A Critical Assessment of Theories of Strain Gradient Plasticity

(Article begins on next page)

A Critical Assessment of Theories of Strain Gradient Plasticity

A.G. Evans[1] and J.W. Hutchinson[2]

[1]Departments of Materials and of Mechanical Engineering, University of California, Santa Barbara, CA 93106

[2] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Abstract

Theories to extend plasticity to the micron scale have been in existence for over a decade, complemented by a growing body of experimental data. Here, materials and mechanics aspects of two prominent strain gradient theories of plasticity, due Nix and Gao and to Fleck and Hutchinson, have been assessed within the context of simple bending. Differences between the theories have been highlighted. The theories predict different trends relative to the size-dependence of initial yielding and rate of hardening. The dislocation mechanics underpinning the two theories is addressed. Distinctions between lower order theories and higher order theories are also drawn emphasizing the flexibility of higher order theories to solve problems for a wide range of boundary conditions, especially those where, locally, the dislocations are blocked (pile-up) and the plastic strain is zero.

1. Introduction

The elevation of the plastic flow stress in small volumes is well-documented [1-11]. Effects of configuration size have been presented for indentation [1-6], torsion [7], bending and thin film extension [8,9]. Micro-structural size effects include those due to grain boundaries (Hall-Petch) [10] and particle reinforcements [11]. In all of these cases, strain gradients are involved and, in such instances, there is general agreement that the size effect, which generally manifests itself at the micron scale, can be attributed to hardening enabled by geometrically-necessary dislocations (GNDs) [12]. The constitutive laws that ensue include a contribution from the uniform plastic strain, ε_p, and another incorporating a length scale, ℓ, in conjunction with a plastic strain gradient, $\varepsilon_p^* \equiv \left| d\varepsilon_p / dx \right|$. The uniform strain is associated with the statistically stored dislocations (SSDs), density ρ_{SSD}, and the gradient with the GNDs, density ρ_{GND} [7]. To establish a viable computational scheme, these contributions must be combined in an appropriate manner. Almost twenty five years after the introduction of the first theory of strain gradient plasticity [13], and following a decade of active research into small scale plasticity, the aim of this article is to provide a critical assessment from the vantage points of both mechanics and materials science by delving into two basic approaches for combining the strain and its gradient. One theory developed by Fleck and Hutchinson [14, 15] (FH) is based on plastic dissipation and considers separate, additive, contributions from the SSNs and GNDs. Another, introduced by Nix and Gao [4, 16-18] (NG), invokes an enhanced flow stress governed by the density of the GNDs. Both formulations have the following commonalities. (i) They are phenomenological in that they employ

isotropic measures of the plastic strain and its gradient. (ii) They aspire to extend the classical J$_2$ theory of plasticity in the simplest meaningful manner into the range where size dependence becomes important. (iii) They reduce to J$_2$ theory when the scale of the gradient is large compared to ℓ. Nevertheless, they predict different trends in the flow stress. The distinctions between them will be highlighted in this article by comparing and contrasting predictions for an especially straightforward loading situation, described below. Thereafter, the merits and limitations of the theories will be discussed and alternative suggestions made.

The distinctions between the theories are most clearly revealed by bending, because the total strain gradient is spatially constant, inducing both SSDs and GNDs as soon as plastic flow commences. Data for the bending moment induced in Ni foils as a function of curvature are used (Fig. 1) [19]. Details of the experiments are given elsewhere [19]. The principle of the test design is to bend a thin foil to a prescribed curvature. When unloaded, the foil relaxes elastically to a smaller, permanent curvature. The decrease in curvature upon unloading provides a measure of the bending moment per width, M, without requiring independent measurement. The results of the bending measurements, conducted for foils of thickness, h, are presented both as plots of the normalized bending moment, M / h^2, as a function the surface strain, $\varepsilon_S = \kappa h / 2$, and in terms of the *reduced surface strain*, defined as the difference between the surface strain and the surface strain due to M for elastic bending: $\varepsilon_S^R = \varepsilon_S - 6(1 - v^2)M / (Eh^2)$. (The elastic strain contribution is of no significance for present purposes). The normalization used in these plots would collapse the data onto a single curve in the absence of a material size effect. The results are presented in a manner that highlights the separate dependencies on foil thickness and grain size. Those summarized on Fig. 1a,c are for foils having the same grain size ($g = 27 \mu m$) but differing thickness (10, 50 and 125 μm), revealing the increase in flow stress with decrease in thickness at fixed grain size. The corresponding results on Fig. 1b,d are all for foils having the same thickness ($h = 50 \mu m$) but four different grain size (6, 14, 27 and 50μm). The increase in flow stress with decrease in grain size is evident. Tensile stress/ plastic strain measurements for the same materials with three of the four different grain sizes are presented in Fig. 2. These measurements will be assessed in the context of the NG and FH theories.

In order to bring-out differences between the theories, two additional sets of measurements are invoked (Fig. 3). One comprises the series of torsion tests conducted on Cu wires of different diameter [7]. Note the strong influence of the diameter on the yield strength, but relatively small effect on the strain hardening. The other is the comparison between the tensile response of thin foils with and without a passivated surface layer. The presence of the passivated layer also increases the yield strength, with a much smaller effect on the strain hardening.

The present assessment continues that initiated by Nix and Gao [4] by focusing on the fidelity and underpinnings of the simplest generalizations of conventional plasticity. It also emphasizes the distinction between lower and higher order gradient plasticity theories, to make the point that only the latter is capable of solving important categories

of problem, such as those involving dislocation pile-ups. The unresolved issues *it does not attempt to address* are as follows. (i) Details of the higher order formulations and the full scope of the boundary conditions to which these apply [20, 21]. (ii) The partitioning of the plastic work due to strain gradients into dissipative and recoverable contributions under highly non-proportional multi-axial stressing [22]: not at issue for pure bending. (iii) Single crystal formulations of strain gradient plasticity, whether cast within a continuum framework [23, 24] or based on discrete dislocations [25].

The article is organized as follows. Basic aspects of plasticity relevant to gradient effects are presented, the two formulations are described and general trends elucidated. Thereafter, the predictions of the theories for bending are compared with the measurements, enabling conclusions to be reached about their respective merits and deficiencies. Finally, the discussion is broadened by introducing other measurements that support the conclusions and by clarifying the scaling, as well as the connection to dislocation-based concepts of hardening.

2. The interaction of plastic strains and strain gradients

The focus is on the manner whereby the plastic strains and their gradients are combined in the strain gradient plasticity (SGP) theories. Issues related to the pertinence of lower or higher order theories (Appendix I) do not explicitly affect the bending assessment. However, because the distinction is crucially important for certain classes of problems, one example requiring the higher order formulation is given in section 4. The simpler deformation (rather than incremental) theories of plasticity are used, consistent with the bending moments being measured at increasing curvature (Fig. 1): involving monotonic plane strain tension or compression.

Definitions: The effective stress and plastic strain are, respectively, $\sigma_e = \sqrt{3 s_{ij} s_{ij} / 2}$, and $\varepsilon_P = \sqrt{2 \varepsilon_{ij}^P \varepsilon_{ij}^P / 3}$, with ε_{ij}^P as the plastic strain tensor and s_{ij} the stress deviator. In uniaxial tension σ_e and ε_P coincide with the stress and the plastic strain, respectively, and the uniaxial stress-strain curve is characterized by the function f according to $\sigma_e = \sigma_Y f(\varepsilon_P)$, with σ_Y the yield strength. This relation holds for proportional multiaxial stressing in the absence of gradients. The associated plastic work/volume is:

$$U_P(\varepsilon_P) = \sigma_Y \int_0^{\varepsilon_P} f(\varepsilon_P) d\varepsilon_P \qquad (1)$$

The Theories: The two theories each provide a length scale, designated ℓ_{NG} and ℓ_{FH} for Nix/Gao and Fleck/Hutchinson, respectively. Definitions for these lengths, within the context of the theories, will be provided in this section (see also Appendix I). The interpretation of the lengths in terms of the governing dislocation phenomena will be deferred to Section 6. The theory introduced by Nix and Gao [4, 16] is inspired by Taylor hardening, with flow stress governed by the interaction of the mobile dislocations with the statistically-stored (SSDs) and geometrically-necessary dislocations (GNDs), through

a linear summation of their densities: $\sigma_e \sim Gb\sqrt{\rho_{SSN} + \rho_{GND}}$, where G is the shear modulus and b the Burgers vector. Namely, interactions between the SSDs and GNDs are not considered. The theory is expressed as a gradient enhanced flow stress:

$$\sigma_e = \sigma_Y \sqrt{f(\varepsilon_P)^2 + \ell_{NG}\varepsilon_P^*} \qquad (2)$$

where ε_P^* is the strain gradient and ℓ_{NG} has dimensions of length. The first term is attributed to the SSDs and reproduces uniaxial tension data. The second is associated with the GNDs. The length parameter, ℓ_{NG}, is chosen to fit experimental data. This formulation may be re-expressed in terms of a reference stress [the tensile flow stress at a strain of unity, $\sigma_{ref} = \sigma_Y f(1)$],

$$\sigma_e = \sigma_{ref} \sqrt{\left(f(\varepsilon_P)/f(1)\right)^2 + \overline{\ell}_{NG}\varepsilon_P^*} \qquad (3)$$

The implications of the re-scaled length parameter, $\overline{\ell}_{NG} = \ell_{NG}/f(1)^2$, will be discussed in Section 6.

The premise of the Fleck and Hutchinson [14, 15] theory is that the movement of the SSDs and GNDs results in plastic dissipation governed by an effective plastic strain, E_P, re-expressed from (1) as:

$$U_P(E_P) = \sigma_Y \int_0^{E_P} f\left(\varepsilon_P\right)d\varepsilon_P \qquad (4)$$

In the ensuing assessment, the preferred measure of E_P is the linear summation,

$$E_P = \varepsilon_P + \ell_{FH}\varepsilon_P^* \qquad (5)$$

with ℓ_{FH} the associated material length parameter. The first term provides *the plastic dissipation caused by the motion of the SSDs and the second that attributed to the GNDs*. This choice reduces to classical theory in the limit when gradients are small, with $E_P \cong \varepsilon_P$. [More generally, E_P has been prescribed by the homogeneous composition

$$E_P = \left(\varepsilon_P{}^\mu + (\ell_{FH}\varepsilon_P^*)^\mu\right)^{1/\mu} \qquad (6)$$

The version with $\mu=2$ has been employed in most studies using FH theory. For reasons elabrated below and in Appendix II, the linear summation with $\mu=1$ is more attractive and will be used hereafter].

General trends can be revealed by considering the non-uniform deformations of an object, size h, that experiences average plastic strain ε_P and average strain gradient $\varepsilon_P^* = c\varepsilon_P / h$ (where $c \approx 1$). Trends are revealed most clearly for a perfectly plastic solid, with $f(\varepsilon_P) = 1$. [Upon approximating the average of the square root by the square root of the average], the NG formulation predicts an average flow stress

$$\sigma \approx \sigma_Y \sqrt{1 + (c\ell_{NG} / h)\varepsilon_P} \tag{7}$$

plotted in Fig. 4b for various h. Note that strain gradients do not elevate the yield strength. Instead, they increase the rate of strain hardening. Moreover, the gradient effect only becomes important when $(c\ell_{NG} / h)\varepsilon_P$ is of order unity. By invoking a similar approximation, the corresponding trends for the FH formulation are

$$E_P \approx \left[1 + (c\ell_{FH} / h)\right]\varepsilon_P \tag{8}$$

and

$$U_P(E_P) = \sigma_Y \left[1 + (c\ell_{FH} / h)\right]\varepsilon_P \tag{9}$$

with average flow stress

$$\sigma \approx \sigma_Y \left[1 + (c\ell_{FH} / h)\right] \tag{10}$$

In this formulation (Fig. 4a), the gradient elevates the yield strength but not the rate of hardening. Moreover, the flow stress increases linearly with $c\ell_{FH} / h$ independent of ε_P.

The foregoing distinction is persistent and represents one of the major differentiators between the formulations. Namely, the NG formulation increases the hardening rate, but not the yield strength, while FH formulation increases the yield strength, with a second order influence on hardening.

3. Application to pure bending

The size dependence of the moment-curvatures relation in pure bending will be explored using the NG and FH models. The tensile stress-strain data for the Ni foils (Fig. 2), over the range of interest, can be accurately represented by

$$\begin{aligned}
\sigma &= E\varepsilon \ (\varepsilon_P = 0), & \sigma \le \sigma_Y \\
\sigma &= \sigma_Y \left[1 + \lambda\left(\varepsilon_P / \varepsilon_Y\right)^N\right], & \sigma > \sigma_Y
\end{aligned} \tag{11}$$

with $\varepsilon_Y = \sigma_Y / E$ the yield strain. The parameters σ_Y, λ and N are chosen to fit data for each grain size (Table I). Elasticity will be taken to isotropic with Young's modulus, $E = 220 GPa$. To simplify the analysis, elastic compressibility will be neglected by taking Poisson's ratio to be 1/2, with little effect on the results of interest in the plastic range. The only other input is the material length scale. For each model, this will be chosen to fit the data.

The bending data are restricted to small strains, justifying use of linear strain-displacement theory. In pure bending, the strain variation across the cross-section is, $\varepsilon_{11} \equiv \varepsilon = \kappa y$, with κ as the imposed curvature and y as the distance from the mid-plane. For plane strain with elastic incompressibility, the effective plastic strain is related to the component parallel to the foil by

$$\varepsilon_P = 2\left|\varepsilon_{11}^P\right| / \sqrt{3} \qquad\qquad (12)$$

The plastic strain distribution, $\varepsilon_P(y)$, is the primary unknown in all of the theories. The strain at the surface of a foil, thickness h, is, $\varepsilon_S = \kappa h / 2$. The curvature at which initial yield occurs at the surface in the conventional solid ($\ell = 0$) is, $\kappa_Y = \sqrt{3}\varepsilon_Y / h$, and the associated moment/length is, $M_Y = \sigma_Y h^2 / (3\sqrt{3})$.

Foils with fixed grain size ($g \cong 27\,\mu m$) at three different thicknesses. Curves of moment against reduced surface strain based on the NG model are plotted in Fig. 5. Details of the calculations are given in Appendix I. The length parameter was chosen to fit the measurements for the foil having thickness $h = 50\,\mu m$ (for which the data are most extensive) at $\varepsilon_S = 2 \times 10^{-3}$ ($\varepsilon_R \cong 1.6 \times 10^{-3}$, $\kappa / \kappa_Y = 10.5$, $M / h^2 \cong 20 MPa$). Thereafter, the same length is used to predict the remaining curves. To obtain a fit, a large length, $\ell_{NG} = 25mm$, was required. The corresponding fit plotted in Fig. 5 for the FH formulation requires a much smaller length parameter, $\ell_{FH} = 5\,\mu m$ (computational details in Appendix I).

Foils having the same thickness ($h = 50\,\mu m$), but different grain size (Fig. 6). The analysis uses the input parameters summarized in Table I, with the length parameters chosen independently for each grain size to give the fit shown. (Curves for grain size $g = 50\,\mu m$ where not computed because uniaxial stress-strain data were unavailable). Significantly, the length parameters (ℓ_{NG} and ℓ_{FH}) are only modestly dependent on grain size. The trends in Figs. 5 and 6 by which the two theories predict flow stress elevation versus hardening elevation follows the pattern noted earlier.

The reason for the large ℓ_{NG} becomes apparent from (2) upon realizing that, for the size effect to be appreciable, $\ell_{NG}\varepsilon_P^*$ must be comparable to $f(\varepsilon)$: which, in turn, is only

slightly larger than unity. Namely, because the plastic strains are only of order 10^{-3}, the gradient is $\varepsilon_P^* \approx (10^{-3}/h)$, requiring that $\ell_{NG} \approx 10^3 h$ for $\ell_{NG}\varepsilon_P^*$ to be comparable to unity. The existence of such large ℓ_{NG} is unprecedented and attributed to its strong strain dependence. All previous assessments have used stress data at much larger strains, inferring smaller ℓ_{NG}. Notably, results for the torsion of copper wires (Fig. 3) [17] twisted to surface strains of order unity, upon using $\sigma_Y \approx \sigma_{ref}$, gave ℓ_{NG} in the micron range: while indentation results (average strains typically 0.1) typically infer tens of microns [6].

By contrast, ℓ_{FH} is in the micron range for all cases: small strain bending, large strain torsion, indentation. The FH formulation does not give rise to appreciable strain-dependence for a given material because, in the definition of the effective plastic strain, E_P, the plastic strain gradient is "balanced" against the plastic strain. In other words, the *relative* importance of the two contributions does not vary with the overall strain level.

The comparison of the predicted and measured moment/curvature curves in the plastic range (Figs. 7 and 8) reveals discrepancies for both models: notwithstanding that some of the deviations are associated with measurement fidelity at the smallest foil thickness. Namely, these measurements (together with those in Fig. 2) provide a clear indication of a yield strength elevation with thinness, at variance with NG. However, FH is also at variance with the bending measurements. It predicts (sometimes substantially) larger yield strength elevations and lower hardening rates than found experimentally. One implication is that both models, in their simplest (foregoing) manifestation, are inadequate. Another is that, a single comparison is insufficient to distinguish the respective models. Nevertheless, the discrepancies have motivated introduction of a two-parameter, FH model (described in section 5) that achieves closer correspondence between measurement and theory.

For completeness, we note that the FH theory with $\mu = 2$ (Appendix II) has the weakest correlation with the measurements [due to lowest order strengthening proportional to $(\ell\varepsilon_P^*)^2$].

The approach to the size-independent limit. The lowest order strengthening dependence on $\ell\varepsilon_P^*$ is exemplified by the appreciable body of Berkovitch nano-indentation data [4]. These data reveal that, as the depth of penetration, d, becomes large, the hardness, H, approaches the limit for large indents, H_0, in accordance with "square root" proportionality $[(H/H_0)^2 - 1 \propto 1/d]$. The analog in bending is a plot of M^2 against $1/h$ at a fixed surface strain ($\varepsilon_s = 2\kappa/h$). The plot (Fig. 9) establishes that both the NG and the FH theories (with $\mu = 1$) are consistent with this trend: a direct consequence of the lowest order dependence of the formulations on $\ell\varepsilon_P^*$.

4. An illustration of the importance of a higher order theory.

The foregoing bending solutions have assumed that the plastic deformation at the surfaces is unconstrained (Appendix I). Suppose, instead, that dislocation motion is blocked by a thin passivation film: whereupon the plastic strain vanishes at the surfaces (Fig. 3) [9]. Such a boundary condition cannot be modeled by any lower order theory. But a higher order theory includes the possibility of specifying an additional boundary condition $\varepsilon_P = 0$ at the surfaces (Appendix I). The basis for the difference is that, in the lower order theory, the gradient only affects the incremental moduli: Otherwise it involves conventional measures of the stress and strain. As a result, only conventional boundary conditions can be enforced. By contrast, any theory based on the plastic work is inherently higher order since the work depends on the strain and its gradient, and both must be retained as essential variables. There are two consequences. (i) The theory embraces additional boundary conditions. (ii) An additional stress-like quantity arises in with dimensions *stress x length* denoted by T. This new quantity is referred to as a "higher-order stress" or, in some contexts, as a couple- or moment-stress; T works through the strain gradient to generate the extra dissipation caused by the GNDs (Appendix I).

The influence on the moment-surface strain relation predicted by FH is plotted on Fig. 10 for foils with grain size $g = 27 \mu m$ and thickness $h = 50 \mu m$. The corresponding distributions of effective plastic strain (at $\kappa / \kappa_Y = 10.3$) are also shown in Fig. 10. Note that the effect of passivation in bending is large because yielding starts at the surfaces where plasticity is severely constrained.

We re-emphasize that no conventional plasticity theory, or lower order theory of any kind, could distinguish between passivated and unpassivated surfaces. The higher order formulation provides the flexibility needed to specify boundary conditions not encompassed by lower order theories. This flexibility is illustrated through schematics of the spatial organizations of the GNDs in bending (Fig. 11) along with the associated continuum plastic strain distributions, $\varepsilon_{11}^p(y)$ (c.f. Fig. 10). For these distributions, the dislocation density is, $\rho_{GND} = b^{-1} d\varepsilon_{11}^p / dy$. The dislocations exit through an unpassivated surface (a dislocation pile-up cannot be sustained): whereupon, at the surface, $\rho_{GND} = b^{-1} d\varepsilon_{11}^p / dy = 0$, and T vanishes. By contrast, dislocations pile-up at the passivated surface, with non-zero ρ_{GND} and T; while the plastic strain vanishes. The mathematical structure of the higher order theory, including the definition of T and how it relates to the boundary conditions, is presented in Appendix I.

5. Formulation with a strain-dependent length scale.

An attractive feature of both the NG and FH models is that they require the specification of only a single material length parameter in addition to the standard inputs to conventional J_2 plasticity theory (Appendix I). As evident from the comparison between

predictions and data in Figs 7 and 8, this simplicity also constrains the ability of the theory to reproduce measurements. Both models can incorporate additional parameters, permitting greater flexibility. The process is illustrated for the FH model by generalizing (7), through a strain dependent length scale, according to

$$E_P = \varepsilon_P + \ell(\varepsilon_P)\,\varepsilon_P^*$$
(13)

as described in Appendix I. The specific example in Fig. 12 for the foil with $g = 27\,\mu m$ and $h = 50\,\mu m$ introduces two parameters, ℓ_0 and a_0, as

$$\ell(\varepsilon_P) = \ell_0 \left[1 + a_0(\varepsilon_P / \varepsilon_Y) \right]$$
(14)

where a_0 clearly invokes a linear dependence of the length scale on the plastic strain. By decreasing the initial strengthening and increasing the rate of hardening, the parameters chosen in Fig. 12 clearly give a better fit to the data. While adding another parameter is hardly desirable, nevertheless it is comforting to establish that the theories can be extended in a relatively straightforward manner to more closely replicate a wide range of measurements incorporating strain gradients.

6. The interpretation of the length scale.

Strain gradient theories have not been comprehensively embraced for several reasons. Most importantly, the ambiguities surrounding the length parameters that emerge upon fitting predictions from different theories to measurements are a source of confusion, which the foregoing bending tests highlight. A further hindrance is the lack of clarity associated with the interpretation of the theories within an explicit dislocation dynamics context [26]. The following discussion provides rudimentary connections between the length scales and dislocation-based phenomena.

Before proceeding we recall that both the NG and FH formulations appeal to GNDs as the source of the size effects, and both relate the density of GNDs to the plastic strain gradient in essentially the same manner (see Appendix I). Divergences between the two theories arise from the manner in which the strain and the strain gradient are combined, as expounded upon in Section 2, leading to the huge differences between the length parameters. Neither theory makes the distinction that the GNDs induced by the strain gradients are closely coordinated, whereas the SSDs are relatively uncoordinated. Moreover, neither theory accounts for interactions among the dislocation categories.

Nix and Gao [4] provide a connection between their inferred length scale and basic dislocation interactions as follows. Based on Taylor hardening, the length scale is related to the flow stress by (2):

$$\ell_{NG} = \chi (G / \sigma_Y)^2 b$$
(15)
$$\overline{\ell}_{NG} = \chi (G / \sigma_{ref})^2 b$$

with $\chi = m^2 \alpha^2 \bar{r}$, where m=3.06 is the Taylor factor connecting tension and shear in an FCC polycrystal, $\alpha \cong 0.2 - 0.4$ is the Taylor hardening coefficient, and $\bar{r} \cong 1.85$ is the Nye factor. If the yield strength is set by the current density of SSDs, Taylor hardening gives $\sigma_Y \approx Gb\sqrt{\rho_{SSD}} \approx Gb/s$, with s representing an effective spacing between obstacles (dislocation cell size or precipitate spacing). Combining with (15) [apart from numerical factors] gives,

$$\ell_{NG} \approx b\left(G/\sigma_Y\right)^2 \approx s^2/b \tag{16}$$

The spacing estimated from (16) in conjunction with the inferred length scale $\ell_{NG} = 25mm$ is credible: $s \approx 2.5\mu m$. Nevertheless, we note a persistent discrepancy. For example, for the foregoing bending measurements with $g = 27\mu m$ ($\sigma_Y = 42MPa$) formula (15) predicts, $\ell_{NG} = 2.8mm$ (with $\alpha = 0.4$): much smaller than $\ell_{NG} = 25mm$ inferred by reproducing the data. [Alternatively, assuming $\sigma_{ref} \approx 10\sigma_Y$, (15) gives, $\bar{\ell}_{NG} = 28\mu m$: while the fit to the experimental data gives, $\bar{\ell}_{NG} = 250\mu m$.]

The clearest microscopic interpretation of ℓ_{FH} emerges upon letting $\Delta\rho_{SSD}$ and $\Delta\rho_{GND}$ be the densities of mobile SSDs and GNDs associated with the plastic strain ε_P and gradient ε_P^*, respectively, and $\bar{d}$ the average distance they travel (again set by dislocation cell size or precipitate spacing). The plastic dissipation is

$$U_P \approx \sigma_Y \left(\Delta\rho_{SSD} b\bar{d} + \Delta\rho_{GND} b\bar{d}\right) \tag{17}$$

Given that, $\varepsilon_P \approx \Delta\rho_{SSD} b\bar{d}$ and $\varepsilon_P^* \approx \Delta\rho_{GND} b$, then

$$U_P \approx \sigma_Y \left(\varepsilon_P + \bar{d}\varepsilon_P^*\right). \tag{18}$$

Based on (4) and (5),

$$U_P \approx \sigma_Y E_P = \sigma_Y \left(\varepsilon_P + \ell_{FH}\varepsilon_P^*\right),$$

The identity, $\ell_{FH} \approx \bar{d}$, implies that the length scale is *determined by the average distance between obstacles*. This outcome is consistent with the statement that ℓ_{FH} sets the scale of the gradient: whereupon the motions of the GNDs and SSDs contribute equally to the dissipation. It is also notable that, for all of the cases interpreted using FH, there is a strong inverse correlation between ℓ_{FH} and the yield strain (Fig. 13). That such a correlation exists is consistent with the proposed dependence of ℓ_{FH} on the distance moved by the dislocations: which, in turn, dictates the plastic strain.

The two microscopic interpretations of the length scale are not necessarily in conflict. Both estimates are within the realm of possibility. Recall that, instead, the important divergences between the formulations are two-fold. (i) The strong strain dependence of ℓ_{NG} (or $\overline{\ell}_{NG}$) contrasts with the relative strain invariance of ℓ_{FH}. (ii) The size effect in the NG theory involves increases in rates of hardening, with little initial strength increase: whereas the FH theory predicts a substantial increase in yield strength with smaller increase in the rate of strain hardening. The strong strain dependence of ℓ_{NG} (or $\overline{\ell}_{NG}$) can also be problematic. For example, indentation tests are one of the simplest and most robust ways to determine the material length parameter. Without accurately accounting for the strain dependence through independent measurements, it would not be possible to use ℓ_{NG} ascertained from indentation tests to predict the small strain bending of foils.

7. Concluding Remarks

By pursuing the two prominent strain gradient theories within the context of simple bending, the following important differences have been highlighted. The strong strain dependence of ℓ_{NG} (or $\overline{\ell}_{NG}$) contrasts with the relative strain invariance of ℓ_{FH}. The size effect in NG theory involves an increase in hardening, with little effect on initial yield: whereas FH theory predicts a substantial increase in yield strength with only a small increase in strain hardening. Both theories deviate from the bending measurements but, in a broader context, other measurements reveal a larger effect of the gradient on yielding than on strain hardening. By adding an extra coefficient, the theories can be adjusted to attain levels of strengthening and strain hardening that more closely match the measurements. This capability has been illustrated for the FH theory.

The length scales inferred by fitting to experimental measurements have the following rudimentary connections to dislocation phenomena. The strengthening in the NG theory is governed by the spacing between the GNDs. The FH theory is based on the plastic dissipation, enabling the length scale to be related to the average distance between obstacles moved by the GNDs. This correlation is the basis for the inverse dependence of ℓ_{FH} on the yield strain.

The difference between lower order (conventional NG) and higher order (FH) theory has been emphasized and illustrated by the example of surface passivated foils in bending. The distinction is that the low order theory is restricted in terms of the boundary conditions that can be imposed, excluding the solution of various important problems. The higher order theory does not have this restriction because it introduces a new quantity (with dimensions *stress x length*) governing the plastic work within the strain gradient: analogous to the dissipation within dislocation pile-ups. Accordingly, higher order theory can solve problems involving locations where the plastic strain is zero, due to the blockage of dislocations.

The version of FH pursued in this article has used the linear summation of the plastic strains and their gradients ($\mu = 1$) because this choice correlates with well-established

"square root" size scaling trends found in hardness and other tests. Most prior applications conducted using FH have used $\mu = 2$, largely because the numerical implementation is most straightforward for this case. However, given the disconnect between this version and scaling trends, we propose that it be abandoned or restricted to the mid-range of the size effect. The ensuing challenge is that, to further progress with the linear version of FH, non-standard numerical methods must be devised to obtain solutions to generalized problems.

Appendix I: Synopsis of Low and High Order Theories

The starting point for all phenomenological isotropic theories is the definition of a measure of the gradient of plastic strains, ε_P^*. The FH formulation uses the most general measure based on the three quadratic invariants, I_i $(i = 1,3)$, of the gradients, $\varepsilon_{ij,k}^P$:

$$\sqrt{\ell_1^2 I_1 + \ell_2^2 I_2 + \ell_3^2 I_3} \equiv \ell_{FH} \varepsilon_P^* \tag{19}$$

Definitions of these invariants, which are non-negative and have dimensions (length)$^{-2}$, are given in [14, 15]. Three length parameters, ℓ_i $(i = 1,3)$, arise. For bending in plane strain,

$$\sqrt{\ell_1^2 I_1 + \ell_2^2 I_2 + \ell_3^2 I_3} = \sqrt{\tfrac{4}{5}\ell_1^2 + \ell_2^2 + \tfrac{6}{5}\ell_3^2} \left|\varepsilon_P'\right| \equiv \ell_{FH} \varepsilon_P^* \tag{20}$$

The NG formulation [16, 17] also employs (19) but the number of length parameters is reduced to one by fixing the ratios as: $\ell_1 = \ell_2 = \ell_3$.[1] The measure of plastic strain gradients is taken as

$$\eta_p = (1/2)\sqrt{I_1 + I_2 + I_3} \tag{21}$$

and the contribution to the flows stress in (2) is $\ell \eta_p$. In plane strain bending,

$$\ell \eta_p = \ell \frac{\sqrt{3}}{2}\left|\varepsilon_P'\right| \equiv \ell_{NG} \varepsilon_P^* \tag{22}$$

Note that the length parameter, ℓ, as defined by Gao, et al. [16, 17] is $2/\sqrt{3}$ times the length employed in this paper, ℓ_{NG}.

---

[1] This restriction is in accord with the fact that the three lengths are not associated with distinct physical mechanisms. While a single length reduces somewhat the flexibility to fit data from different types of deformations (e.g. wire torsion and indentation [5]), the restriction taking the three parameters to be equal is a good compromise as discussed in [17].

Distinctions between lower and higher order gradient theories are illustrated within the context of pure bending. For an elastically incompressible material in plane strain and subject to $\kappa > 0$, the relations introduced earlier hold for $y \geq 0$:

$$\varepsilon_{11} \equiv \varepsilon = \kappa y, \quad \varepsilon_P = 2\varepsilon_{11}^P / \sqrt{3}, \quad \varepsilon_P^* = |d\varepsilon_P / dy| \qquad (23)$$

In both formulations the conventional stresses are given by

$$\sigma_{11} = (4/3)E\left(\varepsilon_{11} - \varepsilon_{11}^P\right) = (4/3)E\left(\kappa y - \sqrt{3}\varepsilon_P / 2\right) \qquad (24)$$

with $\sigma_{33} = \sigma_{11}/2$ and the effective stress is $\sigma_e = \sqrt{3}\sigma_{11}/2$. Because the two non-zero stress components are functions only of y, the conventional equilibrium equations, $\sigma_{ij,j} = 0$, are identically satisfied.

Lower order gradient theory: NG as an example. Most applications of NG have made use of the lower order theory [18] (albeit that higher order versions have been proposed [16, 17]). For problems more complicated than pure bending, a numerical technique (typically finite elements) is needed to obtain solutions for the stresses, based on the incremental equations governing stress-strain, equilibrium and strain-displacement. These equations have same form as conventional plasticity, except that the incremental moduli relating stress and strain increments in the plastically deforming regions would have an additional dependence on the gradient of plastic strain. The effects of the gradients are innocuous in that they do not fundamentally alter the structure of conventional plasticity [27]. In particular, they do not require, or permit, specification of extra boundary conditions, such as the constraint on plastic flow imposed by passivation described in section 4.

Bending subject to increasing κ is sufficiently simple that the following direct approach is the most straightforward way to produce the solution. Let $\varepsilon_Y = \sigma_Y / E$ be the initial tensile yield strain and $\kappa_Y = \sqrt{3}\varepsilon_Y / h$ be the curvature at which yielding starts at the surface ($y = h/2$). When $\kappa > \kappa_Y$, the region $y \leq Y \equiv \sqrt{3}\varepsilon_Y / (2\kappa)$ is elastic ($\varepsilon_P = 0$) while yielding ($\varepsilon_P \geq 0$) occurs for $y > Y$. By (24),

$$\varepsilon_{11}^P \equiv \frac{\sqrt{3}}{2}\varepsilon_P = \kappa y - \frac{\sqrt{3}}{2}\frac{\sigma_Y}{E}\sqrt{f\left(\varepsilon_P\right)^2 + \ell_{NG}\varepsilon_P^*} \qquad (y \geq 0) \qquad (25)$$

where the gradient enhanced NG flow stress σ_e from (2) has been introduced. Solving for $\varepsilon_P^* = d\varepsilon_P / dy$ and introducing $\bar{\varepsilon}_P = \varepsilon_P / \varepsilon_Y$ and $\eta = 2y/h$, gives a first order ordinary differential equation for $\bar{\varepsilon}_P$:

$$\frac{d\bar{\varepsilon}_P}{d\eta} = \frac{1}{(2\ell_{NG}/h)\varepsilon_Y}\left[\left(\frac{\kappa}{\kappa_Y}\eta - \bar{\varepsilon}_P\right)^2 - f\left(\bar{\varepsilon}_P\right)^2\right] \tag{26}$$

The only boundary condition that can be enforced is the requirement that $\bar{\varepsilon}_P = 0$ at the elastic-plastic boundary at $y = Y$ ($\eta = \eta_Y \equiv \kappa_Y/\kappa$). It is not possible to additionally constrain the plastic strain at the surface. The results presented in body of the paper are obtained by integrating (26) numerically from η_Y to the surface at $\eta = 1$. For $\kappa > \kappa_Y$, the moment/length is

$$M = 2\int_0^{h/2}\sigma_{11}y\,dy = M_Y\left\{\left(\frac{\kappa_Y}{\kappa}\right)^2 + 3\int_{\eta_Y}^1\left(\frac{\kappa}{\kappa_Y}\eta - \bar{\varepsilon}_P\right)\eta\,d\eta\right\} \tag{27}$$

where $M_Y = \sigma_Y h^2/(3\sqrt{3})$ is the moment/length at the onset of plastic yielding. The equation governing the plastic strain distribution for conventional plasticity is obtained from (26) in the limit $\ell_{NG} \to 0$, i.e., $\bar{\varepsilon}_P + f(\bar{\varepsilon}_P) = (\kappa/\kappa_Y)\eta$

Higher order gradient theory: FH as an example. The FH formulation employs the plastic work required to deform a material element, $U_P(E_P)$, with E_P given by (5). Attention is again focused on the upper half of the beam. For the deformation theory formulation, the energy/length required to deform the beam to curvature $\kappa > \kappa_Y$ is

$$\begin{aligned}\Phi &= 2\int_0^{h/2}\left[\frac{1}{2}\sigma_{11}\varepsilon_{11}^e + U_P(E_P)\right]dy \\ &= 2\int_0^{h/2}\left[\frac{2}{3}E\left(\kappa y - \frac{\sqrt{3}}{2}\varepsilon_P\right)^2 + U_P(E_P)\right]dy\end{aligned} \tag{28}$$

The first term is the elastic energy and the second is the plastic work. For the case of no constraint on plastic flow at the surface, the plastic strain distribution, $\varepsilon_P(y)$, is determined by minimizing Φ with respect to all continuous distributions ε_P that vanish for $y \leq Y$ and are non-negative for $y > Y$. The first variation of Φ with respect to ε_P is

$$\begin{aligned}\delta\Phi &= 2\int_Y^{h/2}\left[-\frac{2}{\sqrt{3}}E\left(\kappa y - \frac{\sqrt{3}}{2}\varepsilon_P\right)\delta\varepsilon_P + \frac{\partial U_P}{\partial\varepsilon_P}\delta\varepsilon_P + \frac{\partial U_P}{\partial\varepsilon_P'}\delta\varepsilon_P'\right]dy \\ &= 2\int_Y^{h/2}\left[-\sigma_e\delta\varepsilon_P + \frac{\partial U_P}{\partial\varepsilon_P}\delta\varepsilon_P + \frac{\partial U_P}{\partial\varepsilon_P'}\delta\varepsilon_P'\right]dy\end{aligned} \tag{29}$$

with $(\ )' = d(\ )/dy$. The inevitable consequence of the dependence of energy on the gradient of plastic strain is *that a new quantity with dimensions stress x length (N / m) arises*, not present in conventional plasticity. This parameter, designated T, is the work conjugate to the plastic strain gradient, ε'_P and identified as

$$T = \partial U_P / \partial \varepsilon'_P \qquad (30)$$

That is, $T \delta \varepsilon'_p$ is interpreted as the contribution to the plastic work increment due to the geometrically necessary dislocations. Integrating the third term in (29) by parts, gives

$$\delta \Phi = 2 \int_Y^{h/2} \left[-\sigma_e + Q - T' \right] \delta \varepsilon_P \, dy + 2 \left[T \delta \varepsilon_P \right]_Y^{h/2} \qquad (31)$$

where $Q = \partial U_P / \partial \varepsilon_P$. The requirement that $\delta \Phi = 0$ for all $\delta \varepsilon_P$, satisfying $\varepsilon_P = 0$ at $y = Y$, provides the equilibrium equation for the higher order stress and the extra boundary condition at the surface:

$$T' - Q + \sigma_e = 0 \quad \text{on } Y \le y \le h/2 \qquad (32)$$

and

$$T = 0 \text{ at } y = h/2. \qquad (33)$$

For the case where plastic flow is blocked at the surface, the admissible distributions must be additionally constrained such that $\varepsilon_P = 0$ at $y = h/2$; then, T does not vanish at the surface. Because $d\Phi = Md\kappa$, one can show that M is again given by (27).

In summary, for bending, in the lower order theory, a first order, ordinary differential equation governs ε_p and only the condition that can be enforced is that ε_p vanish at the elastic-plastic boundary. The equilibrium equation for the higher order theory, (32), when expressed in terms of ε_p, becomes a second order, ordinary differential equation. Boundary conditions at both the elastic-plastic boundary and at the surface can be imposed, modeling either blocked or unblocked dislocation motion (or even intermediate conditions with further embellishment [21]).

Inclusion of a strain dependence of the length parameter, $\ell_{FH}(\varepsilon_P)$, as illustrated by (14), creates no difficulties. The dependence, $\ell_{FH}(\varepsilon_P)$, appears in E_P and in $U_P(E_P)$ and is directly accounted for in the quantities $Q = \partial U_P / \partial \varepsilon_P$ and $T = \partial U_P / \partial \varepsilon'_P$. In other words, the additional flexibility afforded by the strain-dependent length parameter creates no theoretical or computational obstacles [28].

Results presented in the figures were obtained by minimizing Φ with respect to $\varepsilon_P(y)$ subject to the constraints enumerated above, with

$$U_P(E_P) \equiv \int_0^{E_P} \sigma_e d\varepsilon_P = \sigma_Y \left[E_P + \frac{\lambda \varepsilon_Y}{(N+1)} \left(\frac{E_P}{\varepsilon_Y} \right)^{N+1} \right] \tag{34}$$

The minimization process was implemented numerically. For $\mu > 1$ various methods can be used, including a one-dimensional finite element method with a piecewise linear approximation to $\varepsilon_P(y)$. For $\mu = 1$, the choice of methods narrows, because the minimum is not necessarily analytic and the solution for $\varepsilon_P(y)$ can have discontinuities in its first derivative. In this paper, the interval $Y \le y \le h/2$ was subdivided into equal elements, a piecewise linear approximation to $\varepsilon_P(y)$ was used, and the minimum of Φ was obtained using a standard routine for constrained optimization. To apply this formulation to two- and three-dimensional problems, efforts will have to be made to identify efficient and robust numerical solution methods.

Appendix II: The Exponent in the FH Theory

The FH formulation employing the effective plastic strain (6) with exponent $\mu = 1$ is used in the body of the paper because it captures the size-dependence with more fidelity than the more widely used version with $\mu = 2$. The original studies of Fleck and Hutchinson [14, 15] and others favored $\mu = 2$ over $\mu = 1$ primarily on mathematical grounds and not on the basis of physical considerations. The following difficulties emerge when using the choice $\mu = 2$. A length parameter $\ell_{FH} = 12 \mu m$ fits the data for the foil with grain size, $g = 27 \mu m$, and thickness, $h = 50 \mu m$. However, the behavior predicted for the other two thicknesses is poorly represented and, in particular, the bending moment of the thinnest foil is substantially overestimated. Conversely, if ℓ_{FH} had been chosen to fit the data for the thinnest foil, the strengthening for the thicker foils would be underestimated. As already noted, this inadequacy arises because, when the gradient contribution is relatively small, the version with $\mu = 2$ predicts strength increases on the order of $(\ell \varepsilon_P^*)^2$, while the NG formulation and the FH version with $\mu = 1$ both predict strengthening of order $\ell \varepsilon_P^*$.

References

[1] Stelmashenko N A, Walls M G, Brown L M, Minman Y V. Acta Met. Mater. 1993; 41: 2855.

[2] Ma Q, Clarke D R. J. Mater. Res. 1995; 10: 853.

[3] Swadener J G, George E P, Pharr G M. J. Mech. Phys. Solids 2002; 50: 681.

[4] Nix W D, Gao H. J. Mech. Phys. Solids 1998; 46: 411.

[5] Begley M R, Hutchinson J W. J. Mech. Phys. Solids 1998; 46: 2049.

[6] Huang Y, Xue Z, Gao H, Nix W D, Xia Z C. J. Mater. Res. 2000; 15: 1786.

[7] Fleck N A, Muller G M, Ashby M F, Hutchinson J W. Acta Met. Mater 1994; 42: 475.

[8] Stolken J S, Evans A G. Acta Mater. 1998; 46: 5109.

[9] Xiang Y, Vlassak J J. Acta Mater. 2006; 54: 5449.

[10] Petch N J J, Iron Steel Inst. 1953; 174: 25.

[11] Lloyd D J. Int. Mater. Rev. 1994; 39: 1.

[12] Ashby M F. Phil. Mag. 1970; 21: 399.

[13] Aifantis E C. J. Engng. Mater. Tech. 1984; 106: 326.

[14] Fleck N A, Hutchinson J W. Adv. Appl. Mech. 1997; 33: 295.

[15] Fleck N A, Hutchinson J W. J. Mech. Phys. Solids 2001; 49: 2245.

[16] Gao H, Huang Y, Nix W D, Hutchinson J W. J. Mech. Phys. Solids 1999;47: 1239.

[17] Huang Y, Gao H, Nix W D, Hutchinson J W. J. Mech. Phys. Solids 2000;48: 99.

[18] Huang Y, Qu S, Hwang K C, Li M, Gao H. Int. J. Plasticity 2004; 20: 753.

[19] Zhu T T, Bushby A J, Dunstan D J. J. Mech. Phys. Solids, in press.

[20] Gudmundson, P. J. Mech. Phys. Solids 2004; 52: 1379.

[21] Aifantis K, Soer W, De Hosson J , Willis J R. Acta Mater. 2006; 54: 5077.

[22] Gurtin M E. Int. J. Plasticity 2003; 19: 47.

[23] Gurtin, M E. J. Mech. Phys. Solids 2002; 50: 5.

[24] Evers L P, Brekelmans W A M, Geers M G D. J. Mech. Phys. Solids 2004; 52: 2379.

[25] Cleveringa H H M, van der Giessen E, Needleman A. Acta Mater. 1997; 45: 3163.

[26] Mughrabi H. Mat. Sci. Engng. A 2004; 387-389: 209.

[27] Bassani J L. J. Mech. Phys. Solids 2001; 49: 1983.

[28] Voyiadjis G Z, Abu Al-Rub R K. Int. J. Solids Struct. 2005; 42: 3998.

[29] Hemker K J, Mendis B G, Eberl C. Mater. Sci. Engng. A 2008; 483-484: 727.
(Indentation data supplied by private communication).

Table I—Specification of uniaxial data

Grain Size g (μm)	σ_Y (MPa)	λ	N
6	66	0.078	0.49
14	57.5	0.043	0.60
27	42	0.050	0.57
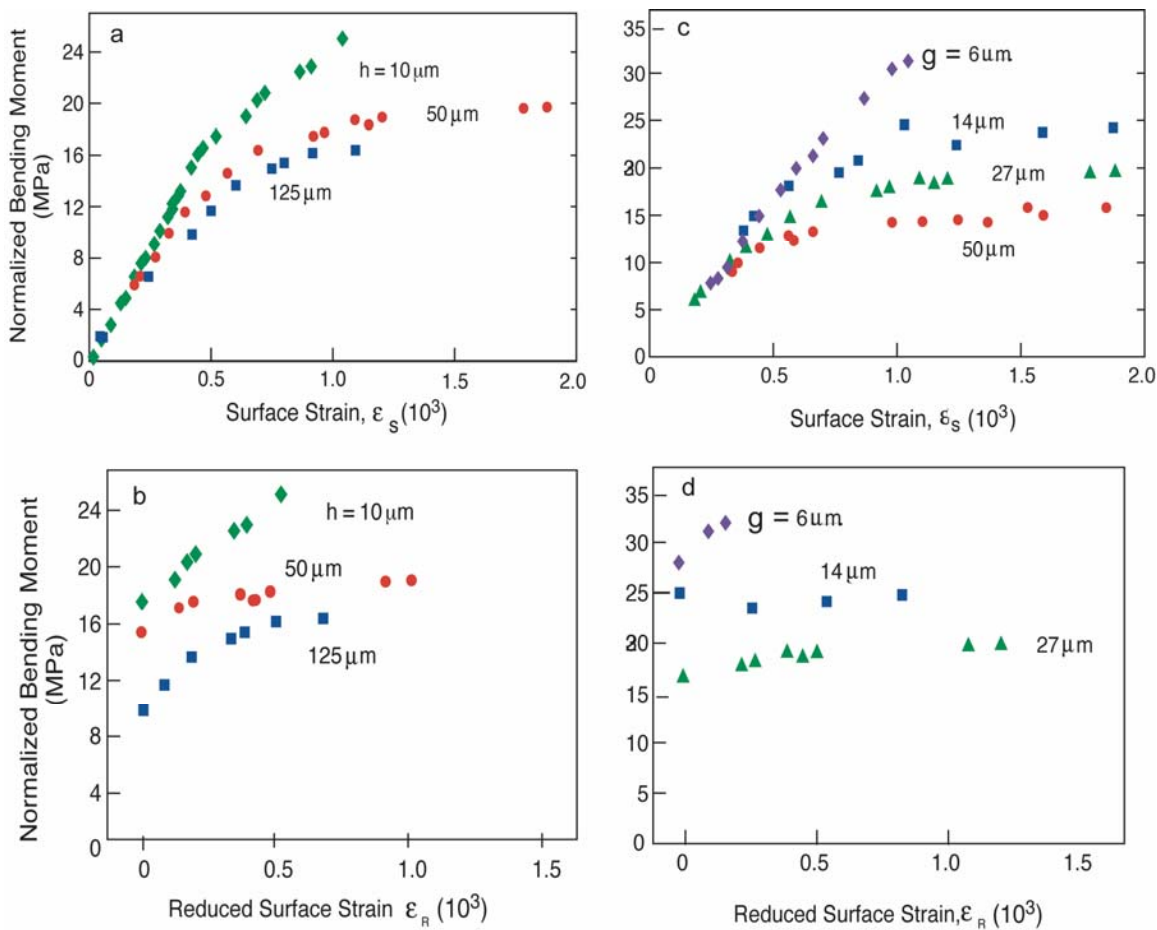
Fig. 1 Normalized moment, M/h^2 as a function of surface strain, ε_S, and reduced surface strain, ε_R, for Ni foils. (a) and (b) For three thicknesses with grain size $g = 27\mu m$. (c) and (d) For three grain sizes with foil thickness $h = 50\mu m$ [19].
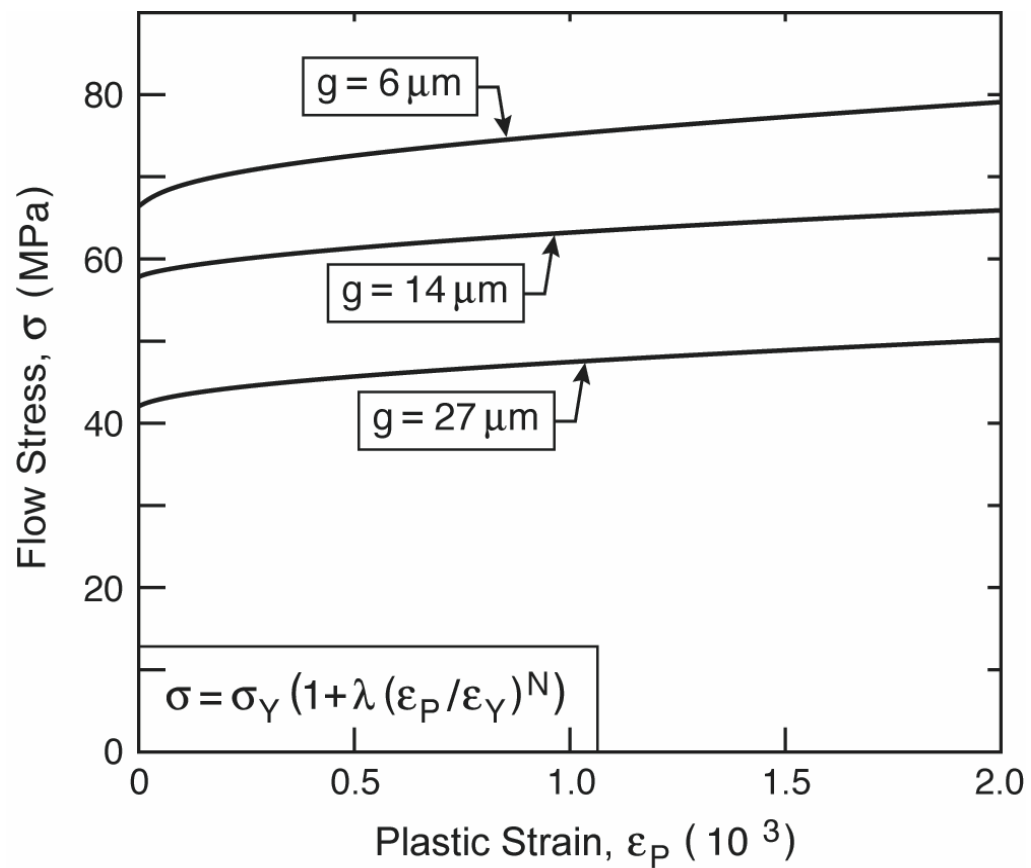
Fig. 2 Tensile stress-plastic strain data for foils with three grain sizes [19].
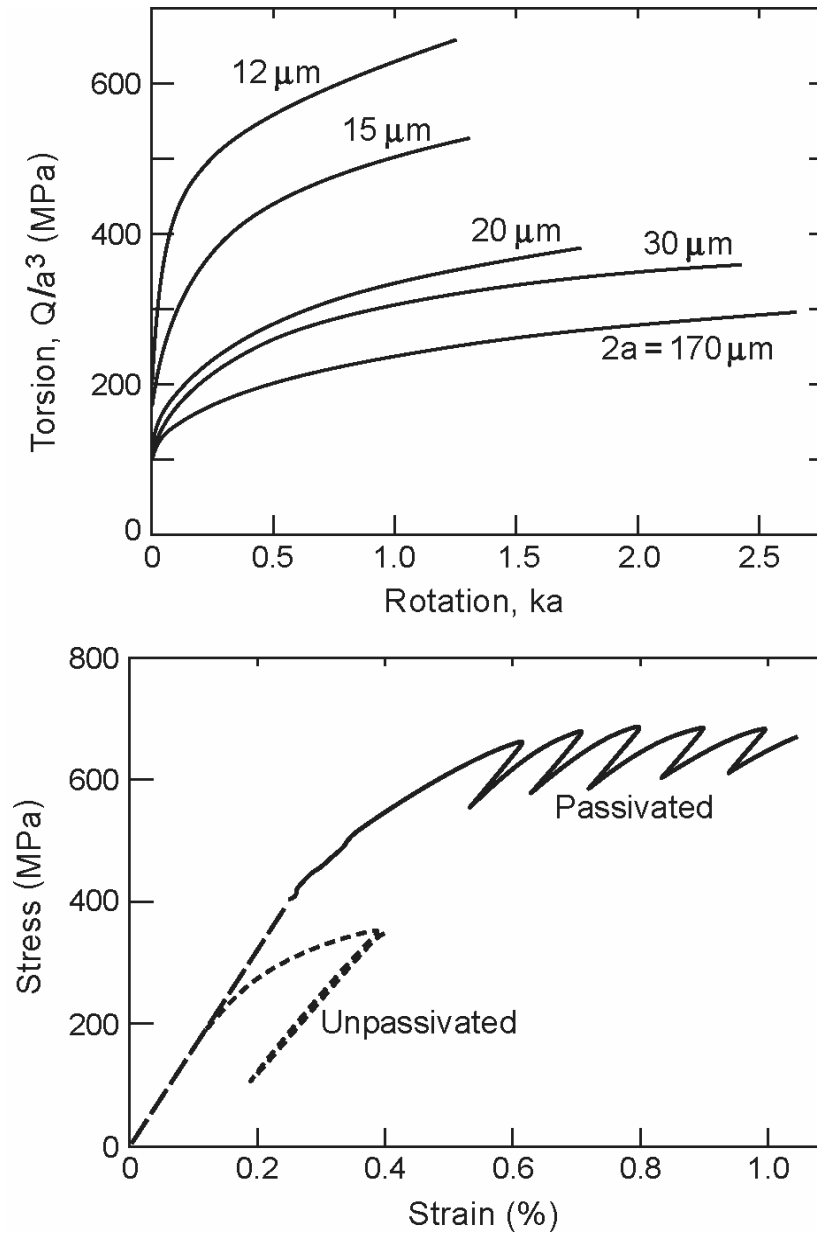
Fig. 3 Top plot: Normalized torque versus normalized rotation for copper wires of various radius, *a* [7]. This normalization would collapse the data to a single curve if there were no material size-dependence. Lower plot: Tensile stress-strain data for thin copper films ($0.34\mu m$) that are unpassivated and passivated (on one surface) [9].
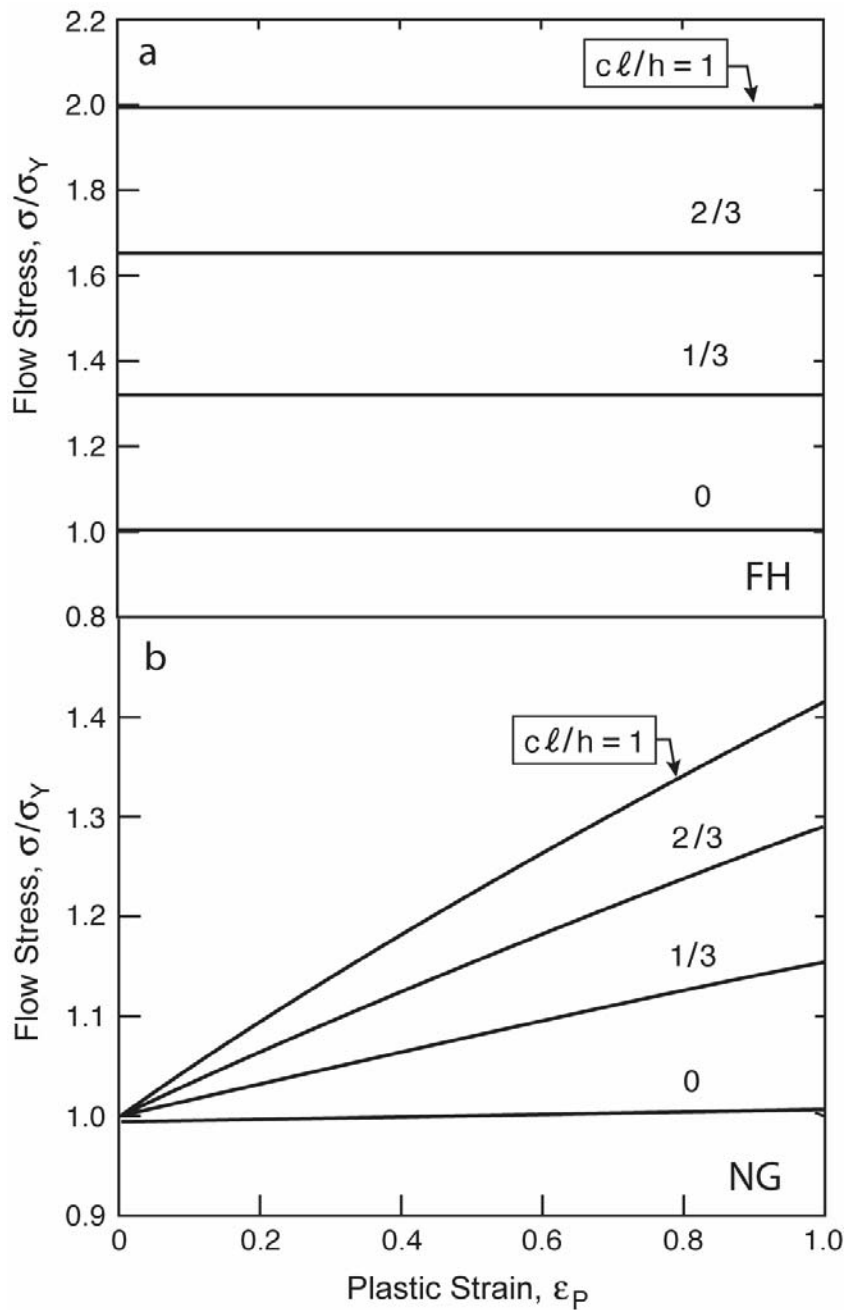
Fig. 4 Trends of average flow stress as a function of average plastic strain, ε_P, for an object of size h subject to an average gradient of plastic strain $\varepsilon_P^* = c\varepsilon_P / h$. (a) FH and (b) NG formulations.
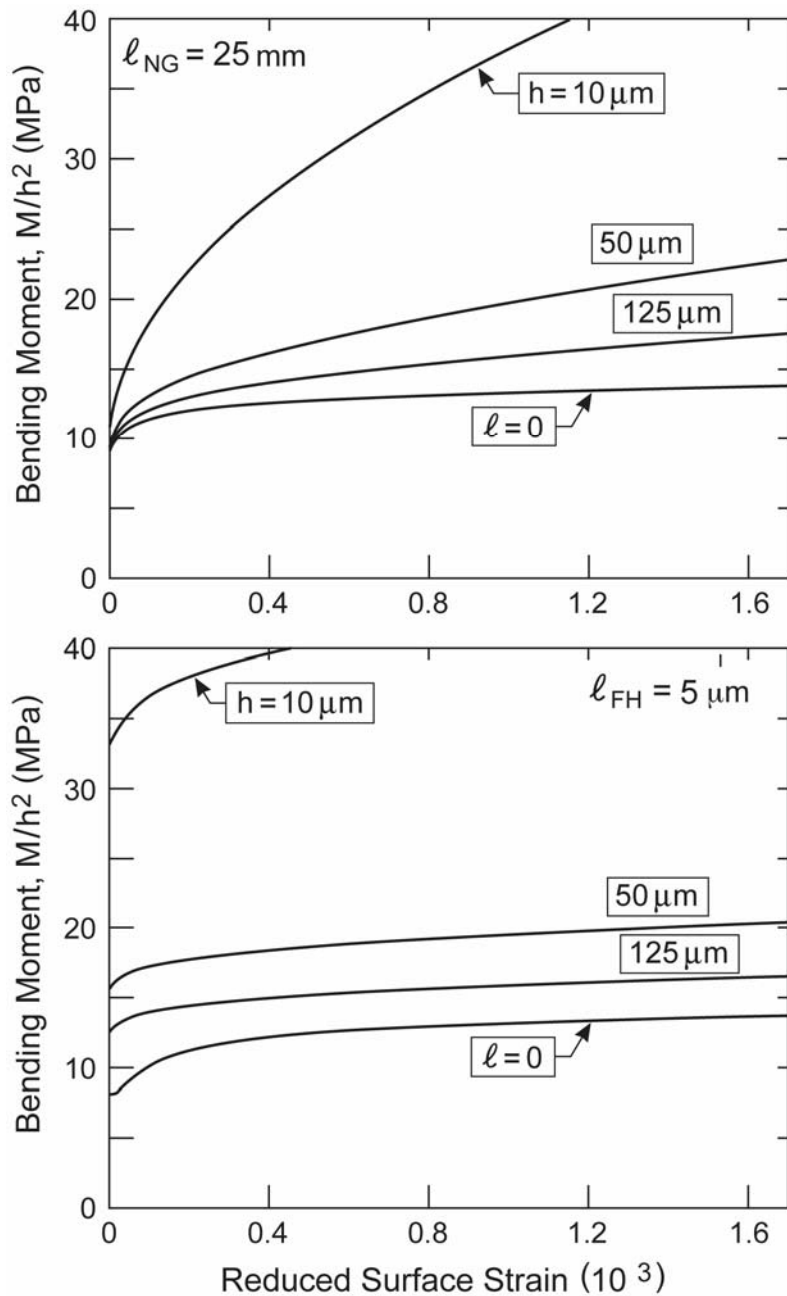
Fig. 5. Predictions of the normalized bending moment, M/h^2 as a function of the reduced surface strain, $\varepsilon_S^R = \varepsilon_S - 6(1-v^2)M/(Eh^2)$, for three foil thicknesses all having grain size, $g = 27\mu m$. In each case the length parameter was chosen to fit the experimental data in Fig. 1 for $h = 50\mu m$. Upper plot: Predictions of NG theory with $\ell_{NG} = 25mm$. Lower plot: Predictions of FH theory ($\mu = 1$) with $\ell_{FH} = 5\mu m$.
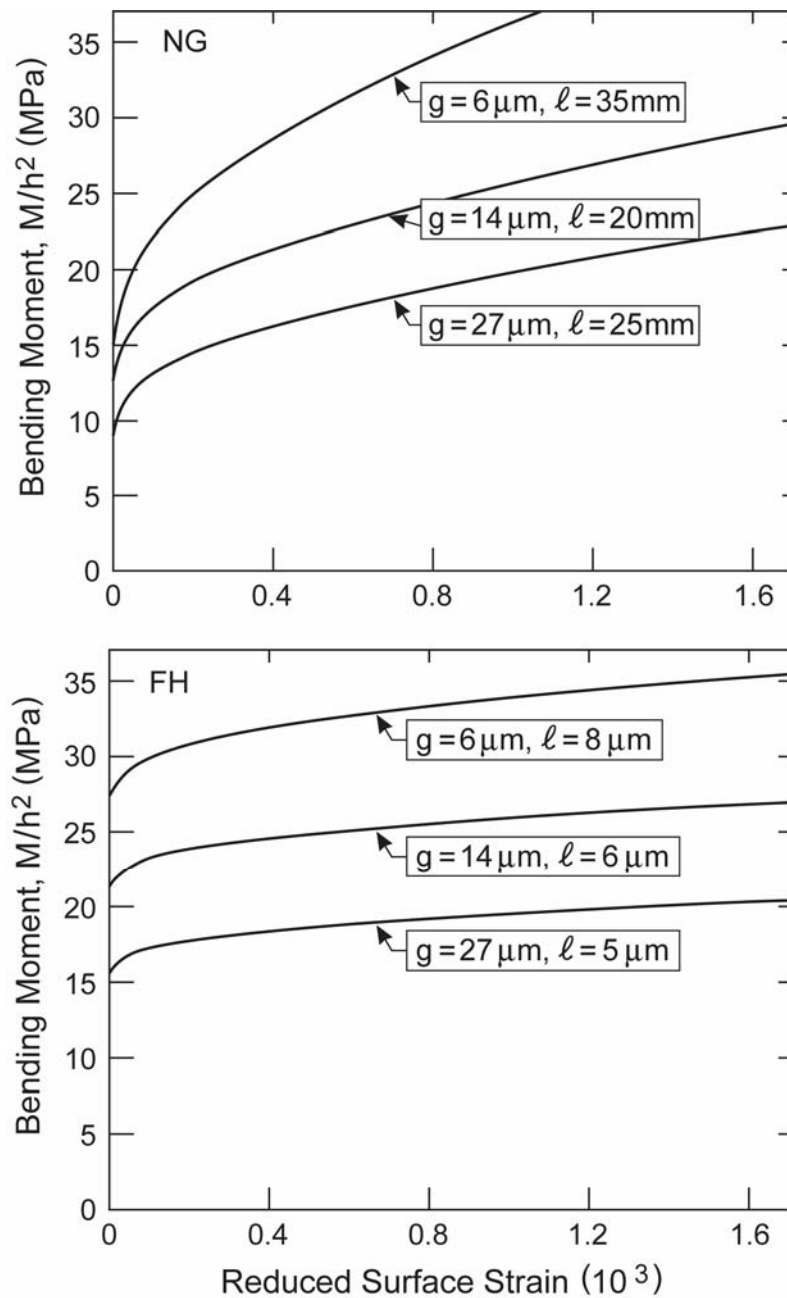
Fig. 6 Predictions of the normalized bending moment, M/h^2, as a function of the reduced surface strain for three grain sizes all for foils having thickness $h = 50\mu m$. In each case, the length parameter was chosen independently to give the best fit to the data in Fig. 1. Upper figure: NG theory and Lower figure: FH theory.
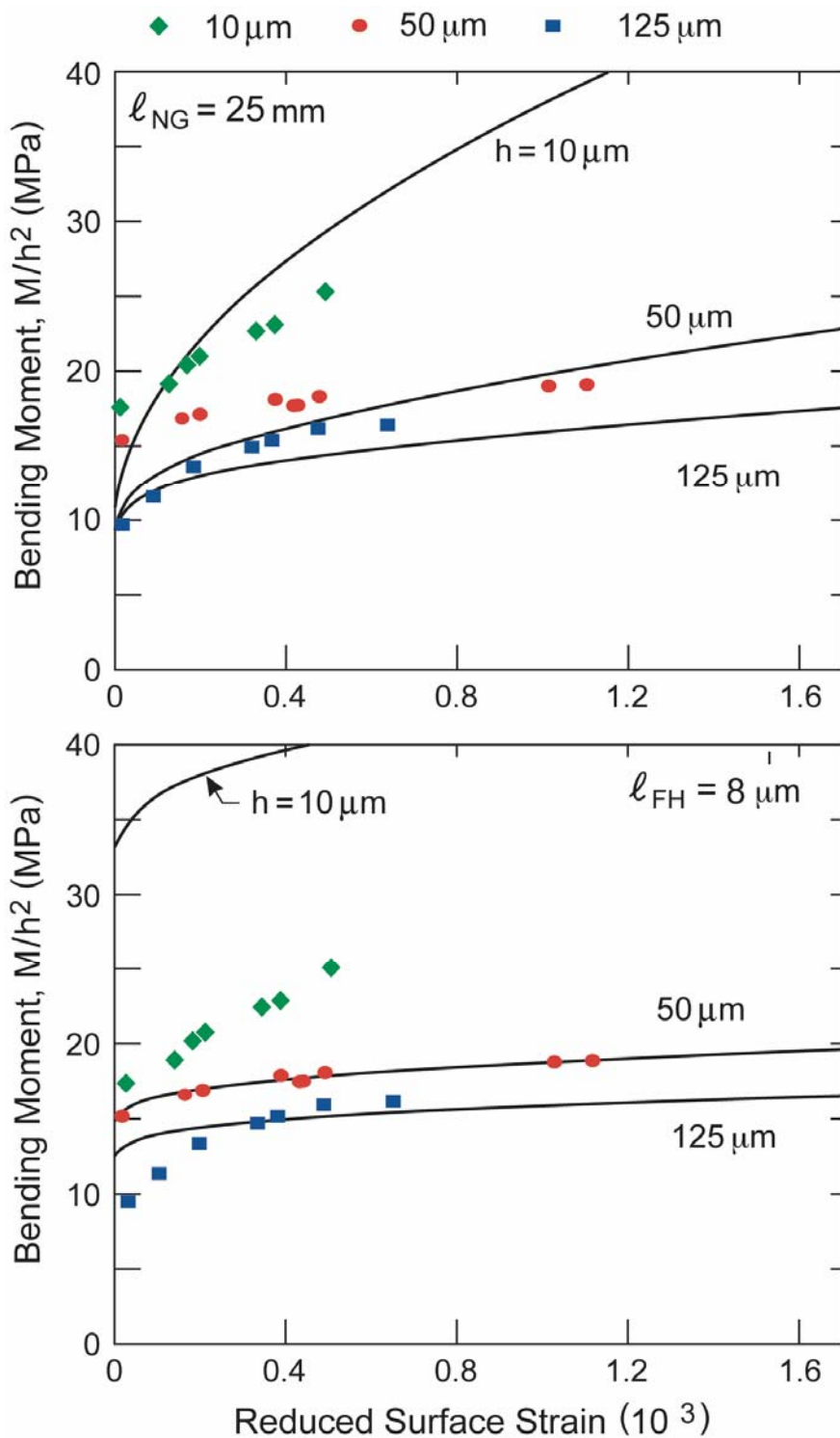
Fig. 7 Comparison between experimental data from Fig. 1 for three foils with grain size, $g = 27\,\mu m$, and theoretical predictions from Fig. 5. Upper plot: NG; Lower plot FH.
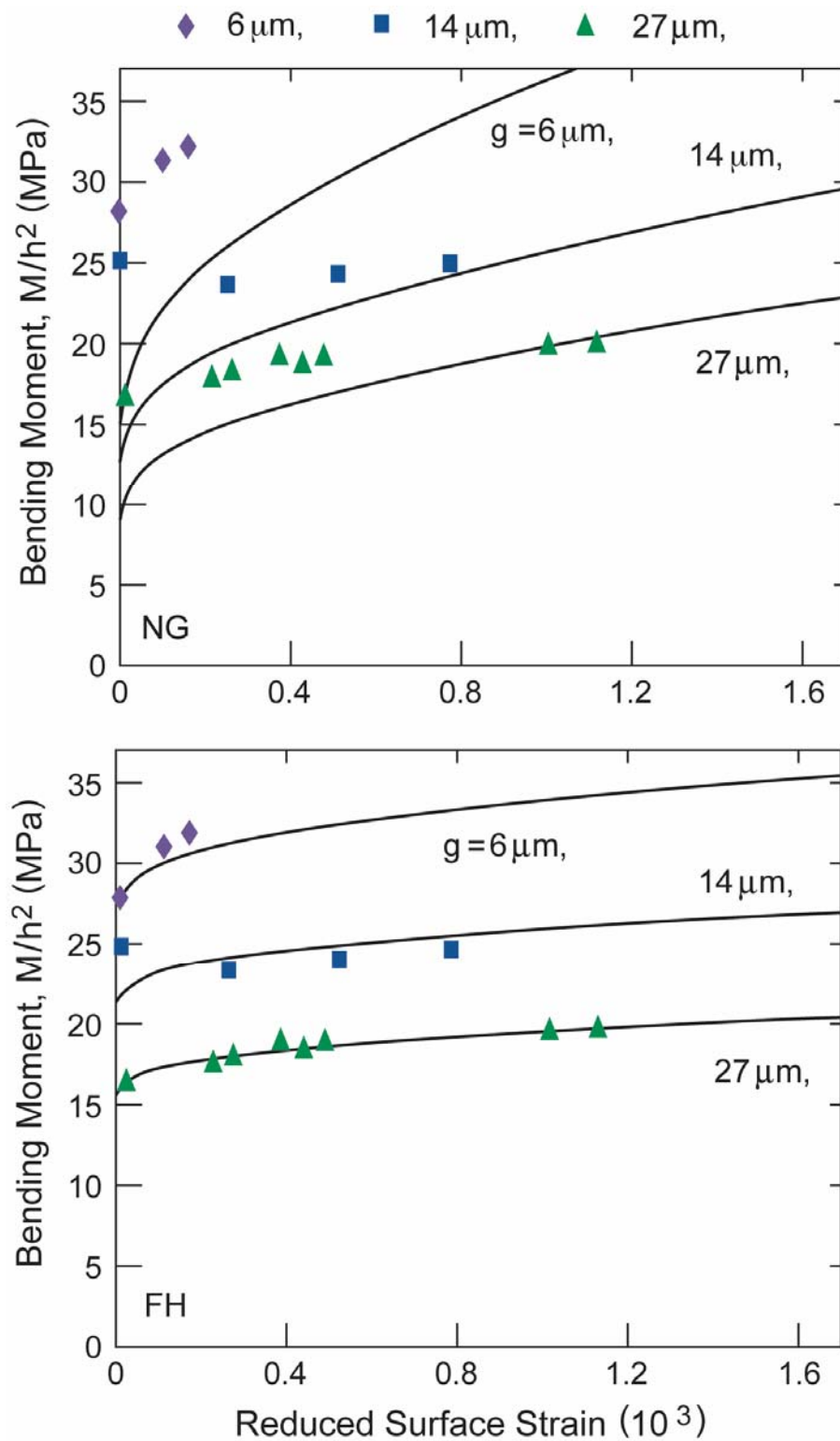
Fig. 8 Comparison between experimental data from Fig. 1 for three grain sizes for foils with thickness, $h = 50\mu m$, and theoretical predictions from Fig. 6. Upper plot: NG; Lower plot FH.
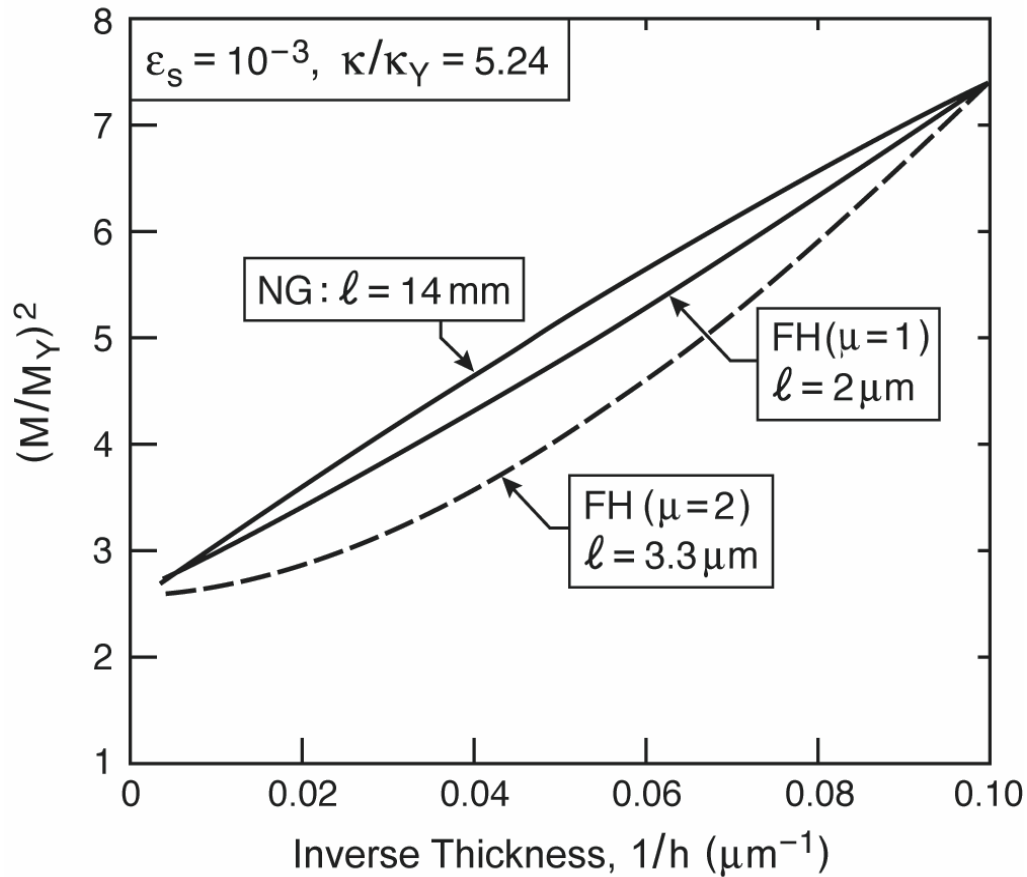
Fig. 9 A plot illustrating the predicted size-dependence in bending with emphasis on the approach to the size-independent limit ($1/h \rightarrow 0$). This example was computed using uniaxial stress-strain data for the foil with $g = 27\,\mu m$ at imposed curvature $\kappa/\kappa_Y = 5.24$ ($\varepsilon_S = 10^{-3}$) with size-independent limit, $(M/M_Y)^2 = 2.6$. The length parameter for each theory was calibrated such the predictions coincide with $(M/M_Y)^2 = 7.4$ at $\kappa/\kappa_Y = 5.24$ for foils having $h = 10\,\mu m$ (at the right hand side of the plot).
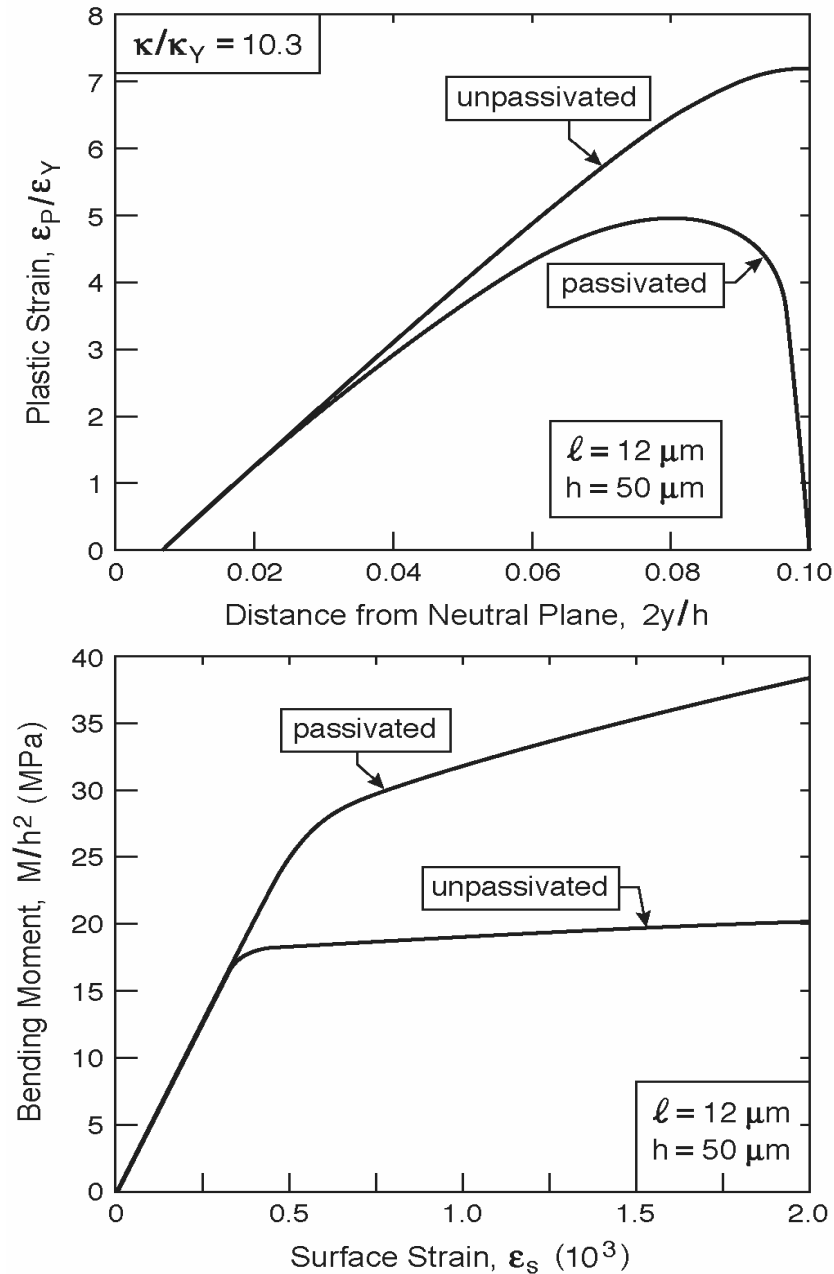
Fig. 10 The role of passivated surfaces, as predicted by FH theory with $\mu = 2$. The passivation layer blocks dislocations such that $\varepsilon_p = 0$ at the surfaces, but is otherwise assumed to have negligible thickness. There is no constraint on ε_p for the unpassivated surface, as in all the previous figures. Lower figure: Normalized moment versus surface strain, ε_s. Upper figure: Distribution of ε_p across the upper half of the foil at $\kappa / \kappa_Y = 10.3$.
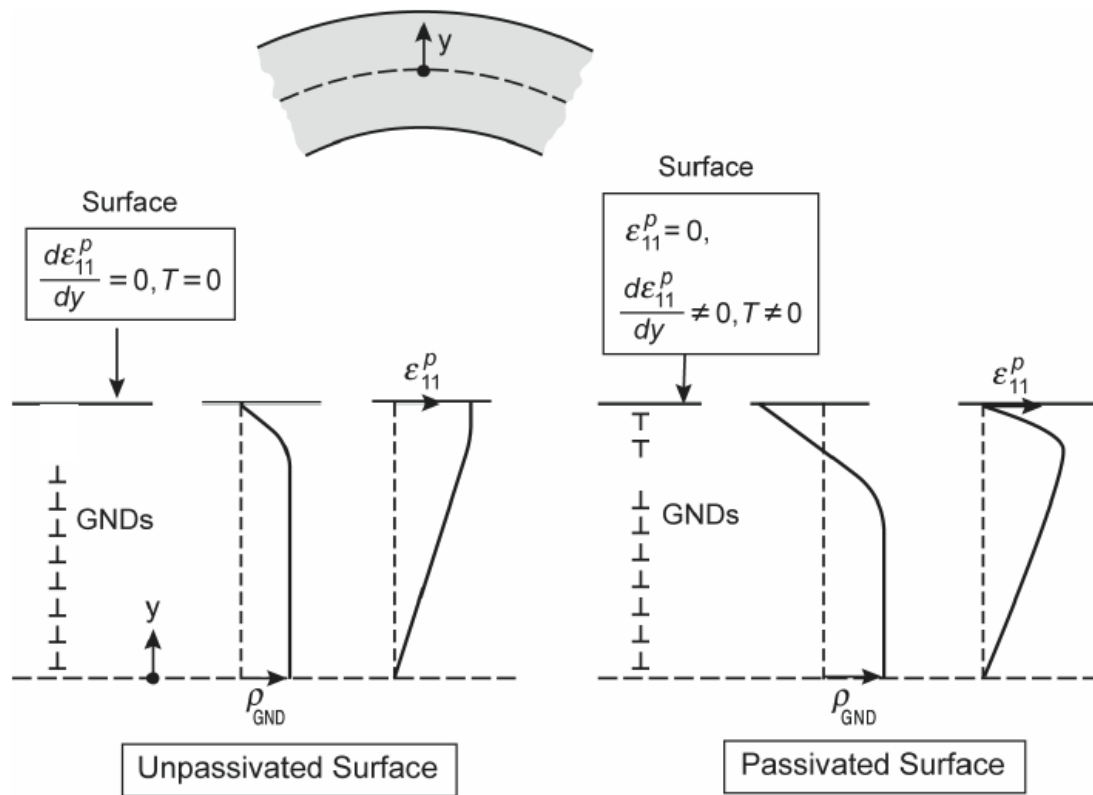
Fig. 11 Schematic of dislocation and plastic strain distributions in bending for foils with unpassivated and passivated surfaces.
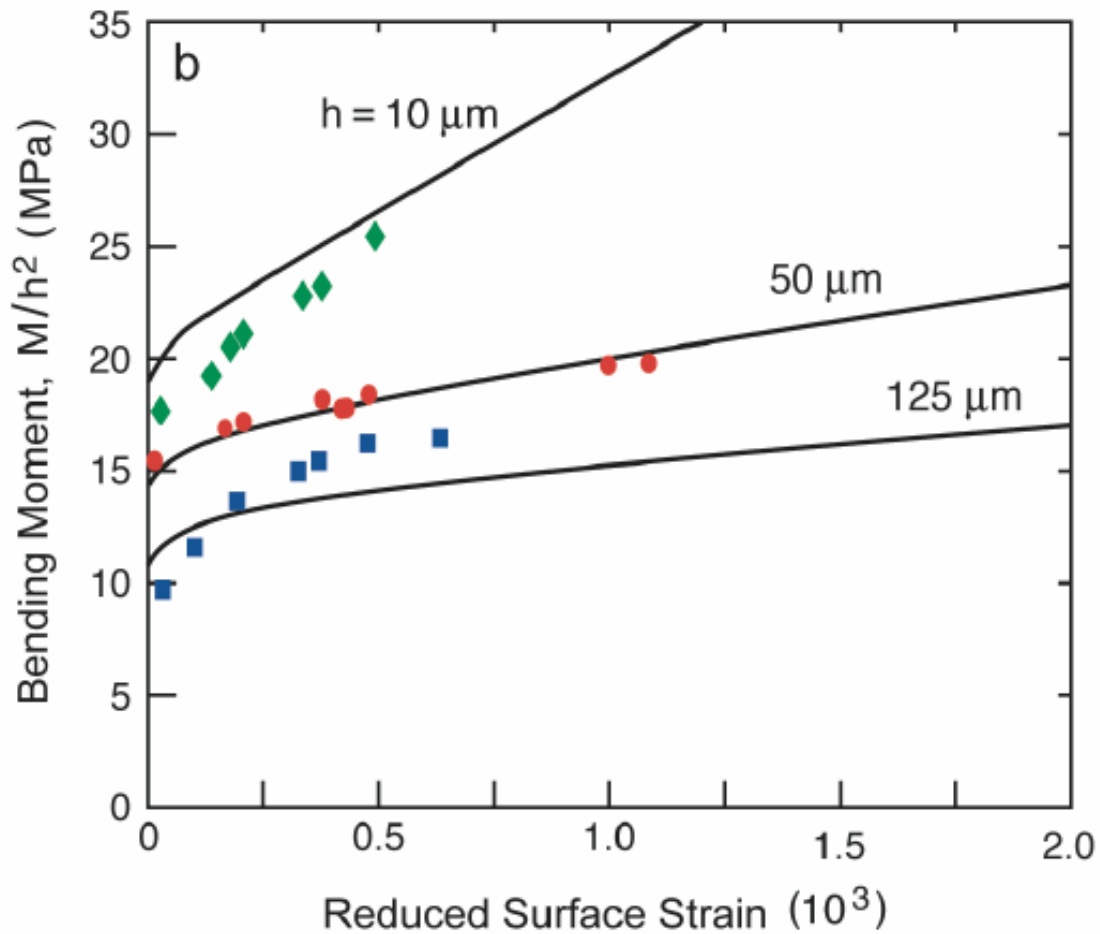
Fig. 12 Prediction from FH theory incorporating a strain-dependent length parameter, $\ell(\varepsilon_P) = \ell_0 \left(1 + a_0(\varepsilon_P / \varepsilon_Y)\right)$. The values, $\ell_0 = 2\mu m$ and $a_0 = 0.2$, were chosen to both fit the data for the foil with $h = 50\mu m$ at $\varepsilon_R = 10^{-3}$ and to better represent the initial increase in strength and hardening of the foil with $h = 10\mu m$.
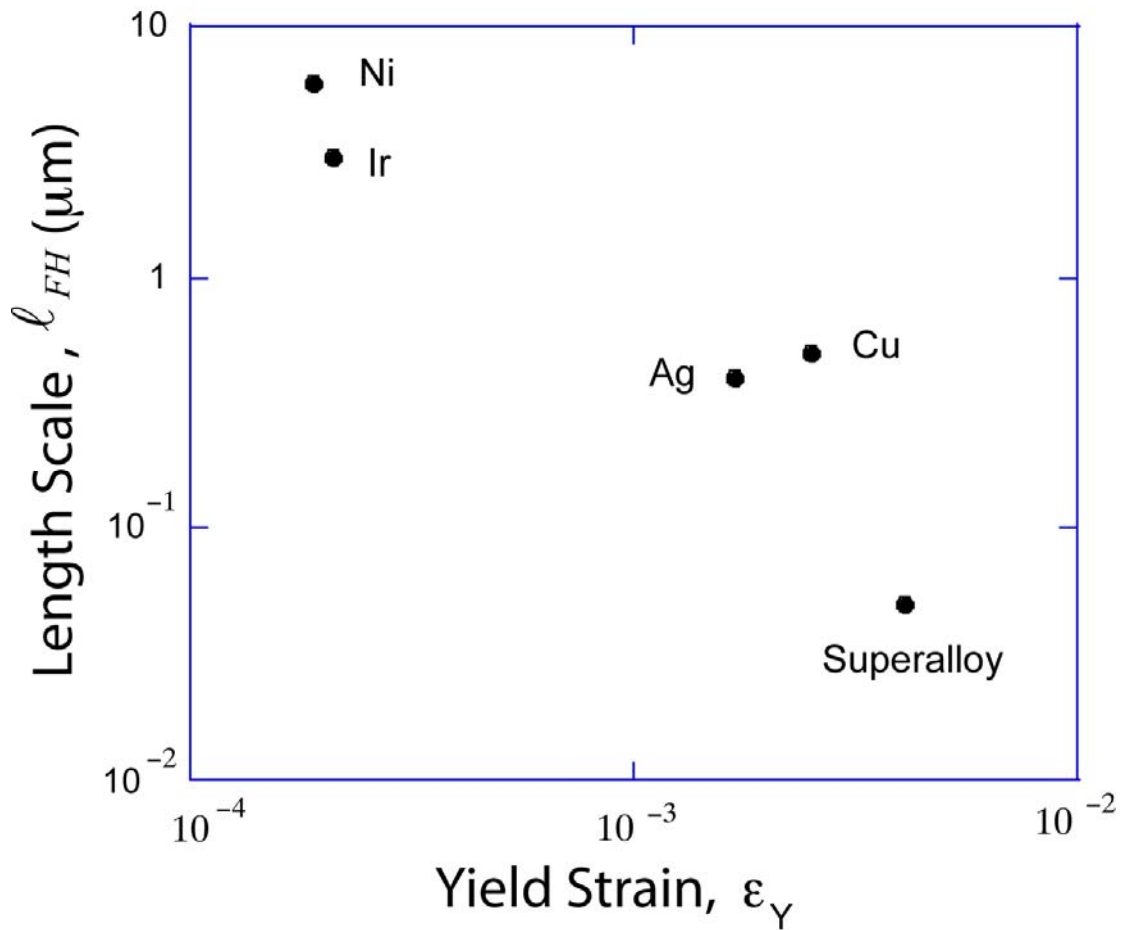
Fig. 13 A plot of the length scale, ℓ_{FH}, against the size-independent tensile yield strain. The length scale was ascertained using FH theory using indentation data for Iridium [3], a superalloy [29], copper [5], and silver [2], together with the present finding for bending of Ni foils.