi}_{M\to A}$, or

$$\dot{\xi} = \beta_M \frac{(1-\xi)\langle -F_1 F_2\rangle}{|F_1 F_2| F_1^2}\langle \dot{F}\rangle + \beta_A \frac{\xi\langle -F_3 F_4\rangle}{|F_3 F_4| F_4^2}\langle -\dot{F}\rangle. \tag{10}$$

Equation (10), combined with the relation $\lambda = \lambda_M \xi$ and substituted in eqn (6), gives an explicit equation for the inelastic strain rate. These equations can be integrated, numerically if necessary, to give the response to an arbitrary stress–temperature input.

4.4. *Numerical examples*

For simple loading paths the equations can be integrated in closed form. In what follows we consider three examples in which closed-form solution to eqn (10) can be found.

For the material parameters we use numerical values based on (though not exactly equal to) those reported by Chrysochoos (1993) for a Cu–Zn–Al alloy. Specifically, we set

$$E = 7\,\text{GPa}, \quad \varepsilon_M = 10\%, \quad C_M = C_A = 1\,\text{MPa}/^\circ\text{C} \quad \beta_A = \beta_M = 3\,\text{MPa},$$
$$M_f = 5^\circ\text{C}, \quad M_s = 40^\circ\text{C}, \quad A_s = 60^\circ\text{C}, \quad A_f = 90^\circ\text{C},$$

where E is the elastic modulus. All numerical tests are begun with the specimen in the parent phase ($\xi = 0$). The figures corresponding to the examples show stress against strain.

Example 1 (shape-memory effect). The initial temperature is set to $T = 50$ ($M_s < T < A_s$). The specimen is subjected first to a stress cycle, keeping the temperature constant, and then to a thermal cycle at zero stress, consisting of heating to $T = 100 > A_f$

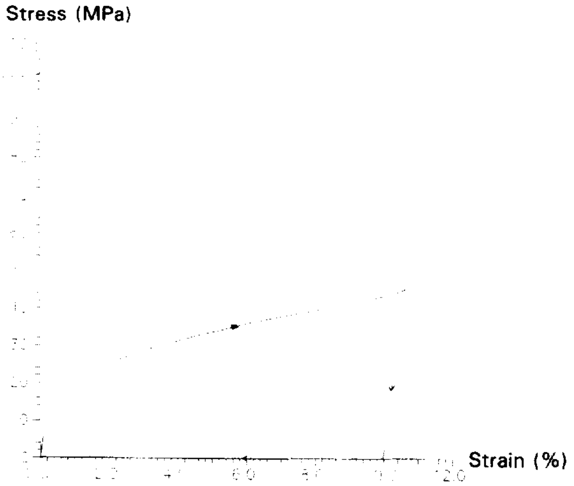

Fig. 3. Shape-memory effect (stress cycle at constant temperature between M_s and A_s, followed by heating above A_f and cooling at zero stress).

and cooling back to the initial temperature. The results are shown in Fig. 3. At the end of the stress cycle the material is completely in the martensitic state (since $T < A_f$, the martensite is stable at zero stress) and accordingly shows some permanent deformation. However, this permanent deformation is recovered after the thermal cycle; in the course of heating, the martensite is completely transformed into austenite. This analysis shows that the model can predict the shape-memory effect.

Example 2 (*pseudoelasticity*). Keeping the temperature constant ($T = 100 > A_f$), we load the specimen so as to have a complete stress-induced transformation (from austenite to martensite); upon unloading, a complete reverse transformation occurs (from martensite to austenite), since martensite is unstable at temperatures greater than A_f and zero stress. The results are shown in Fig. 4.

Example 3 (*pseudoelasticity with partial unloading and reloading*). We test the behavior of the model under multiple stress cycles, while keeping the temperature constant, i.e. $T = 100 > A_f$. In particular, we consider the case of partial unloading and reloading where *partial reloading* implies an incomplete direct transformation (from austenite to martensite), while *partial unloading* implies an incomplete reverse transformation (from martensite to

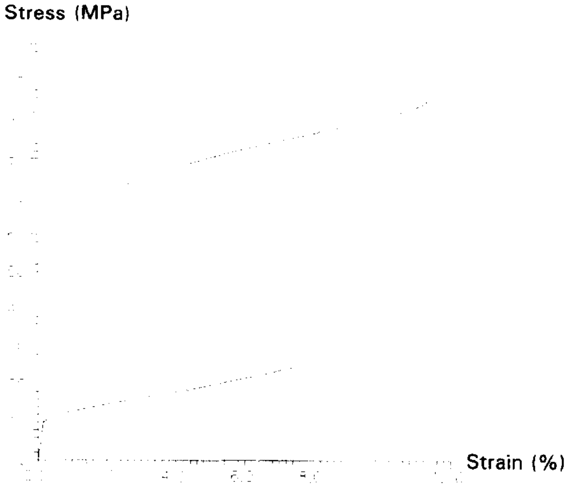

Fig. 4. Pseudoelasticity (stress cycle at constant temperature above A_f).

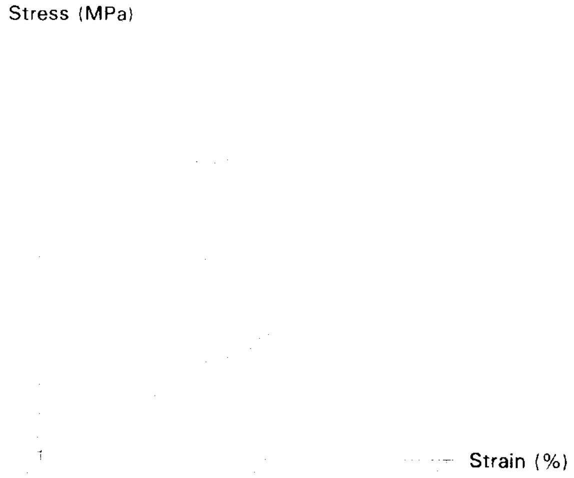

Fig. 5. Pseudoelasticity with partial unloading and reloading.

austenite). The results are shown in Fig. 5. The model presents the appropriate qualitative behavior, as experimentally described by several investigators, such as Cory and McNichols (1987) and Müller and Xu (1991). In particular, the stress–strain curve describes a series of loops, which are internal to the complete loading–unloading cycle; the internal loops exhibit ratcheting, which stabilizes after a few cycles.

4.5. *Conclusion*

In the present work we have reviewed the theory of generalized plasticity and have shown it to be a convenient framework for modeling materials undergoing phase transitions. In particular, we specialized the theory to model the martensitic transformation (austenite ↔ martensite) that occurs in shape-memory alloys, and validated its applicability to the modeling of such materials by means of examples showing the simulation of the pseudoelastic and shape-memory effects.

In future work we shall discuss the construction of more sophisticated rate equations for the martensite fraction and the extension of the model to the case of multiple transformations (such as detwinning and R-phase transformation).

REFERENCES

Abeyaratne, R. and Knowles, J. K. (1993). *J. Mech. Phys. Solids* **41,** 541–571.
Auricchio, F. and Taylor, R. L. (1994). *Int. J. Plasticity* **11,** 65–98.
Brandon, D. and Rogers, R. C. (1992). *J. Intell. Mater. Sys. Structures* **3,** 255–267.
Brinson, L. C. (1993). *J. Intell. Mater. Sys. Structures* **4,** 229–242.
Chrysochoos, A. (1993). *C.R. Ac. Sci. Ser. II.* **316,** 1031–1036.
Coleman, B. D. and Gurtin, M. E. (1967). *J. Chem. Phys.* **47,** 597–613.
Cory, J. S. and McNichols, J. L. Jr (1985). *J. Appl. Phys.* **58,** 3282–3294.
Cory, J. S. and McNichols, J. L. Jr. (1987). *J. Appl. Phys.* **61,** 972–984.
Darve, F., Boulon, M. and Chambon, R. (1978). *J. Mec.* **17,** 679–716.
Darve, F. and Labanieh, S. (1982). *Int. Numer. Analyt. Meth. Geomech.* **6,** 243–275.
Eisenberg, M. A. and Philips, A. (1971). *Acta Mech.* **11,** 247–260.
Falk, F. and Konopka, P. (1990). *J. Phys.: Condensed Matter* **2,** 61–77.
Fiosdick, R. L. and Serrin, J. (1975). *Arch. Ration. Mech. Anal.* **59,** 97–109.
Funakubo, H. (1987). *Shape Memory Alloys* (tr. J. B. Kennedy). Gordon and Breach, New York.
Ivshin, Y. and Pence, T. J. (1993). In *International Conference on Martensitic Transformations* (*ICOMAT*-92) (eds C. M. Wayman and J. Perkins), pp. 389–394. Monterey Institute of Advanced Studies, Carmel, CA.
Ivshin, Y. and Pence T. J. (1994). *Int. J. Engng Sci.* **32,** 681–704.
Kelley, J. L. (1975). *General Topology*. Springer, New York
Khachaturyan, A. G. (1983). *Theory of Structural Transformations in Solids*. Wiley, New York.
Liang, C. and Rogers C. A. (1990). *J. Intell. Mater. Sys. Structures* **1,** 207–234.
Liang, C. and Rogers, C. A. (1992). *J. Engng Math.* **26,** 429–443.
Lubliner, J. (1972). *Int. J. Non-Linear Mech.* 7, 237–254.
Lubliner, J. (1984). *Acta Mech.* **52,** 225–237
Lubliner, J. (1987). In *Thermomechanical Coupling in Solids* (eds H. D. Bui and Q. S. Nguyen), pp. 121–133. Elsevier, Amsterdam.

Lubliner, J. (1990). *Plasticity Theory*. Macmillan, New York.
Lubliner, J. (1991). *Int. J. Solids Structures* **28**, 769–778.
Lubliner, J., Taylor, R. L. and Auricchio, F. (1993). *Int. J. Solids Structures* **30**, 3171–3184.
Melton, K. N. (1990). In *Engineering Aspects of Shape Memory Alloys*, (eds T. W. Duerig *et al.*), pp. 21–35. Butterworth Heinemann, London.
Müller, I. and Xu, H. (1991). *Acta Metallurg. Mater.* **39**, 263–271.
Owen, D. R. (1968). *Arch. Ration. Mech. Anal.* **31**, 91–112.
Owen, D. R. (1970). *Arch. Ration. Mech. Anal.* **37**, 85–110.
Patoor, E., Eberhardt, A. and Berveiller, M. (1988). *Arch. Mech.* **40**, 755–794.
Pipkin, A. C. and Rivlin, R. S. (1965). *Z. Angew. Math. Mech.* **16**, 313–327.
Raniecki, B. and Lexcellent, C. (1994). *Eur. J. Mech. A* **13**, 21–50.
Sun, Q. P. and Hwang, K. C. (1993a). *J. Mech. Phys. Solids* **41**, 1–17.
Sun, Q. P. and Hwang, K. C. (1993b). *J. Mech. Phys. Solids* **41**, 19–33.
Tanaka, K. (1986). *Res. Mech.* **18**, 251–263.
Tanaka, K. and Iwasaki, R. (1985). *Engng Fract. Mech.* **21**, 709–720.
Tanaka, K. and Nagaki, S. (1982). *Ing.-Arch.* **51**, 287–299.
Tanaka, K., Hayashi, T., Itoh, Y. and Tobushi, H. (1992). *Mech. Mater.* **13**, 207–215.
Tobushi, H., Iwanaga, H., Tanaka, K., Hori, T. and Sawada, T. (1991). *Contin. Mech. Thermodyn.* **3**, 79–93.
Wilmanski, K. (1993). *Int. K. Engng Sci.* **31**, 1121–1138.

APPENDIX

In what follows, all topological notions regarding subsets of $\mathscr{S}_{\mathbf{q}}$ will refer to the relative topology induced by $\mathscr{S}_{\mathbf{q}}$.

When a mapping is defined on $\mathscr{S}_{\mathbf{q}}$ with values that are sets, such as $\mathscr{E}(\cdot,\mathbf{q})$ or $\partial\mathscr{E}(\cdot,\mathbf{q})$, we may wish to specify that such a mapping is continuous, and for that purpose we need a topology on a space of sets, namely, the space of non-empty closed subsets of $\mathscr{S}_{\mathbf{q}}$; this space is often denoted $2^{\mathscr{S}_{\mathbf{q}}}$. Fortunately, such a topology exists; it is defined by a metric introduced by Hausdorff and named for him [see, for example, Kelley (1975)]. If d denotes a metric on $\mathscr{S}_{\mathbf{q}}$ (e.g. the natural finite-dimensional metric), then the corresponding Hausdorff metric $\bar{d}$ on $2^{\mathscr{S}_{\mathbf{q}}}$ may be defined by

$$\bar{d}(A,B) = \max\left\{\sup_{\mathbf{G}\in B} d(\mathbf{G},A),\ \sup_{\mathbf{G}\in A} d(\mathbf{G},B)\right\},$$

for any $A, B \in 2^{\mathscr{S}_{\mathbf{q}}}$, where $d(\mathbf{G},A) \overset{\text{def}}{=} \inf_{\mathbf{G}^*\in A} d(\mathbf{G},\mathbf{G}^*)$.

We now prove the following.

Proposition 1. If $\mathbf{G}\in\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ and if the mapping $\partial\mathscr{E}(\cdot,\mathbf{q}):\mathscr{S}_{\mathbf{q}}\to 2^{\mathscr{S}_{\mathbf{q}}}$ is continuous at $\mathbf{G}$, then $\mathbf{G}$ is an interior point of $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$.

Proof. Assume the contrary, that is, $\mathbf{G}\in\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$. Since $\mathbf{G}\in\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ by hypothesis, we have $\mathbf{G}\in\mathring{\mathscr{E}}(\mathbf{G},\mathbf{q})$ by definition and, therefore, if we define $B=\partial\mathscr{E}(\mathbf{G},\mathbf{q})$ to simplify the notation, we have $r=d(\mathbf{G},B)>0$. But if $\mathbf{G}\in\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ then any neighborhood of $\mathbf{G}$ (in $\mathscr{S}_{\mathbf{q}}$) contains a $\mathbf{G}^*\notin\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$; that is, $(\mathbf{G}^*,\mathbf{q})$ is a plastic state, so that by definition $\mathbf{G}^*\in B^*\overset{\text{def}}{=}\partial\mathscr{E}(\mathbf{G}^*,\mathbf{q})$. By the triangle inequality, $d(\mathbf{G},B)\leqslant d(\mathbf{G}^*,B)+d(\mathbf{G},\mathbf{G}^*)$; hence $d(\mathbf{G}^*,B)\geqslant r-d(\mathbf{G},\mathbf{G}^*)$. Furthermore, $d(B,B^*)\geqslant d(\mathbf{G}^*,B)$; consequently $\lim_{d(\mathbf{G},\mathbf{G}^*)\to 0} d(B,B^*)\geqslant r$. Thus the assumption that $\partial\mathscr{E}(\cdot,\mathbf{q})$ is continuous is violated.

Corollary. If, at each $\mathbf{q}$, $\partial\mathscr{E}(\cdot,\mathbf{q})$ is continuous in $\mathscr{S}_{\mathbf{q}}$, then $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ is open in $\mathscr{S}_{\mathbf{q}}$.

Before stating the next proposition, we need the following purely topological result:

Lemma. Let X be a topological space, A a non-empty open set in X, and B a closed set in X whose interior is connected. If $A\subset B$ and $\partial A\subset\partial B$, then $A=\mathring{B}$.

Proof. Clearly $A\subset\mathring{B}$, because if there were a point $x\in A\cap\partial B$, then any neighborhood of x would contain points that do not belong to B and, therefore, do not belong to A, so that A would not be an open set. Now $X-(\mathring{B}-A)=(X-\mathring{B})\cup A=(X-\bar{B})\cup\partial B\cup A$; but $\partial B=\partial B\cup\partial A$ by hypothesis, so that $X-(\mathring{B}-A)=(X-\bar{B})\cup\partial B\cup\partial A\cup A=(X-\mathring{B})\cup\bar{A}$, the union of two closed sets and hence a closed set. Therefore, $\mathring{B}-A$ is open. If, however, $\mathring{B}-A$ is not empty, then $\mathring{B}$ is the union of two disjoint non-empty open sets and is, therefore, not connected. Consequently $A=\mathring{B}$.

Proposition 2. If $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ is path-connected, then for every $\mathbf{G}\in\overline{\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}}$, $\mathscr{E}(\mathbf{G},\mathbf{q})=\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$.

Proof. For every $\mathbf{G},\mathbf{G}^*\in\overline{\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}}$ there is, by the hypothesis of path-connectedness, a curve lying entirely (except possibly the end points) in $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ and joining $\mathbf{G}$ and $\mathbf{G}^*$. Consequently $\mathbf{G}^*\in\mathscr{E}(\mathbf{G},\mathbf{q})$ and $\mathbf{G}\in\mathscr{E}(\mathbf{G}^*,\mathbf{q})$, so that $\mathscr{E}(\mathbf{G},\mathbf{q})$ is the same for all $\mathbf{G}\in\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$; call it $\mathscr{E}_{\mathbf{q}}^0$. Clearly $\overline{\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}}\subset\mathscr{E}_{\mathbf{q}}^0$. Now consider $\mathbf{G}\in\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$; then $(\mathbf{G},\mathbf{q})$ is a plastic state, hence $\mathbf{G}\in\partial\mathscr{E}(\mathbf{G},\mathbf{q})=\partial\mathscr{E}_{\mathbf{q}}^0$, so that $\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}\subset\partial\mathscr{E}_{\mathbf{q}}^0$. We deduce (with the help of the preceding lemma) that $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}=\mathring{\mathscr{E}}_{\mathbf{q}}^0$.

Corollary. If $\mathscr{E}(\mathbf{G},\mathbf{q})$ equals the closure of its interior, then for every $\mathbf{G}\in\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$ we have $\mathscr{E}(\mathbf{G},\mathbf{q})=\overline{\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}}$ and $\partial\mathscr{E}(\mathbf{G},\mathbf{q})=\partial\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}$.

The following two-sided inclusion result is useful for the transition to classical plasticity.

Proposition 3. At every possible $\mathbf{q}$,

$$\bigcap_{\mathbf{G}\in\mathscr{S}_{\mathbf{q}}}\mathscr{E}(\mathbf{G},\mathbf{q})\subset\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}\subset\bigcup_{\mathbf{G}\in\mathscr{S}_{\mathbf{q}}}\mathring{\mathscr{E}}(\mathbf{G},\mathbf{q}).$$

Proof. The intersection is the set $\{\mathbf{G}\mid\mathbf{G}\in\mathscr{E}(\mathbf{G}^*,\mathbf{q})$ for *all* $\mathbf{G}\in\mathscr{S}_{\mathbf{q}}\}$; the union is the set $\{\mathbf{G}\mid\mathbf{G}\in\mathring{\mathscr{E}}(\mathbf{G}^*,\mathbf{q})$ for *any* $\mathbf{G}^*\in\mathscr{S}_{\mathbf{q}}\}$; and $\mathscr{S}_{\mathbf{q}}^{\mathrm{E}}=\{\mathbf{G}\mid\mathbf{G}\in\mathring{\mathscr{E}}(\mathbf{G}^*,\mathbf{q})$ for a particular $\mathbf{G}^*$, namely $\mathbf{G}^*=\mathbf{G}\}$.