PERGAMON

Journal of the Mechanics and Physics of Solids
47 (1999) 1239–1263

JOURNAL OF THE
MECHANICS AND
PHYSICS OF SOLIDS

Mechanism-based strain gradient plasticity—I. Theory

H. Gao[a,*], Y. Huang[b], W.D. Nix[c], J.W. Hutchinson[d]

[a] *Division of Mechanics and Computation, Stanford University, Stanford, California 94305, U.S.A.*
[b] *Department of Mechanical and Industrial Engineering, University of Illinois, Urbana, Illinois 61801, U.S.A.*
[c] *Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, U.S.A.*
[d] *Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, U.S.A.*

Received 22 May 1998; received in revised form 28 September 1998

Abstract

A mechanism-based theory of strain gradient plasticity (MSG) is proposed based on a multiscale framework linking the microscale notion of statistically stored and geometrically necessary dislocations to the mesoscale notion of plastic strain and strain gradient. This theory is motivated by our recent analysis of indentation experiments which strongly suggest a linear dependence of the square of plastic flow stress on strain gradient. While such linear dependence is predicted by the Taylor hardening model relating the flow stress to dislocation density, existing theories of strain gradient plasticity have failed to explain such behavior. We believe that a mesoscale theory of plasticity should not only be based on stress–strain behavior obtained from macroscopic mechanical tests, but should also draw information from micromechanical, gradient-dominant tests such as micro-indentation or nano-indentation. According to this viewpoint, we explore an alternative formulation of strain gradient plasticity in which the Taylor model is adopted as a founding principle. We distinguish the microscale at which dislocation interaction is considered from the mesoscale at which the plasticity theory is formulated. On the microscale, we assume that higher order stresses do not exist, that the square of flow stress increases linearly with the density of geometrically necessary dislocations, strictly following the Taylor model, and that the plastic flow retains the associative structure of conventional plasticity. On the mesoscale, the constitutive equations are constructed by averaging microscale plasticity laws over a representative cell. An expression for the effective strain gradient is obtained by considering models of geometrically necessary dislocations associated with bending, torsion and 2-D axisymmetric void growth. The new theory differs

* Corresponding author. Fax: 001 415 723 1778; e-mail: gao@am-sun2.stanford.edu

PII: 0022–5096(98)00103–3

from all existing phenomenological theories in its mechanism-based guiding principles, although the mathematical structure is quite similar to the theory proposed by Fleck and Hutchinson. A detailed analysis of the new theory is presented in Part II of this paper.

Keywords: A. Dislocations; Strengthening mechanisms; B. Constitutive behavior; Principles; Elastic–plastic material; Strain gradient plasticity

1. Introduction

1.1. *Strain gradient plasticity*: *An overview*

Recent experiments have shown that materials display strong size effects when the characteristic length scale associated with non-uniform plastic deformation is on the order of microns. For example, Fleck et al. (1994) observed in the twisting of thin copper wires that the scaled shear strength increases by a factor of three as the wire diameter decreases from 170–12 microns, while the increase of work-hardening in simple tension is negligible. In bending of ultra thin beams, Stolken and Evans (1998) observed a significant increase in the normalized bending hardening as the beam thickness decreases from 100–12.5 microns, while data from simple tension displays no size dependence. For an aluminum–silicon matrix reinforced by silicon carbide particles, Lloyd (1994) observed a substantial strength increase when the particle diameter was reduced from 16–7.5 microns with the particle volume fraction fixed at 15%. More convincing experimental evidence of the size dependence of material behavior at the micron level comes from the micro-indentation or nano-indentation hardness test. The measured indentation hardness of metallic materials increases by a factor of two as the depth of indentation decreases from 10 microns to 1 micron (Nix, 1989; De Guzman et al., 1993; Stelmeshenko et al., 1993; Ma and Clarke, 1995; Poole et al., 1996; McElhaney et al., 1998).

The classical plasticity theories can not predict this size dependence of material behavior at the micron scale because their constitutive models possess no internal length scale. The predictions based on the classical plasticity theories for non-uniform deformation (e.g., twisting, bending, deformation of composites, micro- or nano-indentation) do not exhibit a size dependence after normalizations. However, there is an impending need to deal with design and manufacturing issues at the micron level, such as in thin films where film thickness is on the order of 1 micron or less; sensors, actuators and microelectromechanical systems (MEMS) where the entire system size is less than 10 microns; microelectronic packaging where features such as vias are smaller than 10 microns; advanced composites where particle or fiber size is on the order of 10 microns; as well as in micromachining. The current design tools, such as finite element analysis (FEA) and computer aided design (CAD), are based on classical continuum theories, which may not be suitable at such a small length scale.

On the other hand, it is still not possible to perform quantum and atomistic simulations on realistic time and length scales required for the micron level structures. A continuum theory for micron level applications is thus necessary to bridge the gap between conventional continuum theories and atomistic simulations.

Another objective that warrants the development of a micron level continuum theory is to link macroscopic fracture behavior to atomistic fracture processes in ductile materials. In a remarkable series of experiments, Elssner et al. (1994) measured both the macroscopic fracture toughness and atomic work of separation of an interface between a single crystal of niobium and a sapphire single crystal. The macroscopic work of fracture was measured using a four-point bend specimen designed for the determination of interfacial toughness, while the atomic value was inferred from the equilibrium shapes of microscopic pores on the interface. The interface between the two materials remained atomistically sharp, i.e. the crack tip was not blunted even though niobium is ductile and has a large number of dislocations. The stress level needed to produce atomic decohesion of a lattice or a strong interface is typically on the order of 0.03 times Young's modulus, or 10 times the tensile yield stress. Hutchinson (1997) pointed out that the maximum stress level that can be achieved near a crack tip is not larger than 4–5 times the tensile yield stress of metals, according to models based on conventional plasticity theories. This clearly falls short of triggering the atomic decohesion observed in Elssner et al.'s (1994) experiments. Attempts to link macroscopic cracking to atomistic fracture are frustrated by the inability of conventional plasticity theories to model stress–strain behavior adequately at the small scales involved in crack tip deformation,

Apparently, some microscopic understanding of plasticity is necessary in order to accurately describe deformation at small length scales. When a material is deformed, dislocations are generated, moved, and stored, and the storage causes the material to work harden. Dislocations become stored for one of two reasons: they accumulate by trapping each other in a random way, or they are required for compatible deformation of various parts of the material. In the former case the dislocations are referred to as statistically stored dislocations (Ashby, 1970), while in the latter case they are called geometrically necessary dislocations and are related to the gradients of plastic shear in a material (Nye, 1953; Cottrell, 1964; Ashby, 1970). Plastic strain gradients appear either because of the geometry of loading or because of inhomogeneous deformation in the material, as in the aforementioned experiments. These considerations have motivated Fleck and Hutchinson (1993, 1997) and Fleck et al. (1994) to develop a phenomenological theory of strain gradient plasticity intended for applications to materials and structures whose dimension controlling plastic deformation falls roughly within a range from a tenth of a micron to ten microns. This theory has been applied to many problems where strain gradient effects are expected to be important, including analyses of crack tip fields (Huang et al., 1995, 1997; Xia and Hutchinson, 1996). The Fleck–Hutchinson theory fits the mathematical framework of higher order continuum theories of elasticity (Toupin, 1962; Koiter, 1964; Mindlin, 1964, 1965), with the strain gradients represented either in terms of the gradients of rotation in the couple-stress theory of strain gradient plasticity (Fleck and Hutchinson, 1993; Fleck et al., 1994) or in terms of both rotation and stretch gradients in a more general isotropic-hardening theory based on all the quadratic invariants of the strain gradient tensor (Fleck and Hutchinson, 1997). The couple stress theory used by Fleck and Hutchinson (1993) also bears some resemblance to the early work of Kroener (1963) who studied the connection between lattice curvature associated with dislocations and couple stresses and developed a non-local continuum theory based on that connection. The work-conjugate of the rotation and/or stretch gradient of deformation defines the higher order stress which is required for this class of strain gradient theory to satisfy the Clausius–Duhem thermodynamic restrictions on the constitutive model for second deformation gradients (Gurtin, 1965a, b; Acharya and Shawki, 1995). In comparison, no work conjugate of strain gradient has been defined in the alternative gradient theories (Aifantis, 1984; Zbib and Aifantis, 1989; Muhlhaus and Aifantis, 1991) which represent the strain gradient effects as the first and second Laplacian of effective strain. Acharya and Bassani (1995) have considered possible formulations of strain gradient plasticity which retain the essential structure of conventional plasticity and obey thermodynamic restrictions; they conclude that the only possible formulation is a flow theory with strain gradient effects represented as an internal variable which acts to increase the current tangent-hardening modulus. However, there has not been a systematic way of constructing the tangent modulus so as to validate this framework. The internal variable formulation has also been adopted by Dai and Parks (1998) in their formulation of a rate-dependent, single crystal theory of strain gradient plasticity.

From a dimensional consideration, an internal constitutive length parameter, l, was introduced to scale the rotational gradient terms in the couple-stress theory of strain gradient plasticity (Fleck and Hutchinson, 1993; Fleck et al., 1994). This length scale is thought of as an internal material length related to the storage of geometrically necessary dislocations, and is found to be approximately 4 microns for copper from Fleck et al.'s (1994) twisting of thin wire experiments, and 6 microns for nickel from Stolken and Evans' (1998) bending of ultra-thin beam experiments. The contribution of the strain gradient could be symbolically represented as $l\,\mathrm{d}\varepsilon/\mathrm{d}x \sim \varepsilon(l/D)$ where D represents the characteristic length of the deformation field usually corresponding to the smallest dimension of geometry (e.g., thickness of a beam, radius of a void, depth of indentation). When D is much larger than the material length, l, the strain gradient terms become negligible in comparison with strains, and strain gradient plasticity then degenerates to the conventional plasticity theory. However, as D becomes comparable to l as in the aforementioned experiments, strain gradient effects begin to play a dominating role. The couple-stress theory of strain gradient plasticity has had some success in estimating the size dependence observed in the aforementioned torsion of thin wires (Fleck et al., 1994) and bending of thin beams (Stolken and Evans, 1998). However, its prediction of indentation hardness (Shu and Fleck, 1998) falls short of agreement with the significant increase of 200 or even 300% observed in micro-indention or nano-indentation tests (Nix, 1989; De Guzman et al., 1993; Stelmashenko et al., 1993; Ma and Clarke, 1995; Poole et al., 1996; McElhaney et al., 1998). For this reason, Fleck and Hutchinson (1997) proposed an extended theory of strain gradient plasticity theory which includes both rotation gradient and stretch gradient of the deformation in the constitutive model. The work-conjugates of rotation and stretch gradients of deformation are couple stress and higher order stress, respectively. Accordingly, two more internal material lengths are introduced in addition to l. Begley and Hutchinson (1998) determined these new lengths to be in the range of 0.22–0.6 microns by fitting indentation data.

1.2. *An experimental law of strain gradient plasticity*

Recent analysis of indentation experiments by Nix and Gao (1998) has shed some light on both the material length l introduced by Fleck and Hutchinson (1993) and the experimental law needed to advance a mechanism-based theory of strain gradient plasticity. Nix and Gao started from the Taylor relation between the shear strength and dislocation density in a material,

$$\tau = \alpha\mu b\sqrt{\rho_T} = \alpha\mu b\sqrt{\rho_S + \rho_G} \tag{1}$$

where ρ_T is the total dislocation density, ρ_S is the density of statistically stored dislocations, ρ_G is the density of geometrically necessary dislocations, μ is the shear modulus, b is the Burgers vector and α is an empirical constant usually ranging from 0.2–0.5. A gradient in the strain field is accommodated by geometrically necessary dislocations, so that an effective strain gradient η can be defined as

$$\eta = \rho_G b. \tag{2}$$

This expression allows η to be interpreted as the curvature of deformation under bending and twist per unit length under torsion.

If the von Mises rule is used, the tensile flow stress can be written as

$$\sigma = \sqrt{3}\tau = \sqrt{3}\alpha\mu b\sqrt{\rho_S + \eta/b}. \tag{3}$$

In the absence of the strain gradient term, we identify the uniaxial stress–strain law

$$\sigma = \sqrt{3}\alpha\mu b\sqrt{\rho_S} = \sigma_Y f(\varepsilon) \tag{4}$$

as the hardening due to the statistically stored dislocations alone. For convenience, we define the state of plastic yield as

$$\sigma = \sigma_Y, \quad \varepsilon = \varepsilon_Y, \quad f(\varepsilon_Y) = 1, \tag{5}$$

where ε_Y is usually taken as 0.2% for ductile metals. Combining (3) and (4) leads to a law for strain gradient plasticity (Nix and Gao, 1998),

$$\sigma = \sigma_Y\sqrt{f^2(\varepsilon) + l\eta}, \tag{6}$$

where

$$l = 3\alpha^2\left(\frac{\mu}{\sigma_Y}\right)^2 b \tag{7}$$

is identified as the material length introduced by Fleck and Hutchinson (1993, 1997). In terms of macroscopic properties of structural metals, the ratio between μ and σ_Y

is typically in the order of 100, suggesting that l is in the order of 10^4b, which is indeed in the order of microns, in agreement with the estimation from twisting of thin copper wires (Fleck et al., 1994) and bending of ultra-thin beams (Stolken and Evans, 1998).

Nix and Gao (1998) developed a dislocation model to estimate the density of geometrically necessary dislocations underneath a conical indenter, and found that the strain gradient law of (6) implies the following relation for the indentation hardness H

$$\frac{H}{H_0} = \sqrt{1 + \frac{h^*}{h}}, \tag{8}$$

where H_0 is the hardness in the absence of strain gradient effects, h is the depth of indentation, and

$$h^* = \frac{81}{2} b\alpha^2 \tan^2\theta \left(\frac{\mu}{H_0}\right)^2, \tag{9}$$

θ being the angle between the surface of the indenter and the plane of the surface to be indented. The prediction of (8) agrees remarkably well with McElhaney et al.'s (1998) microindentation hardness data for single crystal and cold worked polycrystalline copper, as well as with Ma and Clarke's (1995) microindentation hardness data for single crystal silver. A re-examination of Poole et al.'s (1996) experimental data also confirms the linear relation between the square of indentation hardness and the inverse of indent depth. Figure 1 shows the excellent agreement of (8) with the depth dependence of the hardness of (111) single crystal copper measured by McElhaney et al. (1998). In contrast, conventional plasticity theories predict the indentation

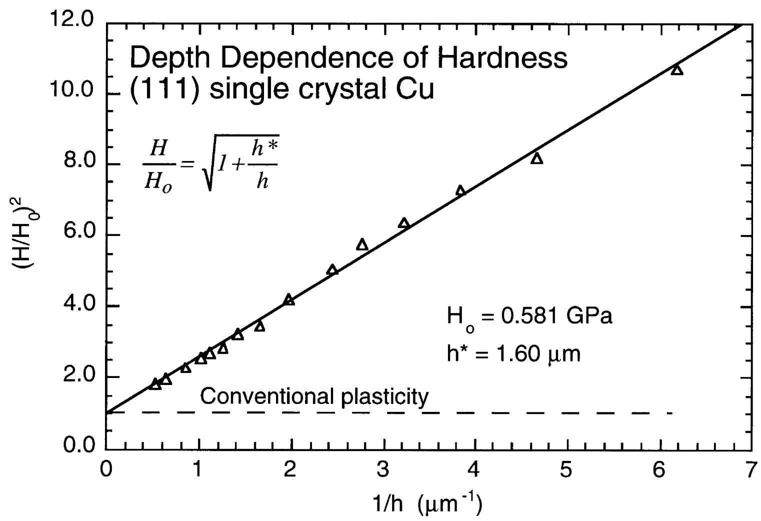

Fig. 1. Depth dependence of the hardness of (111) single crystal copper, plotted according to (8).

hardness to be independent of the indent depth, also shown in Fig. 1. This evidence highlights the need to advance strain gradient plasticity theories for small scale applications.

From a fundamental consideration, the Burgers vector and the dislocation spacing are two physical length scales which control plastic deformation. Nix and Gao (1998) found it helpful to examine an alternative form of (6),

$$\left(\frac{\sigma}{\sigma_0}\right)^2 = 1+\hat{l}\eta, \tag{10}$$

where $\sigma_0 = \sigma_Y f(\varepsilon)$ denotes the flow stress in the absence of strain gradients. The scaling length $\hat{l}$ is found to be related to the Burgers vector b and the mean spacing L_S between statistically stored dislocations,

$$\hat{l} = \frac{L_S^2}{b}. \tag{11}$$

For the case of materials strengthened by dispersoids or precipitates, $\hat{l}$ is on the order of L_P^2/b, where L_P is the mean spacing between particles. The material length l roughly corresponds to the value of $\hat{l}$ at yielding, $l \simeq \hat{l}_{\text{yield}}$.

1.3. *Motivation for mechanism-based strain gradient plasticity*

Fleck and Hutchinson (1993, 1997) used the dislocation theory to motivate their formulation of strain gradient plasticity. However, the actual theory was formulated by replacing effective stresses and strains in conventional plasticity with higher order effective stresses and strains which contain strain gradient terms scaled by a phenomenological material length to be determined from experiments. In other words, the Fleck–Hutchinson theory is developed, primarily based on the macroscopically measured uniaxial stress–strain behavior. Micromechanical experiments such as micro-indentation, micro-torsion and micro-bending were not used at the stage of theory construction, but rather were used to fit the material length l. The remarkable agreement between the strain gradient law of (6) and the micro-indentation data for various materials indicates that the linear relation between the square of indentation hardness and the inverse of indent depth represents a fundamental, intrinsic nature of deformation at the microscale. This provides a strong motivation to develop an alternative formulation in which the strain gradient law of (6) derived from the Taylor relation is incorporated as a fundamental postulate. In this paper, we propose a multiscale, hierarchical framework to facilitate such a marriage between plasticity and dislocation theory. A mesoscale cell with linear variation of strain field is considered. Each point within the cell is considered as a microscale sub-cell within which dislocation interaction is assumed to (approximately) obey the Taylor relation so that the strain gradient law of (6) applies. On the microscale, the η term is to be treated as a measure of the density of geometrically necessary dislocations whose accumulation increases the flow stress strictly following the Taylor model. In other words, microscale plastic flow is assumed to occur as slip of statistically stored dislocations in a background of geometrically necessary dislocations and the microscale plastic deformation is assumed to obey the Taylor work hardening relation and the associative laws of conventional plasticity. The notion of geometrically necessary dislocations is connected to the gradient of the strain field on the level of the mesoscale cell. Higher order stresses are introduced as thermodynamic conjugates of the strain gradients at the mesoscale level on which the plasticity theory is formulated. This ensures that the theory obeys the Clausius–Duhem thermodynamic restrictions of a continuum constitutive model. This hierarchical structure provides a systematic approach for constructing the mesoscale constitutive laws by averaging microscale plasticity laws over the representative cell. The new theory differs from all existing phenomenological theories in its mechanism-based guiding principles, although it fits nicely within the mathematical framework of the phenomenological theory by Fleck and Hutchinson (1997).

2. Basic postulates

Figure 2 shows the multiscale framework which we adopt to construct the MSG theory. On the microscale, the scale of analysis is small compared with the length over which the strain field varies, and the dislocation activities are described by the slip of statistically stored dislocations in a background of geometrically necessary dislocations which influence the microscale flow stress according to the strain gradient

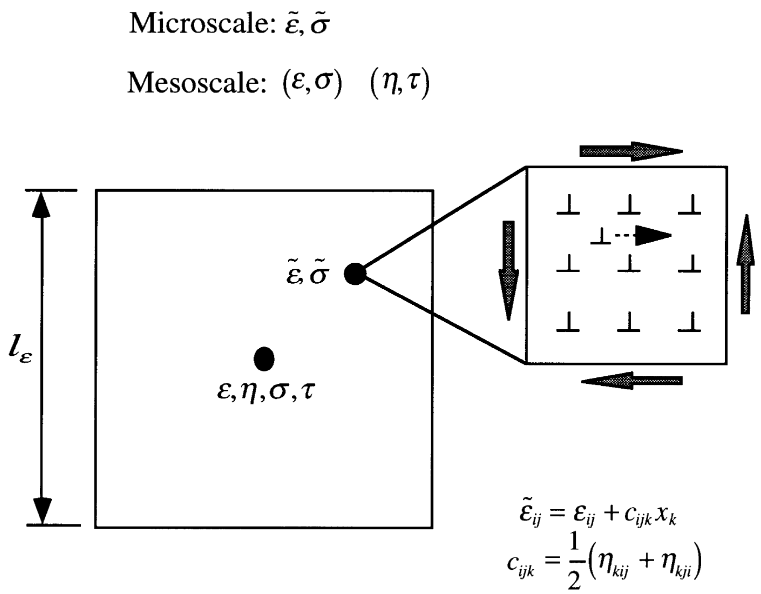

Fig. 2. The multiscale framework for strain gradient plasticity; dislocation interaction is considered on the microscale via the Taylor relation; the higher order continuum theory of strain gradient plasticity is established on the mesoscale representative cell.

law of (6). At this level of analysis, the stress and strain tensors are defined in the classical sense and will be denoted as $\tilde{\boldsymbol{\sigma}}$ and $\tilde{\boldsymbol{\varepsilon}}$, respectively. Concepts associated with strain gradient plasticity such as higher order stresses and strain gradients will be needed to ensure that the constitutive model we develop satisfies the essential thermodynamic restrictions. These concepts are introduced at a higher level of analysis, which will be referred to as the mesoscale analysis. Here the terminology 'mesoscale' is adopted because our continuum strain gradient plasticity theory is formulated based on dislocation activities on a subscale which we have referred to as the 'microscale'.

Why is such hierarchical framework necessary for strain gradient plasticity? The reason is that the microscale picture allows us to make a connection with the scenario envisioned by Taylor, in which the flow stress is defined as the critical stress to move a glide dislocation through a forest of obstacles. On the other hand, the mesoscale picture makes it possible to connect the notion of geometrically necessary dislocations to the gradient of the strain field and to institute a thermodynamically sound constitutive framework. In other words, the multiscale framework allows the concept of geometrically necessary dislocations to become a coherent part of the constitutive law. Our mesoscale variables include the stress $\boldsymbol{\sigma}$, strain $\boldsymbol{\varepsilon}$, the higher order stress $\boldsymbol{\tau}$ and strain gradient $\boldsymbol{\eta}$.

In formulating the MSG theory on the mesoscale, we will adopt the higher order framework laid out by Fleck and Hutchinson (1997), which was developed based on the earlier work of Toupin (1962), Koiter (1964) and Mindlin (1964, 1965). In this framework, the generalized strain variables are the symmetric strain tensor

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \tag{12}$$

and the second gradient of displacement

$$\eta_{ijk} = u_{k,ij}. \tag{13}$$

For simplicity, elastic deformation and compressibility of materials are ignored in the analysis of this paper. The condition of incompressibility can be stated as

$$\varepsilon_{ii} = 0, \quad \eta^H_{ijk} = \frac{1}{4}(\delta_{ik}\eta_{jpp} + \delta_{jk}\eta_{ipp}) = 0, \tag{14}$$

where η^H_{ijk} denotes the hydrostatic part of η_{ijk}, following the notation of Fleck and Hutchinson (1997). Under these restrictions, the work increment per unit volume of solid due to an arbitrary variation of displacement $\boldsymbol{u}$ is

$$\delta w = \sigma'_{ij}\delta\varepsilon_{ij} + \tau'_{ijk}\delta\eta_{ijk} \tag{15}$$

where σ'_{ij} denotes the deviatoric stress components, and τ'_{ijk} are the deviatoric components of the higher order stress tensor conjugated to the strain gradient tensor η_{ijk}.

In order to link the microscale picture of Taylor hardening and the mesoscale picture of strain gradient plasticity, we adopt the following three postulates in our constitutive framework:

(1) We assume that the microscale flow stress is governed by the dislocation motion, and that it obeys the Taylor hardening relation as exhibited by the strain gradient law of (6),

$$\tilde{\sigma} = \sigma_{\mathrm{Y}}\sqrt{f^2(\tilde{\varepsilon}) + l\eta} \tag{16}$$

(2) Strain gradient plasticity is a mesoscale description of dislocation activities and as such it must be derivable from the dislocation-based microscale plasticity laws. We choose a mesoscale cell which will be taken as small as possible so that the variation of strain field can be approximated as linear within the cell, and yet sufficiently large for application of the Taylor model. Higher order strain gradients (higher than the first order) are assumed to be negligible within the meso-cell. The microscale and mesoscale are linked by the plastic work equality

$$\int_{V_{\mathrm{cell}}} \tilde{\sigma}'_{ij}\delta\tilde{\varepsilon}_{ij}\,\mathrm{d}V = (\sigma'_{ij}\delta\varepsilon_{ij} + \tau'_{ijk}\delta\eta_{ijk})V_{\mathrm{cell}}. \tag{17}$$

(3) We assume that the essential structure of conventional plasticity is preserved on the microscale. This can be justified by linking the microscale plastic flow to slip of statistically stored dislocations in a background of geometrically necessary dislocations. If dislocation slip is assumed to be proportional to the resolved Schmid stress along an appropriate slip system, the associative rule of plastic normality holds (Rice, 1970, 1971),

$$\frac{\mathrm{d}\tilde{\varepsilon}_{ij}}{\mathrm{d}\tilde{\varepsilon}} = \frac{3\tilde{\sigma}'_{ij}}{2\tilde{\sigma}_e} \tag{18}$$

where

$$\tilde{\sigma}_e = \sqrt{\tfrac{2}{3}\tilde{\sigma}'_{ij}\tilde{\sigma}'_{ij}} \tag{19}$$

is the effective stress and

$$\mathrm{d}\tilde{\varepsilon} = \sqrt{\tfrac{2}{3}\mathrm{d}\tilde{\varepsilon}_{ij}\,\mathrm{d}\tilde{\varepsilon}_{ij}} \tag{20}$$

is the effective strain increment. For proportional, monotonically increasing deformation, the effective strain in the deformation theory of MSG plasticity can be defined as

$$\tilde{\varepsilon} = \sqrt{\tfrac{2}{3}\tilde{\varepsilon}_{ij}\tilde{\varepsilon}_{ij}}. \tag{21}$$

The microscale yield criterion is

$$\tilde{\sigma}_e = \tilde{\sigma} \tag{22}$$

where $\tilde{\sigma}$ is given by (16).

3. Dislocation models

3.1. *Density of geometrically necessary dislocations and effective strain gradient*

Before the theory of MSG can be established, the density of geometrically necessary dislocations $\rho_{\mathrm{G}} = \eta/b$ must be related to the components of the strain gradient tensor η_{ijk}. Following Fleck and Hutchinson (1997), we define

$$\eta = \sqrt{c_1 \eta_{iik}\eta_{jjk} + c_2 \eta_{ijk}\eta_{ijk} + c_3 \eta_{ijk}\eta_{kji}} \tag{23}$$

as an effective strain gradient which measures the density of geometrically necessary dislocations; the three constants, c_1, c_2 and c_3 scale the three quadratic invariants for the incompressible third order tensor η_{ijk}. Fleck and Hutchinson (1997) attempted to determine c_1, c_2 and c_3 from experimental data. Due to the scarcity of experiments on strain gradient effects, we shall take a different approach. We shall determine the three constants from a series of distinct dislocation models consisting of plane strain bending, pure torsion and 2-D axisymmetric void growth. There may be multiple configurations of geometrically necessary dislocations for a given continuum strain field, in which case the constants cannot be uniquely determined. These configurations are often influenced by the orientations of available slip systems. Without accounting for such complications, we will assume an abundance of slip systems and consider the most efficient dislocation configuration which corresponds to the minimum density of geometrically necessary dislocations required to accommodate a given deformation field.

For plane strain bending, the most efficient dislocation configuration is shown in Fig. 3(a). The plastic bending of a crystal of curvature κ can be accomplished by introducing a density,

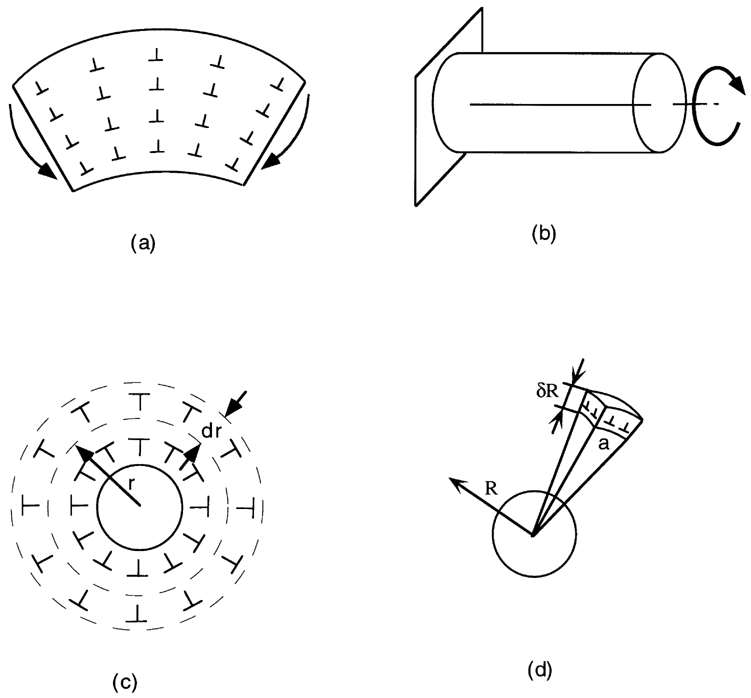

Fig. 3. Models of geometrically necessary dislocations associated with (a) plane strain bending, (b) torsion and (c) 2-D axisymmetric void growth; (d) 3-D spherical void growth.

$$\rho_G = \eta/b \tag{24}$$

of dislocations of Burgers vector b (Nye, 1953; Ashby, 1970; Stolken and Evans, 1998) where

$$\eta = \kappa. \tag{25}$$

Assuming the mean spacing between geometrically necessary dislocations is L_G, this relation can be obtained from the similarity relation

$$b/L_G = \kappa L_G, \quad \rho_G = 1/L_G^2 = \kappa/b. \tag{26}$$

On the other hand, Huang et al. (1998) have shown that the displacement field for plane strain bending consistent with MSG is

$$u_1 = \kappa x_1 x_2, \quad u_2 = -\tfrac{1}{2}\kappa(x_1^2 + x_2^2). \tag{27}$$

The non-zero components of the third-order strain gradient tensor $\eta_{ijk} = u_{k,ij}$ are then

$$\eta_{112} = -\kappa, \quad \eta_{211} = \eta_{121} = \kappa, \quad \eta_{222} = -\kappa \tag{28}$$

with quadratic invariants equal to

$$\eta_{iik}\eta_{jjk} = 4\kappa^2, \quad \eta_{ijk}\eta_{ijk} = 4\kappa^2, \quad \eta_{ijk}\eta_{kji} = 0. \tag{29}$$

Combining (23), (25) and (29), the dislocation model for plane strain bending leads to

$$c_1 + c_2 = \tfrac{1}{4}. \tag{30}$$

For simple torsion of a cylinder of radius R shown in Fig. 3(b), the most efficient dislocation configuration is a screw dislocation lying along the axis of the cylinder. The displacement field associated with such a coaxial screw dislocation is (Eshelby, 1953; Hirthe and Lothe, 1982)

$$u_1 = -\frac{b}{\pi R^2} x_2 x_3, \quad u_2 = \frac{b}{\pi R^2} x_1 x_3, \tag{31}$$

which has the identical form as the displacement of pure torsion (Huang et al., 1998),

$$u_1 = -\kappa x_2 x_3, \quad u_2 = \kappa x_1 x_3 \tag{32}$$

with twist per unit length equal to

$$\kappa = \frac{b}{\pi R^2}. \tag{33}$$

Within a length H of the cylinder, the density of geometrically necessary dislocations is

$$\rho_G = \frac{H}{\pi R^2 H} = \frac{1}{\pi R^2}, \tag{34}$$

so that

$$\rho_G = \eta / b, \quad \eta = \kappa, \tag{35}$$

similar to the bending case. The non-zero strain gradient components are

$$\eta_{231} = \eta_{321} = -\kappa, \quad \eta_{132} = \eta_{312} = \kappa, \tag{36}$$

with quadratic invariants

$$\eta_{iik}\eta_{jjk} = 0, \quad \eta_{ijk}\eta_{ijk} = 4\kappa^2, \quad \eta_{ijk}\eta_{kji} = -2\kappa^2. \tag{37}$$

Combining (23), (35) and (37), the dislocation model for pure torsion suggests

$$c_2 - \tfrac{1}{2}c_3 = \tfrac{1}{4}. \tag{38}$$

Now consider the growth of a 2-D axisymmetric void, in which case the most efficient dislocation configuration is shown in Fig. 3(c). The total number of geometrically necessary dislocations in a ring of radius r and width δr is

$$\delta n_d = \rho_G(2\pi r \delta r). \tag{39}$$

The variation of hoop strain $\varepsilon_{\theta\theta}$ across such a ring from radius r to $r+\delta r$ is

$$\delta\varepsilon_{\theta\theta} = -\frac{\delta n_d b}{2\pi r} = -\rho_G b \delta r. \tag{40}$$

Letting $\delta r \to 0$ yields

$$\rho_G = \frac{\eta}{b}, \quad \eta = -\frac{\partial \varepsilon_{\theta\theta}}{\partial r}. \tag{41}$$

The displacement field associated with 2-D axisymmetric void growth consistent with the MSG theory is (Huang et al., 1998),

$$u_r = u_0 \frac{r_0}{r} \tag{42}$$

where r_0 is the initial size of the void and u_0 is the displacement of the void boundary. This displacement leads to

$$\varepsilon_{\theta\theta} = u_0 \frac{r_0}{r^2}. \tag{43}$$

The non-zero components of strain gradient are

$$\eta_{rrr} = 2u_0 \frac{r_0}{r^3}, \quad \eta_{r\theta\theta} = \eta_{\theta r\theta} = \eta_{\theta\theta r} = -2u_0 \frac{r_0}{r^3}, \tag{44}$$

with quadratic invariants

$$\eta_{iik}\eta_{jjk} = 0, \quad \eta_{ijk}\eta_{ijk} = 16u_0^2\frac{r_0^2}{r^6}, \quad \eta_{ijk}\eta_{kji} = 16u_0^2\frac{r_0^2}{r^6}. \tag{45}$$

Combining (23), (41) and (45), the dislocation model for 2-D axisymmetric void growth suggests

$$c_2 + c_3 = \tfrac{1}{4}. \tag{46}$$

Solving (30), (38), (46) gives

$$c_1 = 0, \quad c_2 = \tfrac{1}{4}, \quad c_3 = 0. \tag{47}$$

Therefore, the form of the effective strain gradient consistent with the notion of most efficient dislocation configuration is

$$\eta = \sqrt{\tfrac{1}{4}\eta_{ijk}\eta_{ijk}}, \tag{48}$$

which is similar to that of the effective strain in conventional plasticity

$$\varepsilon = \sqrt{\tfrac{2}{3}\varepsilon_{ij}\varepsilon_{ij}}. \tag{49}$$

As an independent check, we apply (48) to the growth of 3-D spherical void. It is difficult to construct a dislocation model consistent with spherically symmetric deformation. We will be content with the approximate model shown in Fig. 3(d) where a small block of deformed material near the growing void is assumed to contain a cross grid of edge dislocations. The total number of dislocations in this block is

$$\rho_G a^2 \delta R = 2a\delta n_d, \tag{50}$$

where a is the side length of the block and δn_d denotes the number of geometrically necessary dislocations along a tangential direction; the factor of two arises due to the cross grid assumption. The variation in strain $\varepsilon_{\theta\theta}$ across the width of the block from radius R to $R+\delta R$ is

$$\delta\varepsilon_{\theta\theta} = -\frac{\delta n_d b}{a} = -\frac{1}{2}\rho_G b\delta R. \tag{51}$$

Letting $\delta R \to 0$, we obtain

$$\rho_G = \frac{\eta}{b}, \quad \eta = -2\frac{\partial\varepsilon_{\theta\theta}}{\partial R}. \tag{52}$$

The displacement field associated with a 3-D spherical void consistent with the MSG theory is (Huang et al., 1998),

$$u_R = u_0\frac{R_0^2}{R^2} \tag{53}$$

where R_0 is the initial size of the void and u_0 is the displacement of the void boundary. This displacement leads to

$$\varepsilon_{\theta\theta} = u_0 \frac{R_0^2}{R^3}. \tag{54}$$

The non-zero strain gradient components are

$$\eta_{\mathrm{RRR}} = 6u_0 \frac{R_0^2}{R^4}, \quad \eta_{\mathrm{R}\theta\theta} = \eta_{\theta\mathrm{R}\theta} = \eta_{\mathrm{R}\phi\phi} = \eta_{\phi\mathrm{R}\phi} = \eta_{\theta\theta\mathrm{R}} = \eta_{\phi\phi\mathrm{R}} = -3u_0 \frac{R_0^2}{R^4}. \tag{55}$$

The effective strain gradient calculated from the deformation field according to (48) is

$$\eta = \frac{1}{2}\sqrt{\eta_{ijk}\eta_{ijk}} = 3\sqrt{\frac{5}{2}}u_0 \frac{R_0^2}{R^4}, \tag{56}$$

which is compared to the corresponding quantity calculated from the cross-grid dislocation model

$$\eta_d = -2\frac{\partial \varepsilon_{\theta\theta}}{\partial R} = 6u_0 \frac{R_0^2}{R^4}. \tag{57}$$

The relative difference between the two results is

$$\frac{\eta_d - \eta}{\eta_d} = 0.21. \tag{58}$$

The 21% difference can be partly attributed to the use of a cross-grid of edge dislocations to approximate spherically symmetric deformation and partly to the difficulty in establishing the most efficient dislocation configuration in the present case. Nevertheless, we believe this comparison is encouraging since the effective strain gradient calculated from (48) is expected to be slightly smaller than η_d estimated from a non-optimal dislocation configuration. We are presently unaware of any other cases where appropriate dislocation models can be constructed for further comparison.

3.2. *Comparison with the Fleck–Hutchinson results*

Smyshlyaev and Fleck (1996) and Fleck and Hutchinson (1997) have shown that the incompressible strain gradient tensor η_{ijk} can be decomposed as

$$\boldsymbol{\eta} = \boldsymbol{\eta}^{(1)} + \boldsymbol{\eta}^{(2)} + \boldsymbol{\eta}^{(3)}, \tag{59}$$

where

$$\eta_{ijk}^{(1)} = \eta_{ijk}^{\mathrm{s}} - \tfrac{1}{5}[\delta_{ij}\eta_{kpp}^{\mathrm{s}} + \delta_{jk}\eta_{ipp}^{\mathrm{s}} + \delta_{ki}\eta_{jpp}^{\mathrm{s}}], \tag{60}$$

$$\eta_{ijk}^{(2)} = \tfrac{1}{6}[e_{ikp}e_{jlm}\eta_{lpm} + e_{jkp}e_{ilm}\eta_{lpm} + 2\eta_{ijk} - \eta_{jki} - \eta_{kij}], \tag{61}$$

$$\eta_{ijk}^{(3)} = \tfrac{1}{6}[-e_{ikp}e_{jlm}\eta_{lpm} - e_{jkp}e_{ilm}\eta_{lpm} + 2\eta_{ijk} - \eta_{jki} - \eta_{kij}] + \tfrac{1}{5}[\delta_{ij}\eta_{kpp}^{\mathrm{s}} + \delta_{jk}\eta_{ipp}^{\mathrm{s}} + \delta_{ki}\eta_{jpp}^{\mathrm{s}}]. \tag{62}$$

In the above equations, e_{ikp} is the permutation tensor and $\boldsymbol{\eta}^{\mathrm{s}}$ denotes the fully symmetric part of $\boldsymbol{\eta}$,

$$\eta^{\mathrm{s}}_{ijk} = \tfrac{1}{3}[\eta_{ijk} + \eta_{jki} + \eta_{kij}]. \tag{63}$$

Fleck and Hutchinson (1997) showed that the effective strain gradient defined in (23) can be rewritten as

$$l\eta = \sqrt{l_1^2 \eta^{(1)}_{ijk}\eta^{(1)}_{ijk} + l_2^2 \eta^{(2)}_{ijk}\eta^{(2)}_{ijk} + l_3^2 \eta^{(3)}_{ijk}\eta^{(3)}_{ijk}} \tag{64}$$

where the material length l is introduced so that the coefficients l_1^2, l_2^2 and l_3^2 can be interpreted as three constitutive lengths which are related to constants c_1, c_2 and c_3 by

$$\frac{l_1^2}{l^2} = c_2 + c_3, \quad \frac{l_2^2}{l^2} = c_2 - \frac{1}{2}c_3, \quad \frac{l_3^2}{l^2} = \frac{5}{2}c_1 + c_2 - \frac{1}{3}c_3. \tag{65}$$

Fleck and Hutchinson (1997) intended to determine these three lengths from experiments. In our approach, the choice of (47) implies that

$$l_1 = l_2 = l_3 = \tfrac{1}{2}l. \tag{66}$$

That is, the dislocation models we use suggest that the three constitutive lengths introduced by Fleck and Hutchinson (1997) are identical. This choice does not deviate much from a suggestion of Begley and Hutchinson (1998),

$$l_1 = \tfrac{1}{8}l, \quad l_2 = \tfrac{1}{2}l, \quad l_3 = \sqrt{\tfrac{5}{24}}\,l, \tag{67}$$

which are obtained by fitting the experimental data from bending of ultra-thin beams (Stolken and Evans, 1998), torsion of thin wires (Fleck et al., 1994), and micro-indentation (Stelmashenko et al., 1993; Atkinson, 1995; Ma and Clarke, 1995; Nix, 1997).

At this point, we also mention in passing that the phenomenological couple stress theory of strain gradient plasticity by Fleck and Hutchinson (1993) corresponds to the following choice

$$l_1 = 0, \quad l_2 = \tfrac{1}{2}l, \quad l_3 = \sqrt{\tfrac{5}{24}}\,l, \tag{68}$$

where the effective strain gradient has the form

$$\eta = \sqrt{\tfrac{2}{3}\chi_{ij}\chi_{ij}}. \tag{69}$$

Here, χ_{ij} is the curvature tensor defined as the gradient of material rotation. The relationship between χ_{ij} and η_{ijk} is (Fleck and Hutchinson, 1997)

$$\chi_{ij}\chi_{ij} = \tfrac{3}{8}\eta^{(2)}_{ijk}\eta^{(2)}_{ijk} + \tfrac{5}{16}\eta^{(3)}_{ijk}\eta^{(3)}_{ijk}, \tag{70}$$

$$\chi_{ij}\chi_{ji} = \tfrac{3}{8}\eta^{(2)}_{ijk}\eta^{(2)}_{ijk} - \tfrac{5}{16}\eta^{(3)}_{ijk}\eta^{(3)}_{ijk}. \tag{71}$$

It is thus clear that $\eta^{(2)}_{ijk}$ and $\eta^{(3)}_{ijk}$ are associated with the rotational gradients, while $\eta^{(1)}_{ijk}$ is associated with the stretch gradient (Fleck and Hutchinson, 1997). The three dislocation models we use to derive the effective strain gradient have different fractions of rotation and stretch gradients. The torsion deformation of (32) is a field of pure rotation gradient, the 2-D axisymmetric void deformation of (42) is a field of pure stretch gradient, and the bending deformation of (27) is a field of mixed rotation and stretch gradients.

4. Constitutive equations

Based on the theoretical postulates discussed in Section 2 and dislocation models in Section 3, a new theoretical framework of strain gradient plasticity is proposed below. This theory will be referred to as a mechanism-based theory of strain gradient plasticity (MSG). Attention is focused on the deformation theory of MSG. The flow theory of MSG will be presented in a forthcoming paper.

Consider a unit cell on the mesoscale with the length of all edges equal to l_ε (Fig. 2). The mesoscale cell size l_ε is much smaller than the intrinsic material length l in (7) associated with strain gradient plasticity. Within this cell, the displacement field is assumed to vary as

$$\tilde{u}_k = \varepsilon_{ik}x_i + \tfrac{1}{2}\eta_{ijk}x_ix_j + o(x^3) \tag{72}$$

where x_i denotes the local coordinates with origin at the center of the cell and η_{ijk} is second gradient of the displacement field. When the cell is sufficiently small, higher order displacement gradients can be ignored and the strain field varies linearly as

$$\tilde{\varepsilon}_{ij} = \varepsilon_{ij} + \tfrac{1}{2}(\eta_{kij} + \eta_{kji})x_k. \tag{73}$$

The microscale strain $\tilde{\varepsilon}_{ij}$ is thus related to the mesoscale strain ε_{ij} and strain gradient η_{kij}.

Following classical plasticity theory (Hill, 1950), we define the microscale effective strain as

$$\tilde{\varepsilon}^2 = \tfrac{2}{3}\tilde{\varepsilon}_{ij}\tilde{\varepsilon}_{ij}, \tag{74}$$

$$\tilde{\varepsilon}\delta\tilde{\varepsilon} = \tfrac{2}{3}\tilde{\varepsilon}_{ij}\delta\tilde{\varepsilon}_{ij}. \tag{75}$$

The incremental plastic work can be expressed as $\tilde{\sigma}'_{ij}\delta\tilde{\varepsilon}_{ij}$. Inserting (16) and (75) into the above expression yields

$$\tilde{\sigma}'_{ij} = \frac{2\tilde{\varepsilon}_{ij}}{3\tilde{\varepsilon}}\tilde{\sigma}_e, \quad \tilde{\sigma}_e = \tilde{\sigma} = \sigma_{\mathrm{Y}}\sqrt{f^2(\tilde{\varepsilon}) + l\eta}. \tag{76}$$

We note that the effective stress $\tilde{\sigma}_e$ is always equated to the flow stress $\tilde{\sigma}$ for the deformation theory of MSG.

Substituting the microscale constitutive law (76) into the plastic work equality (17) gives

$$\int_{V_{\rm cell}} \tilde{\sigma}'_{ij}\delta\tilde{\varepsilon}_{ij}\,\mathrm{d}V = (\sigma'_{ij}\delta\varepsilon_{ij} + \tau'_{ijk}\delta\eta_{ijk})V_{\rm cell}. \tag{77}$$

Inserting the kinematic assumption

$$\delta\tilde{\varepsilon}_{ij} = \delta\varepsilon_{ij} + \tfrac{1}{2}(\delta\eta_{kij} + \delta\eta_{kji})x_k \tag{78}$$

into (77) and equating the corresponding coefficients of $\delta\varepsilon_{ij}$ and $\delta\eta_{kij}$ leads to the constitutive equations for the deformation theory of MSG. The coefficient of $\delta\varepsilon_{ij}$ gives

$$\sigma'_{ij} = \frac{1}{V_{\rm cell}}\int_{V_{\rm cell}} \tilde{\sigma}'_{ij}\,\mathrm{d}V, \tag{79}$$

and the coefficient of $\delta\eta_{ijk}$ gives

$$\tau'_{ijk} = \frac{1}{V_{\rm cell}}\,\mathrm{Dev}\left[\frac{1}{2}\int_{V_{\rm cell}} \tilde{\sigma}'_{jk}x_i + \tilde{\sigma}'_{ik}x_j\,\mathrm{d}V\right], \tag{80}$$

where Dev [. . .] denotes the deviatoric part of [. . .].

Since the coordinate origin of x_k is located at the center of the cubic cell, the integral in the above equation can be carried out according to the rules

$$\frac{1}{V_{\rm cell}}\int_{V_{\rm cell}} \mathrm{d}V = 1, \quad \int_{V_{\rm cell}} x_k\,\mathrm{d}V = 0, \quad \frac{1}{V_{\rm cell}}\int_{V_{\rm cell}} x_k x_m\,\mathrm{d}V = \frac{1}{12}l_\varepsilon^2\delta_{km}. \tag{81}$$

Keeping terms to the lowest order in l_ε, we obtain the mesoscale constitutive equations

$$\sigma'_{ij} = \frac{2\varepsilon_{ij}}{3\varepsilon}\sigma, \tag{82}$$

$$\tau'_{ijk} = l_\varepsilon^2\left[\frac{\sigma}{\varepsilon}(\Lambda_{ijk} - \Pi_{ijk}) + \frac{\sigma_{\rm Y}^2 f(\varepsilon)f'(\varepsilon)}{\sigma}\Pi_{ijk}\right], \tag{83}$$

where

$$\sigma = \sigma_{\rm Y}\sqrt{f^2(\varepsilon) + l\eta}, \tag{84}$$

$$\Lambda_{ijk} = \tfrac{1}{72}[2\eta_{ijk} + \eta_{kji} + \eta_{kij} - \tfrac{1}{4}(\delta_{ik}\eta_{ppj} + \delta_{jk}\eta_{ppi})], \tag{85}$$

$$\Pi_{ijk} = \frac{\varepsilon_{mn}}{54\varepsilon^2}\left[\varepsilon_{ik}\eta_{jmn} + \varepsilon_{jk}\eta_{imn} - \frac{1}{4}(\delta_{ik}\varepsilon_{jp} + \delta_{jk}\varepsilon_{ip})\eta_{pmn}\right]. \tag{86}$$

Note that terms of order $o(l_\varepsilon^2)$ in σ'_{ij} and those of order $o(l_\varepsilon^4)$ in τ'_{ijk} have been ignored.

An important question is whether we could write the MSG constitutive law in the form

$$\sigma'_{ij} = \frac{\partial W}{\partial\varepsilon_{ij}}, \quad \tau'_{ijk} = \frac{\partial W}{\partial\eta_{ijk}}, \tag{87}$$

where $W = W(\boldsymbol{\varepsilon}, \boldsymbol{\eta})$ is a strain energy density function. This is clearly impossible because the MSG constitutive equations do not satisfy the reciprocity relation, i.e.

$$\frac{\partial \sigma'_{ij}}{\partial \eta_{kmn}} \neq \frac{\partial \tau'_{kmn}}{\partial \varepsilon_{ij}}, \tag{88}$$

required for the existence of a strain energy function. This has important implications for the application of a J-integral approach to the MSG theory, which we leave to future work.

The flow theory of MSG can be established based on the same guiding principles outlined in this paper. Although we do not discuss it in detail here, the MSG flow theory has been worked out and is currently being subjected to analysis (Huang et al., work in progress). The starting point for the MSG flow theory is the microscale normality relation

$$\frac{\mathrm{d}\tilde{\varepsilon}_{ij}}{\mathrm{d}\tilde{\varepsilon}} = \frac{3\tilde{\sigma}'_{ij}}{2\tilde{\sigma}_e}. \tag{89}$$

Under proportional, monotonically increasing deformation,

$$\frac{\mathrm{d}\tilde{\varepsilon}_{ij}}{\mathrm{d}\tilde{\varepsilon}} = \frac{\tilde{\varepsilon}_{ij}}{\tilde{\varepsilon}}, \quad \tilde{\sigma}_e = \tilde{\sigma}, \tag{90}$$

the microscale flow equation (89) reduces to the deformation equation (76) and the flow theory becomes coincident with the deformation theory. Further discussions of the flow theory of MSG are deferred to a later paper.

5. Discussion

The MSG constitutive equations will be analyzed in Part II (Huang et al., 1999) of this paper for several sample applications, and will be further studied in Part III (Huang et al., work in progress) with a detailed investigation of micro-indentation. Here we give some discussion on a few outstanding issues of the new theory.

5.1. *Taylor hardening model*

Some discussion of the Taylor hardening model might be helpful to gain a deeper understanding of the connection between MSG and the dislocation interaction processes. Dislocation theory indicates that the Peach–Koehler force due to interaction of a pair of dislocations is proportional to

$$\sigma_{\text{pair}} \sim \frac{\mu b}{2\pi(1-\nu)L} \tag{91}$$

where L is the distance between the dislocations. This stress sets a critical value for the applied stress to break or untangle the interactive pair so that slip can occur even if one of the dislocations is pinned by an obstacle. In the Taylor model, this picture is generalized to the interaction of a group of statistically stored dislocations which trap each other in a random way. If the mean dislocation spacing is L, the critical stress required to untangle the interactive dislocations and to induce significant plastic deformation is defined as the Taylor flow stress

$$\sigma = \frac{\alpha\mu b}{L} = \alpha\mu b\sqrt{\rho}, \tag{92}$$

where $\rho = 1/L^2$ is the dislocation density. Alternatively, the Taylor flow stress can also be viewed as the 'passing stress' for a moving dislocation to glide through a forest of tangled dislocations without being pinned. The similarity between the Taylor model and the interaction of a pair of dislocations indicates the potential of using (92) as a fundamental measure of dislocation interaction at length scales close to those of discrete dislocations.

5.2. *Strain gradient length scale l*

The material length l corresponds to the scale at which the effects of strain gradient become comparable to those of strain. In the presence of a strong strain gradient, the total dislocation density ρ is considered to be the sum of statistically stored dislocations $\rho_s = 1/L_S^2$ and geometrically necessary dislocations $\rho_G = \eta/b$. The strain gradient effects become significant when ρ_S and ρ_G are of the same order of magnitude. Equating

$$\rho_S = \rho_G \tag{93}$$

immediately suggests that

$$\eta^{-1} = L_S^2/b = \hat{l} \tag{94}$$

is a fundamental length scale signifying the strain gradient effects. Nix and Gao (1998) have shown that L_S^2/b is a fundamental length scale which is related to the material length l and the uniaxial stress–strain behavior $\sigma_0 = \sigma_Y f(\varepsilon)$ by the following relations

$$\hat{l} = \frac{L_S^2}{b}, \quad L_S \simeq \frac{\mu b}{\sigma_0}, \quad l \simeq \hat{l}_{\text{yield}}, \tag{95}$$

where $\hat{l}_{\text{yield}}$ is the value of $\hat{l}$ at yielding. Therefore, the material length l is a fundamental length scale related to Burgers vector and dislocation spacing at yielding, and is a fundamental measure of the deformation length at which geometrically necessary dislocations constitute a significant fraction of the total dislocation population.

5.3. *Mesoscale cell size l_ε*

The theory of strain gradient plasticity is intended for applications with deformation length scales in the range of 0.1–10 micron (Fleck and Hutchinson, 1997). Since plasticity theories describe the collective behavior of a large number of dislocations, the range of validity of MSG should be larger than the mean dislocation spacing. A dislocation spacing of 0.1 micron corresponds to a dislocation density on the order of $10^{10}\,\mathrm{cm}^{-2}$, which is typical for a deformed crystal. The cell size l_ε in the MSG theory is a resolution parameter which controls the accuracy with which the strain gradient is calculated in each cell. This parameter needs to be sufficiently small to ensure accuracy of the strain gradient. However, since the Taylor hardening model is assumed to govern, this cell also needs to be large enough to contain a sufficient number of dislocations to ensure the accuracy of the flow stress. In other words, there is a fundamental inconsistency in trying to capture both the strain gradient and the flow stress accurately. A compromise between these two conflicting requirements dictates a suitable choice of l_ε. We propose to define

$$l_\varepsilon = \beta L_{\mathrm{yield}} = \beta \frac{\mu b}{\sigma_{\mathrm{Y}}}, \tag{96}$$

where L_{yield} is the mean spacing between statistically stored dislocations at yielding and β is a constant coefficient to be determined from experiments. Dislocation spacing provides a fundamental measure of the 'discreteness' limit for continuum plasticity, similar to atomic spacing as a measure of the discreteness limit for continuum elasticity. For stress levels below the yield point, plastic deformation is negligible and elasticity theory is usually used to analyze the deformation. For stresses above the yield point, the mean dislocation spacing becomes smaller than L_{yield} so that l_ε is always larger than the dislocation spacing for the intended application range. The condition $\beta > 1$ ensures that there are multiple dislocations within the mesoscale cell.

For typical structural metals, we may take the following numbers as representative estimates of the relevant length scales:

$$\mu/\sigma_{\mathrm{Y}} = 100, \quad b = 0.1\,\mathrm{nm}, \quad L_{\mathrm{yield}} = 10\,\mathrm{nm}, \tag{97}$$

$$l_\varepsilon \sim 10\text{–}100\,\mathrm{nm}, \quad l \sim 1\text{–}10\,\mu\mathrm{m}$$

where the cell size parameter β is taken to be from 1–10.

With the exception of single crystal materials, there is usually more than one microstructural length scale in an engineering material. Examples of such length scales include grain size, particle spacing, layer thickness, etc. Dislocations are usually not uniformly distributed. These complications make it difficult to calculate the strain gradient length l directly from microstructure. An important advantage of MSG is that the material length $l = 3\alpha^2(\mu/\sigma_{\mathrm{Y}})^2 b$ is related to macroscopically measurable quantities such as the yield stress σ_{Y}, which account for, in an average sense, the effects of various microstructural features. For example, the Hall–Petch relation indicates that the uniaxial flow stress increases with the reduction in grain size,

$$\sigma_{\mathrm{Y}} = c_0 + c_1 d^{-1/2}, \tag{98}$$

where d is the mean grain diameter. The grain size thus affects the strain gradient plasticity indirectly through its influence on σ_{Y} and on the constitutive length l.

5.4. *A 'quantum mechanical' analogy of MSG*

There may be significant difficulties in understanding the role of the cell size parameter l_ε since a similar concept does not exist in the conventional theories of elasticity and plasticity. We believe that this parameter is not an ad hoc parameter introduced for convenience, but rather a fundamental parameter associated with the very nature of plastic deformation. It may be particularly helpful to draw a loose analogy between MSG and a few concepts in quantum mechanics (e.g., Saxon, 1968) for physics oriented readers. First of all, plastic deformation 'quantized' by dislocation activities may be compared to lattice vibration quantized by phonons, or to electromagnetic radiation quantized by photons. In quantum mechanics, complementary variables such as position-momentum and energy-time obey the Heisenberg uncertainty principle in that both variables in a pair cannot be simultaneously determined to arbitrary accuracy. This uncertainty is dictated by Planck's constant h which measures the strength of quantization. The cell size l_ε of MSG is a consequence of a similar uncertainty associated with plastic deformation. This uncertainty is evident for the complementarity between flow stress and strain gradient, in that accurate determination of flow stress requires large cell size l_ε with many interactive dislocations, while accurate determination of strain gradient requires small cell size l_ε so that the linear expansion of the strain field within the cell is valid. The analogy can be further extended to the mathematical structure of MSG, in the sense that l_ε scales the highest derivative of the deformation field (see further discussions in Part II), similar to the role of Planck's constant in scaling the highest derivative of the Schroedinger equation.

5.5. *Prospects of strain gradient plasticity*

The MSG theory represents an effort to bridge the length scales between conventional mechanics theories which are intended for applications beyond 10 micron and quantum-atomistic simulations which are still restricted to submicron size scales and picosecond time scales. There is still a long way to go before the quantum-atomistic simulations can be routinely applied to engineering applications. In the meantime, efforts must be made to improve the continuum models of deformation and failure so that micro- and nano-structures can be manufactured and designed on a rational basis. These efforts must involve interdisciplinary research activities which combine continuum mechanics, materials science and physics.

The theory of MSG is a true marriage between continuum mechanics and materials science. This marriage is profoundly manifested by the fundamental length scales l and l_ε which are combinations of the elasticity constant μ, the plasticity constant σ_Y and the Burgers vector b. These fundamental lengths can be obtained merely by a dimensional analysis combining elasticity, plasticity and the atomic nature of solids:

$$l_n = \left(\frac{\mu}{\sigma_Y}\right)^n b \tag{99}$$

where n is an arbitrary integer. If we restrict our attention to lengths which are in the size range from interatomic spacing to 10 microns, we are necessarily led to the fundamental lengths $l_\varepsilon \sim l_1$ and $l \sim l_2$. The higher order length scales, $l_n\,(n > 2)$, are above the millimeter range where conventional mechanics theories have proven to be successful.

Historically, the theories of elasticity and plasticity in continuum mechanics were developed largely based on phenomenological behaviors of materials on the macroscopic scale, and quite independent of the advances in materials science and physics on the understanding of the microscopic processes which are responsible for the deformation. The theory of elasticity has been successfully applied to studies of crystal defects, which has led to the theory of dislocations. However, apart from notable exceptions, this interaction between mechanics and materials science has primarily been a one-way process in the past. Despite the fact that the theory of dislocations has provided a fundamental understanding of plastic deformation, much of the conventional theory of plasticity had already been in place when dislocations were discovered in the 1930s. The development of mechanism-based strain gradient plasticity highlights a new era in which advances in modern materials science and physics begin to play an essential part in the founding principles of continuum mechanics.

6. Summary

In this paper, the following objectives have been achieved:

(1) We have proposed a systematic method of formulating mechanism-based strain gradient plasticity for mesoscale applications with deformation lengths in the range of 0.1 microns and above. This method consists of three basic postulates. First, it is assumed that the microscale flow stress is governed by the Taylor hardening model. Second, the microscale and mesoscale variables are linked by a virtual work statement equating plastic work on the two scales. Third, the essential structure of conventional plasticity such as normality is retained at the microscale.

(2) Dislocation models are developed to relate the microscale notion of geometrically necessary dislocations to the mesoscale notion of strain gradients. In particular, we show that the dislocation models indicate that the effective strain gradient η, which measures the density of geometrically necessary dislocations, is related to the strain gradient tensor $\eta_{ijk} = u_{k,ij}$ by the simple expression $\eta^2 = (1/4)\eta_{ijk}\eta_{ijk}$.

(3) It is shown that the three basic postulates of MSG allow us to construct higher order continuum theories of strain gradient plasticity. For conciseness, only the deformation theory of MSG is considered in this paper, without accounting for elasticity and compressibility of materials.

(4) Two length scales are identified in MSG. The first is a material length l which governs the strain gradient effects and measures the interaction between statistically stored and geometrically necessary dislocations. This length is related to elastic modulus, plastic yield stress and Burgers vector by $l \sim (\mu/\sigma_Y)^2 b$ (Nix and Gao, 1998). The second length is a resolution parameter $l_\varepsilon \sim \mu b/\sigma_Y$ corresponding to the size of the mesoscale cell in the theory formulation. The significance of l_ε can be explained from a fundamental uncertainty relationship in the simultaneous

determination of flow stress and strain gradient. Accurate determination of flow stress requires large cell size l_ε with many interactive dislocations within the cell while accurate determination of strain gradient requires small cell size l_ε so that the linear expansion of the strain field within the cell is valid. A loose analogy with quantum mechanics is used to illustrate the significance of the uncertainty concept for continuum plasticity and the cell size l_ε in MSG.

Acknowledgments

The work of H.G. was supported by the NSF through Young Investigator Award MSS-9358093. The work of Y.H. was supported by the NSF through Grants INT-9423964 and CMS-96-10491, and by the NSF of China. The work of W.D.N. was supported by the DOE through Grant DE-FG03-89ER45387. The work of J.W.H. was supported in part by the ONR through Grant N00014-96-10059 and by the NSF through Grant CMS-96-34632. H.G. and Y.H. acknowledge helpful discussions with Keh-Chih Hwang.

References

Acharya, A., Bassani, J.L., 1995. On non-local flow theories that preserve the classical structure of incremental boundary value problems. In: Micromechanics of Plasticity and Damage of Multiphase Materials. IUTAM Symposium, Paris, August 29–September 1, 1995.

Acharya, A., Shawki, T.G., 1995. Thermodynamic restrictions on constitutive equations for second-deformation-gradient inelastic behavior. J. Mech. Phys. Solids 43, 1751–1772.

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

Ashby, M.F., 1970. The deformation of plastically non-homogeneous alloys. Phil. Mag. 21, 399–424.

Atkinson, M., 1995. Further analysis of the size effect in indentation hardness tests of some metals. J. Mater. Res. 10, 2908–2915.

Begley, M.R., Hutchinson, J.W., 1998. The mechanics of size-dependent indentation. J. Mech. Phys. Solids, in press.

Cottrell, A.H., 1964. In: The Mechanical Properties of Matter. Wiley, New York, p. 277.

Dai, H., Parks, D.M., 1997. Geometrically-necessary dislocation density in continuum crystal plasticity theory and FEM implementation, unpublished manuscript.

De Guzman, M.S., Newbauer, G., Flinn, P., Nix, W.D., 1993. The role of indentation depth on the measured hardness of materials. Materials Research Symposium Proceedings 308, 613–618.

Elssner, G., Korn, D., Ruehle, M., 1994. The influence of interface impurities on fracture energy of UHV diffusion bonded metal–ceramic bicrystals. Scripta Metall. Mater. 31, 1037–1042.

Eshelby, J.D., 1953. Screw dislocations in thin rods. J. Appl. Phys. 24, 176.

Fleck, N.A., Hutchinson, J.W., 1993. A phenomenological theory for strain gradient effects in plasticity. J. Mech. Phys. Solids 41, 1825–1857.

Fleck, N.A., Hutchinson, J.W., 1997. Strain gradient plasticity. In: Hutchinson, J.W., Wu, T.Y. (Eds.), Advances in Applied Mechanics, vol. 33. Academic Press, New York, pp. 295–361.

Fleck, N.A., Muller, G.M., Ashby, M.F., Hutchinson, J.W., 1994. Strain gradient plasticity: theory and experiment. Acta Metallurgica et Materialia 42, 475–487.

Gurtin, M.E., 1965a. Thermodynamics and the possibility of spatial interaction in rigid heat conductors. Arch. Ration. Mech. Anal. 18, 335–342.

Gurtin, M.E., 1965b. Thermodynamics and the possibility of spatial interaction in elastic materials. Arch. Ration. Mech. Anal. 19, 339–352.

Hill, R., 1950. Mathematical Theory of Plasticity. Oxford University Press, Oxford, U.K.

Hirth, J.P., Lothe, J., 1982. Theory of Dislocations, 2nd ed. John Wiley and Sons, New York, 1982, p. 62.

Huang, Y., Zhang, L., Guo, T.F., Hwang, K.-C., 1995. Near-tip fields for cracks in materials with strain-gradient effects. In: Willis, J.R. (Ed.), Proceedings of IUTAM Symposium on Nonlinear Analysis of Fracture. Kluwer Academic Publishers, Cambridge, U.K., pp. 231–242.

Huang, Y., Zhang, L., Guo, T.F., Hwang, K.-C., 1997. Mixed mode near-tip fields for cracks in materials with strain-gradient effects. J. Mech. Phys. Solids 45, 439–465.

Huang, Y., Gao, H., Nix, W.D., Hutchinson, J.W., 1999. Mechanism-based strain gradient plasticity—II. Analysis, in preparation.

Hutchinson, J.W., 1997. Linking scales in mechanics. In: Karihaloo, B.L., Mai, Y.-W., Ripley, M.I., Ritchie, R.O. (Eds.), Advances in Fracture Research, Pergamon Press, New York, pp. 1–14.

Koiter, W.T., 1964. Couple stresses in the theory of elasticity, I and II. Proc. K. Ned. Akad. Wet. (B) 67, 17–44.

Kroener, E., 1962. Dislocations and continuum mechanics. Appl. Mech. Rev. 15, 599–606.

Lloyd, D.J., 1994. Particle reinforced aluminum and magnesium matrix composites. Int. Mater. Rev. 39, 1–23.

Ma, Q., Clarke, D.R., 1995. Size dependent hardness in silver single crystals. J. Materials Research 10, 853–863.

McElhaney, K.W., Vlassak, J.J., Nix, W.D., 1998. Determination of indenter tip geometry and indentation contact area for depth-sensing indentation experiments, J. Materials Research 13, 1300–1306.

Mindlin, R.D., 1964. Micro-structure in linear elasticity. Arch. Ration. Mech. Anal. 16, 51–78.

Mindlin, R.D., 1965. Second gradient of strain and surface tension in linear elasticity. Int. J. Solids Struct. 1, 417–438.

Muhlhaus, H.B., Aifantis, E.C., 1991. A variational principle for gradient plasticity. Int. J. Solids Struct. 28, 845–857.

Nix, W.D., 1989. Mechanical properties of thin films. Metall. Trans. 20A, 2217–2245.

Nix, W.D., 1997. Elastic and plastic properties of thin films on substrates: nano-indentation techniques. Mat. Sci. and Engr. A234–236, 37–44.

Nix, W.D., Gao, H., 1998. Indentation size effects in crystalline materials: a law for strain gradient plasticity. J. Mech. Phys. Solids 46, 411–425.

Nye, J.F., 1953. Some geometrical relations in dislocated crystal. Acta Metall. 1, 153–162.

Poole, W.J., Ashby, M.F., Fleck, N.A., 1996. Microhardness of annealed and work-hardened copper polycrystals. Scripta Metall. Mater. 34, 559–564.

Rice, J.R., 1970. On the structure of stress–strain relations for time-dependent plastic deformation in metals. J. Appl. Mech. 34, 728–737.

Rice, J.R., 1971. Inelastic constitutive relations for solids: an internal-variable theory and its application to metal plasticity. J. Mech. Phys. Solids 19, 433–455.

Saxon, D.S., 1968. Elementary quantum mechanics. Holden-Day, San Francisco.

Shu, J.Y., Fleck, N.A., 1998. The prediction of a size effect in micro indentation. Int. J. Solids Struct. 35, 1363–1383.

Smyshlyaev, V.P., Fleck, N.A., 1996. The role of strain gradients in the grain size effect for polycrystals. J. Mech. Phys. Solids 44, 465–495.

Stelmashenko, N.A., Walls, M.G., Brown, L.M., Milman, Y.V., 1993. Microindentation on W and Mo oriented single crystals: an STM study. Acta Metallurgica et Materialia 41, 2855–2865.

Stolken, J.S., Evans, A.G., 1998. A microbend test method for measuring the plasticity length scale, submitted for publication.

Toupin, R.A., 1962. Elastic materials with couple stresses. Arch. Ration. Mech. Anal. 11, 385–414.

Xia, Z.C., Hutchinson, J.W., 1996. Crack tip fields in strain gradient plasticity. J. Mech. Phys. Solids 44, 1621–1648.

Zbib, H., Aifantis, E.C., 1989. On the localization and postlocalization behavior of plastic deformation. Part I. On the initiation of shear bands; Part II. On the evolution and thickness of shear bands; Part III. On the structure and velocity of Portevin–Le Chatelier bands. Res. Mech. 261–277, 279–292 and 293–305.