Validation of a phenomenological strain-gradient plasticity theory

D.J. Dunstan

School of Physics and Astronomy, Queen Mary University of London,

Address: London, E1 4NS, England

E-mail: d.dunstan@qmul.ac.uk
Validation of a phenomenological strain-gradient plasticity theory

Strain-gradient plasticity theories have been developed to account for the size effect in small-scale plasticity in metals. However, they remain of limited use in engineering, for example in standards for nanoindentation, because of their phenomenological nature. In particular, a key parameter, the characteristic length, can only be determined by fitting to experiment. Here it is shown that the characteristic length in one such theory derives directly from known quantities through fundamental dislocation physics. This explains and validates the theory for use in engineering.

Keywords: plasticity of metals; strengthening mechanisms; strained layers; dislocations; strain-gradient theory; critical thickness theory.

The increase in strength (the size effect) when dislocation-mediated plasticity is restricted to small volumes has been extensively documented experimentally over the past 60 years [see, e.g., 1–13]. It is an important effect in many technologies from metallurgy to semiconductors, yet it is not fully understood [12, 14]. In micromechanics, many loading conditions impose a plastic strain gradient, and so theories in which the strain gradient plays a central role have been developed [3–6, 15–19]. In contrast, in semiconductor technology, Matthews critical thickness theory has been largely accepted to explain and predict the effect in terms simply of the size – stronger when smaller [20–22]. The strain-gradient theories have not been comprehensively embraced [23], because of ambiguities about the underlying physics and about the parameters – in particular, the characteristic length – which enter into these theories. One consequence is that there are no satisfactory international standards for comparing nanoindentation data, in which the
size effect plays an important role, with macroscopic indentation data. Here it is shown
that the Fleck-Hutchinson strain-gradient theory [4, 17–19, 23] follows mathematically
and physically directly from critical thickness theory [20–22]. The strain-gradient theory
fits experiment well, but with the characteristic length as a free fitting parameter. This
phenomenological parameter is here derived from known physical quantities via critical
thickness theory. The derivation and the associated re-interpretation validate the strain-
gradient theory for use in practical engineering contexts, as an approximation that
expresses a non-local property as a local property.

Increases in strength (the size effect) due to boundaries imposed on dislocation-
mediated plasticity on scales up to tens of microns have been presented for
nanoindentation [3, 5], thin wires in torsion [4, 9, 10], thin foils in bending [6, 8], and for a
large variety of still smaller structures down to sub-micron sizes mostly created by
focused ion-beam (FIB) milling [e.g. 7, 11, 13]. Microstructural constraints giving rise to
the size effect include sub-grain boundaries [2] and grain boundaries (the Hall-Petch
effect) [1, 12]. Pseudomorphic (strained-layer) heteroepitaxial crystal growth is another
key example [20–22]. In many of these situations, plastic strain gradients are necessarily
or optionally present, and there is widespread agreement that in such situations the size
effect can be attributed to the strain gradient.

In formal continuum mechanics, to set up a strain-gradient plasticity theory
(SGP), the stress is not only a function of plastic strain \( \varepsilon_P \), but also a function of its spatial
gradient \( \ell \varepsilon'_P = \ell d\varepsilon_P / ds \) where \( s \) is position and the characteristic length \( \ell \) is introduced
to give a dimensionless quantity [16–19]. Where a physical interpretation is called for,
appeal is made to the geometrically-necessary dislocations (GNDs) [3] which in a
crystalline material are necessarily associated with plastic strain gradients [15]. Values of
\( \ell \) are found from fitting to experiment (see Fig.1). The major problem for such strain-
gradient theories is to give a reasonable physical interpretation of the values of \( \ell \) that
result. There have been many proposals. See [24] for a recent discussion and a new
proposal.

Evans and Hutchinson [23] gave an appraisal of SGP theories, for brevity
confined to the Nix-Gao (NG) theory [3] and the Fleck-Hutchinson (FH) theory [4, 17,
19]. These two theories illustrate adequately both the successes of SGP theories in
general, and their difficulties. The successes lie in the good fits to experimental data that
these theories give. The major difficulty is that, fitting to experimental datasets for soft
metals, the NG theory gives characteristic lengths \( \ell_{NG} \sim 25\text{mm} \), and the FH theory gives
\( \ell_{FH} \sim 5\mu\text{m} \). Neither is characteristic of any length scale experimentally observed in the
specimens, whether structural or microstructural. For this reason, and because of the lack
of any explicit connection between the theories and dislocation dynamics, Evans and
Hutchinson noted that strain-gradient theories have not been comprehensively embraced
[23].

Here, the FH characteristic length is derived from critical thickness theory. This
reveals a previously unsuspected link between the two theories. In particular, it provides
the explicit connection between the FH theory and the physics of dislocation dynamics
that was previously lacking. It thereby validates the use of the FH theory for prediction in
engineering applications (with due attention to the approximations revealed in it).
It is not necessary to use a full derivation of strain-gradient theory. We take Evans and Hutchinson [23] as a starting-point. They define an effective stress $\sigma$ which is a function of the yield stress and the plastic strain, $\sigma = \sigma_Y f(\varepsilon_P)$. For the FH theory, they state as a premise that the plastic work per unit volume may be written as

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

(1)

The upper integral limit $E_P$ brings in the effect of the strain gradient $\varepsilon'_P$ by the definition

$$ E_P = \varepsilon_P + \ell_{FH} \varepsilon'_P $$

(2)

This is a specific form of the generalized effective plastic strain $E_p$ [19]. Consider an object of size $h$, average plastic strain $\bar{\varepsilon}_P$ and average plastic strain gradient $\bar{\varepsilon}'_P = c \bar{\varepsilon}_P / h$ with $c \sim 1$, and with perfect plasticity, $f(\varepsilon_P) = 1$. From equation (1), the average flow stress is

$$ \bar{\sigma} = \sigma_Y \left( 1 + \frac{c \ell_{FH}}{h} \right) $$

(3)

This is equation (11) of Ref.23. Note that the strengthening is independent of $\varepsilon_P$. The strain gradient increases the yield strength but not the rate of strain-hardening. Using $\ell_{FH} = 5 \mu m$ and adding a work-hardening term, Evans and Hutchinson [23] obtain excellent fits to the data of Ehrler et al. [8] for nickel foils.

We apply equation (3) to simple and very well understood examples of the size effect. These are the plastic relaxation of non-lattice-matched epitaxial strained-layer
structures grown above their critical thicknesses. Growth is in the $z$ direction to a
thickness $h$ above the substrate at $z = 0$. At typical growth temperatures of 600°C for
GaAs-based structures (more than half the melting-point) the intrinsic yield strength is
very low. The ability to support elastic strains of 0.01 and more at thicknesses of tens of
nm comes from the size effect. In good-quality growth, there is little or no evidence of
work-hardening and the material may be taken to be perfectly plastic. Matthews critical
thickness theory [20–22] gives the critical thickness $h_C$ at which misfit dislocations
(GNDs) may form at $z = 0$ to relieve the elastic strain in a simple layer with misfit strain
$\varepsilon_0$. The result, for our purposes here, is best expressed by the geometrical version of
Matthew’s theory [25, 26], as $h_C \sim b/\varepsilon_0$ where $b$ is the relevant (in-plane) component of
the Burgers vector of the misfit dislocations (the GNDs). This version agrees well with
experiment. Moreover, it omits unnecessary detail which is specific to single-crystal
cubic semiconductors and also it omits the ill-defined parameters, the inner and outer cut-
off radii, that appear in the calculation of the dislocation self-energy. The elastic strain $\varepsilon_E$
$= \varepsilon_0$ for $h < h_C$ and the plastic relaxation at greater thicknesses gives $\varepsilon_E \sim b/h$ for $h > h_C$.
The condition for plastic relaxation may be written in terms of the strain-thickness
product as $\varepsilon_E h \sim b$. The theory is readily generalised to more complicated structures
(graded layers with $\varepsilon_0 = g\varepsilon$, multilayers and superlattices) by considering the strain-
thickness integral of $\varepsilon_E(z)dz$ over the thickness and introducing plastic relaxation during
growth as necessary to limit the integral to the value $b$ [27]. Any intrinsic or bulk strength
simply adds to this size-effect strength. In all cases the size effect is due to the energy
required to create the length of GND needed to accommodate the misfit.
For significant plastic deformation (stress relaxation) when the initial dislocation density is low, dislocation multiplication must take place – sources must operate. Beanland showed that this requires a much greater thickness, \( h_R \sim 5 \ h_c \) for simple layers [28, 29]. In this case, the energy required to create the GNDs is small compared with the energy dissipated in source operation. Then the strain-thickness product or integral during plastic deformation is \( \sim 5b \) for \( h > h_R \). Experimentally, these predictions of the theory have been confirmed extensively in simple layers, graded layers and in more complicated structures [30–32]. The theory also predicts the spatial distribution of GNDs and of \( \varepsilon_P \) [32], confirmed by discrete dislocation dynamics simulation [33].

We calculate the average plastic strain, the average plastic strain gradient, the average stress, and the constant \( c \) for three standard epitaxial structures (Table I). For the simple constant-composition strained layer with misfit strain \( \varepsilon_0 \) grown above its relaxation critical thickness the plastic strain \( \varepsilon_P(z) \) throughout the thickness of the layer is constant and so this is also the average, \( \bar{\varepsilon}_P = \varepsilon_P \). The average stress is \( \bar{\sigma} = M \varepsilon_E \) where \( M \) is the relevant elastic modulus. The plastic strain gradient is ideally infinite at the substrate – layer interface and zero elsewhere, but the average comes just from the change of plastic strain, from 0 at the substrate at \( z = 0 \) to \( \varepsilon_P \) at the top at \( z = h \). The constant \( c = 1 \) in this case by definition. Then the average stress (Table I), with a bulk yield stress \( \sigma_Y \) added, may be set equal to the average stress predicted by the FH theory in equation (3) giving,
where \( \varepsilon_Y \) is the yield strain.

In linearly-graded layers, with the misfit increasing as \( g_z \), the strain-thickness integral without plastic relaxation is \( \frac{1}{2}gh^2 \), and the critical thickness \( h_R \) is given by setting this equal to \( 5b \). When growth continues above \( h_R \), the lower material relaxes completely.

A top layer of thickness \( h_R \) has a uniform \( \varepsilon_P \) and stress increasing linearly with the slope \( Mg \). We consider first a thin structure with growth to a thickness \( h = h_R + \delta \) (\( \delta \) small) giving constant plastic strain throughout the grade, except for the thin layer of thickness \( \delta h \) at the bottom (Table I) which we ignore. Again \( c = 1 \). The stress increases linearly so the average stress is half the surface stress (Table I). Again adding a bulk yield stress \( \sigma_Y \) and equating the average stress with the average stress of equation (3) we have

\[
\bar{\sigma} = \sigma_Y \left(1 + \frac{c\ell_{FH}}{h_R}\right) = \sigma_Y + \frac{1}{2}Mgh_R
\]

\[
\ell_{FH} = \frac{\frac{1}{2}Mgh_R^2}{\sigma_Y} = \frac{5b}{\varepsilon_Y}
\]

Graded-layer growth to a much greater thickness \( h \gg h_R \) gives complete plastic relaxation to \( \varepsilon_E = 0, \varepsilon_P = g_z \) throughout the layer except for a thin region at the top of thickness \( h_R \) where \( \varepsilon_P \) is constant and the elastic strain \( \varepsilon_E \) rises from 0 to \( gh_R \) [27, 32].

Neglecting the thin region at the top, the average plastic strain is \( \frac{1}{2} gh \), while the average plastic strain gradient is just \( g \), so that here \( c = 2 \). The stress is zero except in the thin
region at the top where it rises from zero to \( Mgh_R \), so the stress-thickness integral is constant at \( \frac{1}{2}Mgh_R \) and the average stress is obtained by multiplying by \( h_R/h \). Again adding a bulk strength \( \sigma_Y \) and equating the average stress with the average stress of equation (3) we have,

\[
\bar{\sigma} = \sigma_Y \left( 1 + \frac{c \ell_{FH}}{h} \right) = \sigma_Y + \frac{\frac{1}{2}Mgh_C^2}{h}
\]

\[\ell_{FH} = \frac{\frac{1}{2}Mgh_C^2}{c \sigma_Y} = \frac{5b}{2 \varepsilon_Y} \tag{6}\]

All three examples, equations (4-6), give similar results, varying only because of the factor \( c \), so we conclude that

\[\ell_{FH} = \frac{5b}{c \varepsilon_Y} \tag{7}\]

The problem of a linearly-graded layer maps perfectly onto half of the problem of a beam in bending, from the neutral plane to either free surface [33]. Taking typical numerical values for pure nickel and other soft metals, \( M \sim 100 \) GPa, \( b \sim 0.25 \) nm and yield strengths about 20 MPa, gives \( \ell_{FH} = 3.125 \mu m \) from equation (6). This is in good agreement with the results from empirical fits (Fig.1).

Evans and Hutchinson [23] give values (but not error bars) of \( \ell_{FH} \) obtained by fitting the FH theory to data from different authors for indentation of iridium, silver, copper and a superalloy, and to data for bending nickel foils. They note the inverse correlation between the values of \( \ell_{FH} \) and the yield strain \( \varepsilon_Y \) of the material (figure 1), as in equations (4-7). Their tentative interpretation is that \( \ell_{FH} \) represents the distance
moved by dislocations between e.g. cell walls or precipitates, which will be reduced as
\( \sigma_Y^{-1} \) in stronger materials. However, this interpretation overlooks the physical origin of
the size effect. Moreover, equation (7) *predicts* the absolute magnitudes of \( \ell_{FH} \) very well
(figure 1).

The presence of \( c \), the ratio of the peak value of \( \varepsilon_P \) to its average value, in the
denominator of equation (7) is interesting. Gradient theory fits DDD simulation results
better if the characteristic length is allowed to be a variable and to decrease with strain
[24]. The graded layers, equations (5, 6) show that \( c \) varies from 1 at low strain to 2 at
high strain, with a concomitant reduction of a factor of 2 in the characteristic length of
equation (7).

The phenomenological FH and similar strain-gradient theories express the
*outcomes* of the size effect accurately, but using a fitting parameter, the characteristic
length, which is not a true characteristic of the material. Evans and Hutchinson [23]
attribute equation (3) to the summation of the energy dissipation caused by the movement
of statistically-stored dislocations (SSDs) and that due to the movement of GNDs, the
second term.

Our interpretation of equation (3) is different. From figure 1 and equation (7), the
characteristic length is the Matthews critical thickness \( h_C \) or the relaxation critical
thickness \( h_R \) calculated using the elastic yield strain or flow stress of the material.
Equivalently, it is the thickness \( h \) at which the size effect doubles the strength of the
material. *Note that the \( \sigma_Y \) in the denominator of equation (7) permits rewriting equation
(3) as*
\[ \bar{\sigma} = \sigma_y + \frac{c\sigma_y \ell_{FH}}{h} = \sigma_y + \frac{5Mb}{h} \] (8)

so that the inverse dependence of \( \ell_{FH} \) on \( \sigma_y \) is cancelled by the prefactor \( \sigma_y \). This is a very clear indication that the size effect is independent of the phenomena determining the yield strength, such as dislocation and defect densities. The first term does indeed represent whatever dissipative mechanism is responsible for the strength of bulk material without a size effect, such as the movements of SSDs. The second term, however, in the case that source operation is not required (\( \varepsilon_E \sim b/h \)), represents the energy stored (not dissipated) by the creation of GND length – the Matthews model [20–22]. In the case that source operation is required (\( \varepsilon_E \sim 5b/h \)), and this is generally the case for significant plastic deformation, the second term represents mostly the energy dissipated by source operation under the \(~5\times\) greater stress required to operate sources within a restricted size compared with the stress required merely to create extra GND length [29, 31]. In this interpretation, it is clear that neither the presence of GNDs nor the presence of a plastic strain gradient are directly responsible for the increased strength when they are present. The increased strength arises from the energy required to create the GNDs or to operate sources.

In this context, it is interesting to observe that the Matthews theory (\( \varepsilon_E \sim b/h \)) for simple strained layers requires the presence of a substrate, for otherwise misfit dislocations have nowhere to exist. But given the need for dislocation multiplication, the need to operate sources, the relationship \( \varepsilon_E \sim 5b/h \) is independent of the presence or absence of a substrate, since two free surfaces with a separation \( h \) constrain the curvatures...
of dislocations in a source (to more than \( \sim h^{-1} \)) in much the same way as one free surface
and a strained-layer – substrate interface or neutral plane does, or indeed the two
interfaces of a capped layer. Consequently, equation (7) applies as well to a stand-alone
thin foil, wire or micropillar under uniaxial tension or compression as it does to an
epitaxial layer on a substrate, or to a foil under bending or a wire under torsion, as long as
due attention is paid to the appropriate value of \( h \) in each case.

In the applications of equations (1–3) the primary unknown is the plastic strain
distribution. It can be obtained within the strain-gradient theory by analytic means for
very simple cases such as the beam in bending [23], or by numerical methods [19].
However, these methods rely upon the approximation that the stress-strain relationship
implied by equations (1–3) is local. This is an approximation that is severely in error for
the simple strained layer, since only the material at the substrate – layer interface
experiences a plastic strain gradient, yet the full thickness of the layer is capable of
sustaining the stress \( M\varepsilon_E >> \sigma_Y \). Source operation and significant plastic deformation do
not depend upon conditions at a point, but upon conditions over an extended region
(source size) around the point, as recognised in nonlocal plasticity theories. Nevertheless,
the approximation can be good – this is best seen in the beam-bending or graded layer
problems. That is why, as observed by Liu et al. [10], the experimental data cannot test
between critical thickness theory and strain-gradient theory, for both will fit well.

It is worth commenting on the possible application of this analysis to other
gradient theories. Whenever the gradient term is multiplied by the yield or flow stress, as
in equation (3), and then the characteristic length turns out to vary as the inverse of the
yield or flow stress (or plastic strain), the separation we have done in equation (8) is
possible. This gives a gradient coefficient unrelated to yield or flow stress and then interpretations in terms of dislocation or defect spacing become inappropriate. From the review by Zhang and K. Aifantis [34], this seems to be the case for most gradient theories including those based on, or equivalent to, the Aifantis theories [24, 35].

In conclusion, it is demonstrated that the characteristic length in the FH strain-gradient theory can be obtained from known material and structural parameters, \( \ell_{FH} = 5b / c \varepsilon_Y, c \sim 1 \). The derivation shows that this SGP corresponds physically to critical thickness theory. It explains why SGP theories are capable of fitting experimental data. It validates the use of this theory to obtain approximate constitutive laws for use in finite-element calculations. It offers the prospect of understanding in general, on a secure physical basis, why strong metals are strong, and how to include size effects in rigorous engineering modelling and simulation.

**Acknowledgements:** Undergraduate students David Lewis and Viktor Jevdomikov assisted with strain-gradient theory. Discussions with the late Tony (A.G.) Evans, and with Dabiao Liu, Andy J. Bushby and Alan J. Drew are gratefully acknowledged.

**References**


Table I. Parameters in the critical thickness calculations for strained layers with $\sigma_Y = 0$. Symbols are defined in the text.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\varepsilon(z)$</th>
<th>$H_R$</th>
<th>$h$</th>
<th>$\varepsilon_E(z)$</th>
<th>$\bar{\sigma}$</th>
<th>$\varepsilon_P(z)$</th>
<th>$\bar{\varepsilon}_P$</th>
<th>$\bar{\varepsilon}'_P$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple layer</td>
<td>$\varepsilon_0$</td>
<td>$5b/\varepsilon_0$</td>
<td>$&gt; h_R$</td>
<td>$5b/h$</td>
<td>$5M\bar{b}/h$</td>
<td>$\varepsilon_0 - \varepsilon_E$</td>
<td>$\varepsilon_P$</td>
<td>$\varepsilon_P/h$</td>
<td>1</td>
</tr>
<tr>
<td>Thin grade</td>
<td>$g$</td>
<td>$\sqrt{10b/g}$</td>
<td>$h_R + \delta$</td>
<td>$z &lt; \delta$: 0 else: $g(z - \delta)$</td>
<td>$-\frac{1}{2}Mg h_R$</td>
<td>$z &lt; \delta$: $gz$ else: $g\delta$</td>
<td>$\sim g\delta$</td>
<td>$\sim g\delta/h$</td>
<td>$\sim 1$</td>
</tr>
<tr>
<td>Thick grade</td>
<td>$g$</td>
<td>$\sqrt{10b/g}$</td>
<td>$&gt;&gt; h_R$</td>
<td>$z &lt; (h - h_R): 0$ else: $g(z - h + h_R)$</td>
<td>$-\frac{1}{2}Mg h_R^2 h / h$</td>
<td>$z &lt; (h - h_R): g$ else: $g(h - h_R)$</td>
<td>$\sim \frac{1}{2}gh$</td>
<td>$\sim g$</td>
<td>$\sim 2$</td>
</tr>
</tbody>
</table>
Figure Caption

Figure 1. Characteristic lengths $\ell_{FH}$ are plotted against the tensile yield strains $\varepsilon_Y$. The length scales were found by fitting the FH theory to indentation data from the literature for Ir, Ag, Cu and superalloy and to foil-bending data for Ni. After figure 13 of reference 23. The solid line is the prediction of equation (7), for a typical value of $b = 0.25$ nm and with $c = 2$. 